Cincuenta Aniversario 1968 2018
UAM Universidad Autónoma de Madrid

Department of Chemistry

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Department of Chemistry
and Biochemistry

Dynamics of charged biomolecules: Fragmentation, molecular growth and solvation

A thesis submitted for the degree of
Doctor in *Theoretical Chemistry and Computational Modelling*

Candidate:
Giovanna D'Angelo

Supervisors:
Manuel Alcamí Pertejo
Pedro Alexandrino Fernandes

Madrid, December 2018

MARIE CURIE ACTIONS

With the contribution of the Atom Fysik group,
under the supervision of Henning Zettergren

Stockholm University

A mio padre

Abstract

Close range ion/atom collisions may deposit a large amount of energy on the target (molecule or cluster). The interaction projectile/target is dependent on the nature of the projectile, its charge state and its initial energy. Tuning the experimental conditions, thus the kinetic energy of the projectile, one may be in dominant nuclear stopping regime or dominant electronic stopping regime. The former implies that non-statistical processes, where the energy deposited is not evenly distributed among all the vibrational degrees of freedom of the target, are likely to play an important role in the target decay. The latter sees a largely dominant 'slow' statistical decay for the target. Understanding the relative weight of these two processes and unveiling the fine chemistry of the fragmentation pathways involved may found applications in understanding *e.g.*, the origin and evolution of complex molecules in space, the radiosensitivity of biomolecules and the role of the water environment in the process.

In this work we have used classical and *ab initio* molecular dynamics simulations combined with electronic structure calculations to reproduce entire sequences of collisions and to get insights on the fragmentation pathways induced by the energy deposition consequent to the collision. Chapter 3 summarizes the basics of the electronic structure methods used; while chapters 4 and 5 describe the classical and the ab-initio methods employed, respectively.

In this thesis we have investigated four four different systems: butadiene clusters, porphyrin, protonated and deprotonated, and 5-halo-substituted uracils. Polycyclic Aromatic Hydrocarbons (PAHs) are ubiquitous in space. Previous studies have already observed molecular growth, induced by non-statistical decay. In Chapter 6, we have explored the possibilities of forming ring structures, which may be PAHs precursors, from butadiene clusters.

We have used a combination of a reactive force-field potential (AIREBO) and a Coulombian screened potential (ZBL) to reproduce the nuclear scattering dominated processes visible in the experiments. The analysis of the results indicates the prompt atom knockout as the main responsible for the molecular growth observed in the experimental mass spectrum. DFT electronic structure calculations, at B3LYP/cc-pVDZ level of theory, suggest the formation of benzene-like ring structure products.

In the biological world, porphyrins are among the most common molecules, for this reason they have been proposed as a biomarker for assessing the origin of extra-terrestrial planets, meteorites, comets and so on. Chapter 7 explores the influence of knockout on the mass spectrum of 5,10,15,20-tetraphenylporphin (TPP) and its metallate derivatives, $FeTPP^+$ and $ZnTPP^+$, with a computational approach similar to the one adopted in Chapter 6. To the best of our knowledge, fingerprints of heavy atom knockout have not been identified so far in porphyrins mass spectra. The comparison between simulated and experimental mass spectra evidences the presence of heavy atom knockout. The calculations of the heavy atom knockout cross sections made us conclude they constitute the $\sim 40\%(\sim 70\%)$ of the total experimental cross sections of TPP in collision with He (and Ne).

Adenine is the nucleobasis which could best survive in an harsh environment like the Insterstellar Medium. In addition, the study of its protonated and deprotonated forms may give insights on the protonation and deprotonation effects in DNA. In Chapter 8, the stability of protonated and deprotonated adenine has been subject of a combined experimental and theoretical study. The computational approach consisted of a mixed statical and dynamical approach. We have performed *ab initio* molecular dynamics (ADMP type, at B3LYP/6-31++G(d,p) level of theory) in order to quantitatively reproduce the experimental mass spectrum features, such that fragments structures and fragmentation pathways could be identified.

5-bromo-uracil (5BrU) is a known mutagen and may be used as a model system for the radiosensitizer 5-bromo-2-deoxy-uridine. In Chapter 9, we have explored the influence of the water environment in the decay of 5BrU. This work uses a statical-dynamical approach similar to the one already described in Chapter 8 to study the nano-hydrated 5-bromo-uracil cluster. The *ab initio* molecular dynamics (BOMD type, at M06-2X/SVP level

of theory) suggested an unexpected channel: the water insertion on the $[5BrU-H]^+$ residue. The electronic structure PES exploration (at M06-2X/6-311++G(d,p) level of theory) of the neutral, ionized, protonated and $[5BrU-H]^+$ species has provided pathways and structures, which could explain some of the interesting features of the experimental mass spectrum of hydrated 5BrU, provided from our collaborators from Caen University and CNR Rome.

Following the results of Chapter 9, we have investigated the nucleophilic water attack on other halo-substituted uracil (5ClU and 5FU) and uracil (U) itself. As all these molecules are mutagens and/or model systems of radiosensitizers, we aimed, through the observation of similarities and differences in the chemistry of all of them, to shed some light on the chemical mechanisms of mutagenesis and radiosensitivity. The PES sampling involved all the neutral, ionized, protonated and $[5XU-H]^+$ species for each molecule. The results not only suggest an influence of the substituted moiety in position 5 on the reaction mechanism, but also an influence of additional water molecules when bounded to the nucleophile.

Resumen

Las colisiones de iones/átomos a bajo parametro de impacto pueden depositar una gran cantidad de energía en el sistema target (molécula o cluster). La interacción proyectil/target depende de la naturaleza del proyectil, su estado de carga y su energía inicial. Ajustando las condiciones experimentales, es decir la energía cinética del proyectil, uno puede estar en el régimen donde dominan los procesos nucleares para frenar el proyectil (nuclear stopping) o en el régimen donde dominan los procesos electrónicos para frenar el proyectil (electronic stopping). El primero implica que los procesos no estadísticos, donde la energía depositada no se distribuye de manera uniforme entre todos los grados de libertad vibracional del target, es probable que desempeñen un papel importante en la descomposición del mismo. El segundo caso (electronic stopping) implica un decaimiento estadístico 'lento' para el target. Comprender el superficie de energía potencialo relativo de estos dos procesos y descubrir la química de las rutas de fragmentación involucradas puede encontrar aplicaciones en la comprensión de *por ejemplo*, el origen y la evolución de moléculas complejas en el espacio, la radiosensibilidad de biomoléculas y el papel del entorno del agua en los procesos de fragmentación.

En este trabajo se han utilizado simulaciones de dinámica molecular clásica y *ab initio* combinadas con cálculos de estructura electrónica para reproducir secuencias completas de colisiones y obtener información sobre las vías de fragmentación inducidas por la deposición de energía como consecuencia de la colisión. El capítulo 3 resume los conceptos básicos de los métodos de estructura electrónica utilizados; mientras que los capítulos 4 y 5 describen, respectivamente, los métodos clásicos y ab-initio empleados.

En esta tesis hemos investigado cuatro sistemas diferentes: cluster de butadieno, porfirinas, adenina protonada y desprotonada, y uracilos substituidos en posición 5 por halógenos. Los hidrocarburos aromáticos policíclicos

(HAP) son ubicuos en el espacio. Estudios anteriores ya han observado crecimiento molecular, inducido por decaimiento no estadístico. En el Capítulo 6, hemos explorado las posibilidades de formar estructuras de anillos, que ppodrían ser precursores de HAP, a partir de cluster de butadieno. Hemos utilizado una combinación de un potencial de campo de fuerza reactivo (AIREBO) y un potencial Coulombiano apantallado (ZBL) para reproducir los procesos dominados por scattering nuclear que son visibles en los experimentos. El análisis de los resultados indica la eliminación de átomos por colisión directa como el principal responsable del crecimiento molecular observado en el espectro de masas experimental. Los cálculos de la estructura electrónica DFT, a nivel teórico de B3LYP / cc-pVDZ, sugieren la formación de productos de estructura de anillo tipo benceno.

En el mundo biológico, las porfirinas se encuentran entre las moléculas más comunes, por esta razón se han propuesto como biomarcadores para evaluar el origen de planetas extraterrestres, meteoritos, cometas, etc. El capítulo 7 explora la influencia de los knockout en el espectro de masas de 5,10,15,20-tetrafenilporfina (TPP) y sus derivados metálicos, $FeTPP^+$ y $ZnTPP^+$, con una estratégia computacional similar a la adoptada en el Capítulo 6. Hasta donde sabemos, los patrones característicos de knockout de atomos pesados no se han identificado hasta ahora en los espectros de masas de porfirinas. La comparación entre espectros de masas simulados y experimentales evidencia la presencia de knockout de átomos pesados. Los cálculos de las secciones eficaces de knockout de átomos pesados nos han llevado a la conclusión de que estos processos constituyen el $\sim 40\%(\sim 70\%)$ del total de las secciones eficaces experimentales del TPP en colisión con He (y Ne).

La adenina es la nucleobase que mejor podría sobrevivir en un entorno hostil como el medio insterstellar. Además, el estudio de sus formas protonadas y desprotonadas puede dar una idea de los efectos de la protonación y la desprotonación en el ADN. En el Capítulo 8, la estabilidad de la adenina protonada y desprotonada ha sido objeto de un estudio combinado experimental y teórico. El enfoque computacional consistió en una aproximación mixta con cálculos estáticos y dinámicos. Hemos ejecutado la dinámica molecular *ab initio* (tipo ADMP, en el nivel de teoría B3LYP / 6-31++G (d, p)) para reproducir cuantitativamente las características experimentales del

espectro de masas, de manera que se pudiesen identificar las estructuras de fragmentos y las vías de fragmentación de los mismos.

El 5-bromo-uracilo (5BrU) es un mutágeno conocido y también se puede usar como un sistema modelo para el radiosensibilizador 5-bromo-2-deoxi-uridina. En el Capítulo 9, hemos explorado la influencia del entorno de agua en la descomposición de 5BrU. Este trabajo utiliza un enfoque estático-dinámico, similar a lo que ya ha sido descrito en el Capítulo 8, para estudiar el cluster de 5-bromo-uracilo nanohidratado. Las dinámicas moleculares *ab initio* (tipo BOMD, en el nivel teórico M06-2X / SVP) sugirieron la existéncia un canal inesperado: la inserción de agua en el residuo $[5BrU - H]^+$. La exploración de la superficie de energía potencial (en el nivel teórico M06-2X / 6-311++ G(d, p)) de las especies neutras, ionizadas, protonadas y $[5BrU - H]^+$ ha proporcionado vías y estructuras, que podrían explicar algunas de las características interesantes del espectro de masas experimental de 5BrU hidratado, proporcionado por nuestros colaboradores de la Universidad de Caen y del CNR de Roma.

Siguiendo los resultados del Capítulo 9, hemos investigado el ataque nucleofílico del agua en otros uracilos sustituidos con halogenos (5ClU y 5FU) y en el propio uracilo (U). Como todas estas moléculas son mutágenos y/o sistemas modelo de radiosensibilizadores, pretendimos, mediante la observación de similitudes y diferencias en la química de todas ellas, arrojar algo de luz sobre los mecanismos químicos de mutagénesis y radiosensibilidad. El muestreo de superficie de energía potencial involucró a todas las especies neutras, ionizadas, protonadas y $[5XU - H]^+$ para cada molécula. Los resultados no solo sugieren una influencia del resto sustituido en la posición 5 sobre el mecanismo de reacción, sino también una influencia de moléculas de agua adicionales cuando se presentan enlazadas al nucleófilo entrante.

Table of contents

I Introduction **1**

1 On atom/ion Collisions with Molecular Targets **3**

1.1 Statistical and Non-Statistical Decay 5

1.2 Knockout Space Molecules 6

1.3 Principles of Radiation Damage 9

1.3.1 Time scales of the radiation damage 10

1.4 Space and earth molecules 11

1.4.1 Tetraphenylporphyrin 12

1.4.2 Adenine 13

1.5 Modeling radiosensitizers 15

2 Related Experimental Details **19**

2.1 AIRIBE: collisions with atomic ions 21

2.2 EIS lab: collisions with molecular ions 23

II Essentials of Theory **27**

3 Retracing fragmentation pathways from electronic structure calculations **29**

3.1 Potential Energy Surface Sampling 29

3.2 Solving the electronic problem 30

3.2.1 A (super)brief history of quantum theory 30

3.2.2 The Born-Oppenheimer approximation 31

3.2.3 Relying on the unknown: Density Functional Theory . 34

4 From static(s) to dynamic(s) 47
4.1 Classical Molecular Dynamics 49
4.1.1 Tersoff potentials' big family 50
4.2 Ziegler-Biersack-Littmark Potential 55
4.3 Modeling collisions with atom/ion projectiles 56

5 We don't do (just) classics 59
5.1 Pills of DFTB Formalism 60
5.2 *ab initio* Molecular Dynamics 62
5.2.1 Deriving Nuclear Classical Dynamics 65
5.2.2 Born-Oppenheimer and Atom-centered Density Matrix Propagation Molecular Dynamics 68
5.2.3 Main drawbacks of treating fragmentation at DFT level 69

III Results & Discussion 71

6 Ion-induced molecular growth in clusters of small hydrocarbon chains 73
6.1 Collisions with butadiene clusters 75

7 Knockout driven fragmentation of porphyrins 83
7.1 He/Ne collisions with porphyrin isolated molecules 84

8 Dissociative Decay of Protonated and Deprotonated Adenine 91
8.1 Fragmentation cross-sections 92
8.2 Protonated adenine stability 95
8.3 Deprotonated adenine stability 100

9 Direct evidence of water mediated keto - enol tautomerisation in the mutagen 5-bromouracil? 109
9.1 C^{4+} collisions with nano-hydrated clusters of 5BrU 111
9.2 Computational modeling of nano-hydrated 5BrU decay 114
9.2.1 Molecular dynamics simulations 114
9.2.2 PES sampling 120

10 The role of water insertion in uracil and 5-halo-substituted uracil **137**
10.1 Is the nucleophilic addition sensitive to the halo-substitution ?138
10.2 Water network effects 144

IV Final Remarks **151**

11 Conclusions and Perspectives **153**

12 Conclusiones y Perspectivas **157**

References **161**

Part I

Introduction

Chapter 1

On atom/ion Collisions with Molecular Targets

The work presented in this thesis aimed to shed light on the physical and physico-chemical phenomena involved in the energetic processing driven by the interaction of a target molecule or cluster (neutral or charged) with a projectile (charged or neutral).

When dealing with ion/atom collisions, to determine the amount of energy deposited and predict the following decay paths is not straightforward. For example, we might assist to long range weak interactions projectile-target (Figure 1.1, top row). In the case of an ionic projectile, if the electron absortion cross-section of the ion is big enough, one or more electrons will be taken from the electron density of the target. Usually, distant interactions do not deposit much energy on the molecule. On the other hand, we may have a close range collision (Figure 1.1, bottom row). In this case, the amount of energy deposited on the molecule is much bigger and it depends on the nature of the projectile, its charge state and its initial energy [1]. Figure 1.2 stresses this last sentence. The energy loss by the projectile per unit length the projectile passes through the target, the *stopping power*, may be split in two contributions: the *nuclear* and the *electronic* stopping. In Figure 1.2, we show resuts from calculations of nuclear and electronic stopping powers performed for both He (solid lines) and Ar (dashed lines) projectiles colliding with solid pyrene ($C_{16}H_{10}$) [2, 3] at different initial projectile energies. As we can see, the ratio nuclear/electronic stopping changes with the projectile

energy, favouring the nuclear stopping at lower energies and the electronic stopping at higher energies.

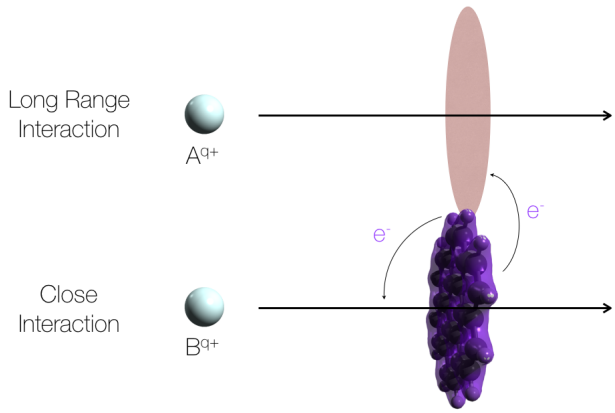


Figure 1.1: A simplified scheme for long range (top) and close range (bottom) interactions between a sample molecule (pyrene in this case) and a particle (A or B) of charge $q+$ with $q \in \mathbb{N}$.

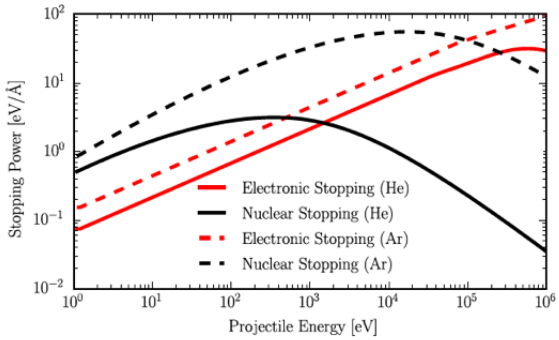


Figure 1.2: Nuclear (black lines) and electronic (red lines) stopping powers as calculated for He (solid lines) and Ar (dashed lines) atoms in collision with solid $C_{16}H_{10}$. These simulations were realised with the SRIM software [2, 3].

In addition, the heavier atom (Ar) deposits on average more energy then the lighter one (He); and the switching point from nuclear stopping dominated to electronic stopping dominated regime appears at higher energies for Ar

with respect to He. Given the mentioned energy dependance of the ratio nuclear/electronic stopping, one may tune the projectile energy to set the conditions to recreate a variety of environments. This considerably widens the number of possibilities one could explore and, consequently, also the possible applications and uses. In the case of this work, we may narrow the number of applications down to: energetic processing of molecules in the universe and the effects of radiation damage.

In this chapter, we will provide details about the physics behind the collisions with molecular targets (section 1.1) and introduce some basics of radiation damage (section 1.3). Afterwards, we will present the context of each work discussed here, starting from astrophysically interesting mechanisms (section 1.2) till exploring more biologically-oriented physico-chemical processes (sections 1.4 and 1.5).

It is important to point out that, in order to get meaningful insights in this field, it is essential to join experiments and theory. In the case of this specific work, the experiments were performed by our co-workers at the AIRIBE facility in Caen University (see section 2.1) and the EIS lab in Stockholm university (see section 2.2).

1.1 Statistical and Non-Statistical Decay

Molecular excitation is usually followed by a process through which the molecules cool down. This decay may occur in many different ways, among them, one is through the emission of photons or electrons with the molecule remaining intact (possibly as a different isomer). However, sometimes there may be remaining excitation energy high enough to produce a fragmentation. Once an excess of energy is deposited on a molecule, it might be:

- Evenly distributed all over the molecule;
- Localized. This event may be considered as a local vibrational excitation.

In both cases, if the excess of energy ends up to be redistributed across all the available vibrational degrees of freedom in the system before it fragments, then the most probable fragmentation channels are those with the lowest

energy profile. These events are referred to as *statistical* fragmentation processes. Because of this statistical nature, the fragmentation mechanism is largely independent of the initial excitation method. When the excitation energy is not redistributed among all the molecular degrees of freedom, the processes induced are referred to as *non-statistical*. Non-statistical fragmentation processes are, essentially, all the fragmentations that occur before the excitation energy of a system has been evenly distributed. This kind of fragmentation takes place when localized excitation energy leads to a rapid dissociation of the system, usually at the site of excitation. Because of this, the particular non-statistical fragmentation that takes place will depend on the excitation method and on the point of interaction. Fragmentation pathways that are weak or even absent in thermally driven processes may become important, and sometimes even dominant, in non-statistical processes. In order to study non-statistical fragmentation processes, it is essential to work in experimental conditions that allow to separate the non-statical from the statistical events. If we consider Figure 1.2, we can expect that working in a nuclear scattering dominated regime would be a good strategy to observe phenomena proceeding from the close interaction of the projectile with a restricted region of the target. The mentioned region might be as small as an atom, when this occurs the *knockout* of a single atom from the target is produced.

1.2 Knockout Space Molecules

Historically, space was thought to be practically empty. However, since the first identification of molecular species in the '30s [4–6], more than 200 different species have been identified [7].

In the *interstellar Medium* (ISM), the region of space between two stars, a good variety of molecules have been detected, from simple carbon-based molecules to more complex organic ones, like aminoacids, till prebiotic molecules [7]. One of the ubiquitous molecules in the universe are *Polycyclic Aromatic Hydrocarbons* (PAHs), basically condensed benzene rings. Figure 1.3 summarizes what has just been mentioned. Here, two different pictures of the galaxy Messier-82 are shown. On the left, an Hubble picture of the galaxy is shown, where the only light detected is the visible one, so one



Figure 1.3: Messier-82 galaxy, also known as cigar galaxy for its shape, as seen in the visible by the Hubble Space Telescope (on the left) and in combined visible-infrared by the Spitzer Space Telescope (on the right).

cannot appreciate the variety of molecular species inhabiting this region. Although, when detecting infrared light as well, a red shade surrounding the galaxy can be spotted (picture on the right). This red shade emission is associated with the presence of PAHs. The importance of this particular class of molecules in the universe seems to be related to its potential to produce H_2, which is essential for the collapse of molecular clouds in the early stages of stellar formation [8]. Chen et al. presented evidences according to which the excess of energy deposited with energetic ionic collision should be able to produce H_2 from PAHs [9]. Another work by Gatchell et al. pointed out the importance of knockout in PAHs molecular growth [10]. Considering that no specific PAH molecular structure has been identified yet in the ISM [7], here the reference system is a cluster of pyrene molecules in collision with Ar^+ and He^+ at 12 and 11 keV, respectively. The experimental results (Figure 1.4, left column) evidenced the presence of peaks due to the molecular species with presumed molecular formula $C_{16+m}H_{10}$ and $m \in \mathbb{N}$. The simulations, performed with a combination of AIREBO and ZBL potentials (see sections 4.1.1 and 4.2 for more information), demonstrated this molecular growth to be the result of knockout processes of C from other pyrene molecules in the cluster.

Taking this as a starting point, we have been wondering on if it was possible to build PAHs with the aid of this very same mechanism acting on clusters of smaller unsaturated hydrocarbons. One first evidence might come

from a provided evidence of the formation of benzene-like ring structures from these clusters (Figure 1.5). Our answers to this question will be discussed in Chapter 6.

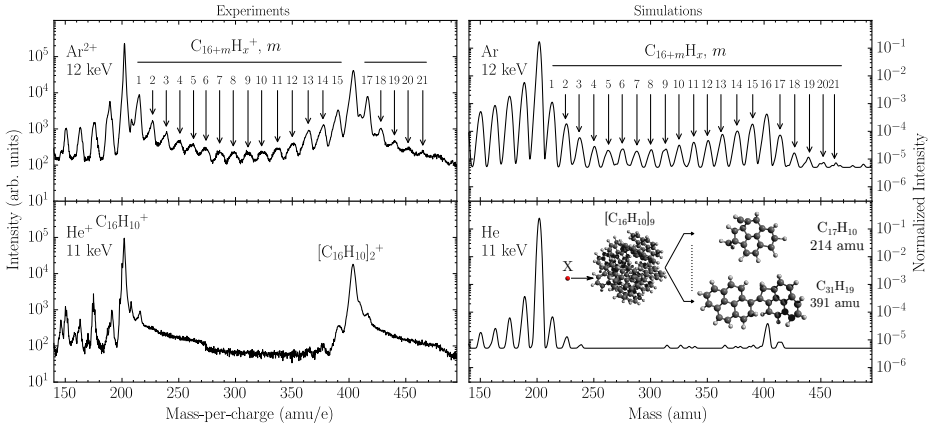


Figure 1.4: Experimental mass spectra (left column) from collisions between clusters of pyrene ($[C_{16}H_{10}]_k$) and ions of Ar^+ at 12 keV (top panel) and He^+ at 11 keV (bottom panel). On the right, the corresponding simulated spectra for collisions of $[C_{16}H_{10}]_9$ with neutral Ar and He are shown. Growth products with 17-37 C additional carbon atoms are observed in the mass spectra with Ar. The $[C_{16}H_{10}]^{+2}$ peak in the experimental spectra is mainly from loosely bound dimers of intact molecules, not included in the simulated spectra. The Figure is a Michael Gatchell's adaptation from [10].

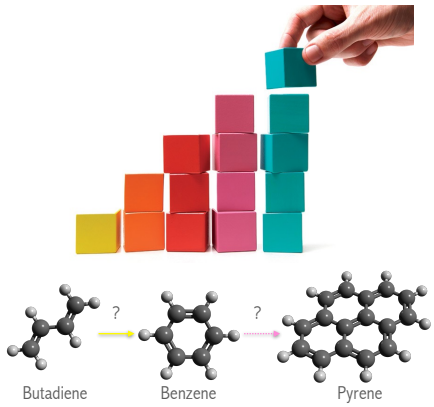


Figure 1.5: A schematic guess of how to build PAHs from butadiene.

1.3 Principles of Radiation Damage

Ionizing radiation may induce damage in biological tissues; nowadays these damages are considered the results of direct or indirect effects on DNA molecules [11]. Direct effects are the consequence of the radiation directly interacting with the DNA molecules; whereas, we refer to indirect effects as the result of the interaction with the surrounding media, generally water molecules. In particular, radiolysis of water produces *reactive oxygen species* (ROS), which may further react with DNA [12]. The DNA damage induced by tightly bound water molecules seems to be due to charge transfer in a phenomenon called the *quasi-direct effect* [13, 14]. The fact that no hydroxyl radicals and trapped electrons were detected below about 10 and 21 water molecules per nucleotide, respectively, has been presented as evidence to this effect [15]. However, one must stress that most of the commonly used radiations, like high-energy photons or swift light ions, generate a significant amount of ionization in the water molecules surrounding the DNA. All of this, together with the incredible complexity of a cell environment, makes the experimental decoupling of the ionization of water molecules from the direct DNA ionization not an easy task to accomplish.

Modern radiotherapy makes large use of ionizing radiation [16]. In particular, proton- [17] and hadron-therapies [18] have emerged in these last years. The reason of this success may be found in Figure 1.6. Here, the released dose (of radiation) is reported with respect to the depth in tissue. Looking at the H^+ (yellow area) and C^{6+} (purple area) beams, it is evident that they deposit most of their energy at a specific depth. This peak of energy loss is called the *Bragg peak*, which is specific to each ionizing radiation. The most interesting feature, although, pops up when we compare these peaks with the one of the X-ray radiation (green area), which is the most traditional wavelength for radiotherapy. In this case, in fact, the Bragg peak is wider and it is located at lower depths with respect to the H^+ and the C^{6+} ones.

As a result of this, the ion beams are more selective towards the cancer, and their use makes possible to limit the damages to the surrounding healthy tissues.

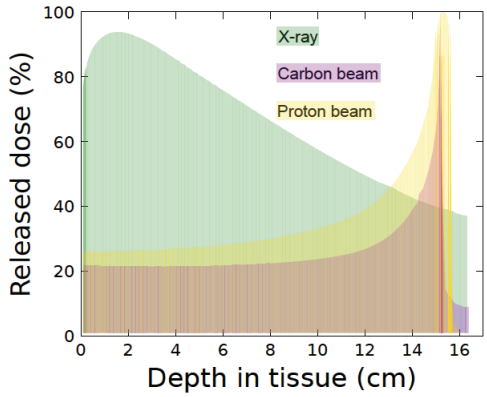


Figure 1.6: Plot of the released dose (in percentage) vs the depth in tissue (in cm) reached by X-ray (green area), proton beam (yellow area) and C^{6+} beam (purple area).

1.3.1 Time scales of the radiation damage

Reconstructing the causality chain that connects the biological damage (in its extremes: cellular death, the desired process when treating cancer, or DNA mutation and possibly cancer, negative consequence of exposure to radiation) to the initiating physico-chemical event, *i.e.* ionization, is not as straightforward as it might look. When trying to computationally simulate the processes involved in the interaction ionizing radiation-biomolecule, one has to mind the time scale of each event. The following tentative list quantifies some significant time-scales :

1. Collision, 10^{-15} s;
2. Physico-chemical events that may lead to DNA damage, 10^{-15} to 10^{-12} s
3. Chemical events responsible of the formation of new species, up to 1 s;
4. Enzymatic attempt to repair the DNA, min to hours.

Step 4 is crucial for the life of the modified cell. As a matter of fact, if the enzymatic repair is unsuccessful, the cell may die in a few hours. Alternatively, if the mutation is incorporated in the cell DNA and it does not provoke its death, it makes the cell dysfunctional. Accumulating dysfunctional cells in years generates cancer (Figure 1.7).

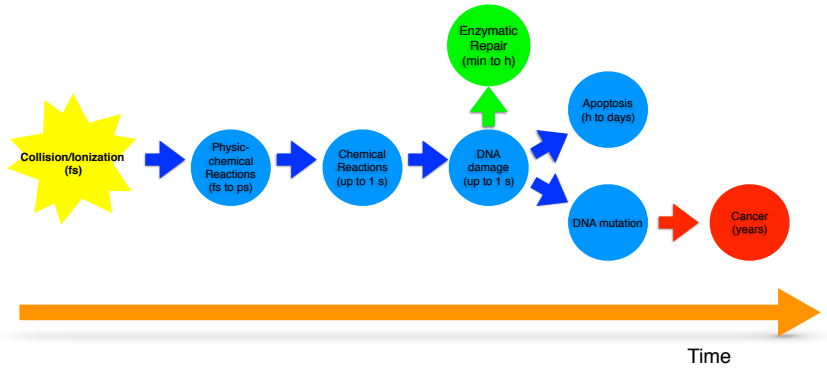


Figure 1.7: Tentative events chain starting from the interaction radiation-tissue.

Computational chemistry, especially through molecular dynamics simulations, may be extremely useful in this context. Understanding the fine chemistry of the processes involved in this chain is useful, in first place, to both prevent the induced damages and plan possible solutions to it. Additionally, it follows naturally that understanding the physico-chemical mechanisms of the ion-therapy action may help designing new and more effective therapies. However, our possibilities are limited by the time-scales and the system size. In fact, *ab initio* molecular dynamics reaches a few picoseconds simulation time and may treat just a hundred atoms, with high computational cost. In order to end calculations in the range of human life time, the accuracy of the methods employed to treat bigger systems must be reduced accordingly (Figure 1.8).

1.4 Space and earth molecules

Tetraphenylporphyrin (TPP) and *adenine* in this project sit between systems of pure astronomical interest and the ones of pure biological interest. Such that, understanding their energetic processing is useful to deepen the knowledge on their interesting physico-chemical mechanisms in the ISM or extra-terrestrial environments, on one side. On the other side, considering them as simpler and computationally more affordable models of bigger biological structures, their study provides interesting insights on their response to the exposure to an ion beam.



Figure 1.8: Scheme illustrating the progressive reduction of the level of theory accordingly to the increase of the time scale and the size of the system under investigation.

1.4.1 Tetraphenylporphyrin

Even if there is no definitive evidence of the presence of porphyrins in the interstellar medium, Johnson et al. postulate the existence of tetrabenzoporphyrins ($MgTBP$ and H_2TBP) within interstellar paraffin grain matrices [19]. In addition, Suo et al. proposed porphyrins as an ideal biomarker to unambigously identify the presence of molecules originally biological in the extra-terrestrial environment [20]. Last but not least, substituted porphines are common in nature and have different biological functionalities. In this latter context, they usually exist as metal coordinating compounds (*metalloporphyrins*, MP). As a matter of fact, iron containing porphyrins (hemes) are, for instance, fundamental for oxygen transport and storage in hemoproteins [21, 22]. Magnesium porphyrins are essential to photosynthesis [23]. We decided to use 5,10,15,20-tetraphenylporphin (TPP or simply tetraphenylporphyrin) and its metalled counterparts, 5,10,15,20-tetraphenyl- 21H,23H-porphinato iron(III) and 5,10,15,20-tetraphenyl- 21H,23H-porphinato zinc (III)(MTPP, with $M = Fe$ and Zn, respectively) as model systems for our studies on porphyrins energetic processing induced by collisions.

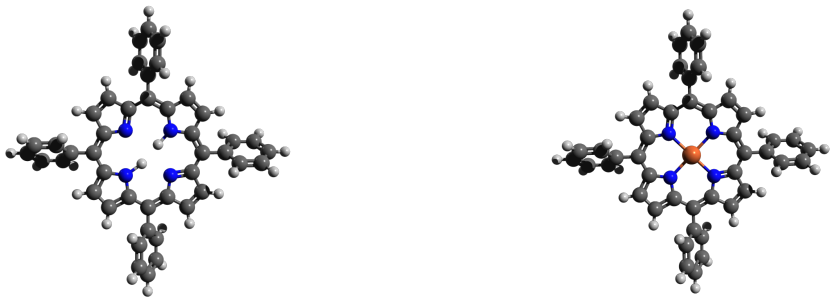
A substantial amount of studies on MTPPs electron impact (EI) [24–28], collisions with ions or atoms [29–31], or photoinduced dissociation (PID) [32, 33] exists. All these experiments may contribute to a better understanding of damage to biomolecules. However, previous studies suggest that the energetic ion traversal of a biological medium may produce hyperthermal scattering of ions (kinetic energy range: 1-100 eV), which may, in turn, induce severe damage to biological tissues [34–36]. This seems to be especially true at penetration depths just before the primary beam is stopped (at the tail of the Bragg peak) [31]. Moreover, to the best of our knowledge, direct evidence of knockout driven fragmentation has, so far, not been considered in connection with collision induced dissociation of biomolecules. This may be due, on one side, to the low dissociation energies of many biomolecules and their fragments, which is responsible for several further fragmentation steps following the initial interaction. Additionally, often biomolecules have irregular three-dimensional structures, this type of molecules is particularly subjected to secondary knockouts, which make the interpretation of the data so involved that identifying the fingerprint of knockout is not straightforward. Finally, the peculiar rigidity of porphyrin structures, caused by their π delocalized orbitals (Fig. 1.10), together with the prevalence of nuclear stopping in the energy regime employed in the experiments, makes them suitable for studies of non-statistical (knockout driven) fragmentation [37].

The studies carried out on collision driven reactions of TPP and MTPP (M = Fe and Zn) are discussed in Chapter 7.

1.4.2 Adenine

Several organic compounds, *i.e.* aminoacids [7], have been identified as components of the interstellar medium, but the presence of nucleobases has never been detected. However, experimental evidences demonstrated that nucleobases are more stable than amino acids when irradiated by VUV or soft X-rays. This is most probably due to their heterocyclic structures and unsaturated bonds [38]. Among all the possible nucleobases, adenine has shown to be the most likely to survive in such a harsh environment as the ISM [39]. Not only, precursors of nucleobases (*e.g.* HCN, pyrimidine, pyridine, imidazole) have been found in dense molecular clouds [39], meteorites [40, 41], on surfaces of comets [42, 43] and in Titan atmosphere [39, 44]. Assuming

Figure 1.9

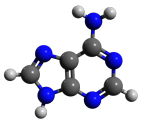


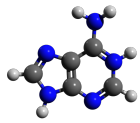
(a) TPP (b) FeTPP

Figure 1.10: (a) Tetraphenylporphyrin (TPP) and (b) Tetraphenylporphyrin(Fe)$^+$ structures. Porphyrins are a group of heterocyclic macrocycles, composed of four modified pyrrole subunits interconnected at their α carbon atoms via methine bridges (=CH-). In TPP the meso positions are substituted by phenyl groups (-C_6H_5).

that these precursors are the building blocks of nucleobases, then how do we form them?

Because of the already mentioned extra-resistance, adenine has been considered an interesting starting point for this type of study. As a matter of fact, several studies already exist that aim to answer the previous question for the case of adenine. Some suggest adenine comes from the HCN pentamerization through photoactivation [45], and ammonia or water catalysis [46]. Others indicate pathways that do not involve HCN at all [47]. The reconstruction of the molecular formation mechanism in gas phase through the identification of the inverse reaction is a commonly used strategy. In this context, *collision induced dissociation* (CID) is a well known technique used to investigate the structure of the nucleobases [48]. The protonated molecule has been largely investigated. In precedent studies, protonated and radical cation adenine fragments have reported sequential losses of HCN [49–53], independently from the ionization technique and/or fragmentation process, supporting the idea of HCN as an adenine precursor. Fewer studies investigated the fragmentation of the deprotonated adenine molecule [39, 48]. In these previous studies, the fragmentation of this anionic molecule appear to proceed mainly through the loss of HCN and NCHNH [48]/HNCNH [39], not very different with respect to the protonated case.





(a) Ad (b) [Ad-H]$^-$ (c) [Ad+H]$^+$

Figure 1.11: Structures of the most stable isomers [39, 54] for: (a) adenine (Ad); (b) deprotonated adenine ([Ad-H]$^-$); (c) protonated adenine ([Ad+H]$^+$).

Finally, it has been shown that slow electrons [55, 56] or ions [57, 58] are responsible of single and/or double strand breaks [59] and the consequent cell death or cancer [60]. Adenine is a DNA basis. In this other scenario, the study of protonated and deprotonated adenine (Figure 1.11) cross-sections may be useful to compare the fragility of the original molecule to the cases when a proton is added or removed from the DNA bases. Our study on deprotonated and protonated adenine aimed to contribute to this animated discussions. The results obtained are reported in Chapter 8.

1.5 Modeling radiosensitizers

We have already discussed the importance of understanding the fine mechanisms of radiation damage in order to improve the performance of ion therapy (section 1.3). As it usually occurs, there is not just a unique way to pursue this aim. For example, there are some compounds available in medicine, which are able, if injected in a patient, to enhance the cells radiosensitivity [61, 62]. Such a compound amplifies the capability of radiation to induce a mortal DNA mutation to the cells, often because it is a mutagen itself. Among all the available molecules eligible for this function, we decided to focus on *5-bromouracil* (5BrU) as a model system of the radiosensitizer 5-bromo-2-deoxy-uridine. 5BrU is a base analog, a molecule structurally similar to a natural occurring DNA basis. Apart from the nowadays famous Watson-Crick complementary base pairing scheme [63], other more unusual pairing scheme, *i.e.* with base analogs, may occur. This natural phenomenon goes under the name of *base mispairing* [64]. In particular, 5-bromouracil is

a thymine (T) analog and, in its keto-form, it pairs with adenine (A). Its mutagenicity seems to be due to the fact that 5BrU is more often in its enolic form than thymine is, and in this form it can base pair with guanine (G). Because of this, it also affects base pairing during DNA replication, which leads to point mutations [64, 65]. 5BrU is also more reactive to ultraviolet light than T, and this may also increase the level of mutation. Experiments have shown that 5BrU typically seems to cause $GC \rightarrow AT$ transitions [66].

Considering the importance of this molecule and its effects, lot of works already exist which try to elucidate the fine chemistry of the mutagenic action of 5-bromo-uracil [67–74].

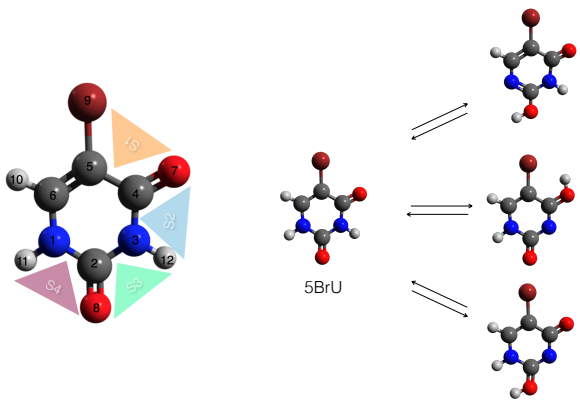


Figure 1.12: On the left: a scheme of 5BrU with atoms numbered and the molecules divided into four hydration zones: S1, S2, S3, S4, as in [70]. On the right: Tautomerization of 5BrU in its enolic forms.

The keto-enolic equlibria shifted by the presence of water assisting the proton transfer have been looked with interest since very early [63, 75]. This impulsed many later studies on the tautomers and their stability in water [72, 76, 77]. However, the absence of sufficient experimental evidence of this mechanism lead to different theories on the mechanism of 5BrU mutagenicity [67–74]. One of the mechanisms proposed involves the deprotonation of N3 (Figure 1.12), which, increasing the acidity of 5BrU compared to U, would favour the formation of complexes with guanine [71]. On the other hand, Adams et al. postulated that the electronegativity of the bromine atom increased the cross section for trapping of radiation-produced electrons [78]. He suggested that the negative bromouracil species thus produced

dissociates to a bromide ion and a uracilyl radical. Zimbrick et al. confirmed the dissociative electron attachment of hydrated electrons to 5-bromouracil [79]. The resulting uracilyl radical is so reactive it can abstract hydrogens from the 2'-carbon atom of the neighboring sugar. As a consequence of this hydrogen atom abstraction, a strand break will be produced and a uracil moiety will be introduced into the DNA.

We decided to participate in this debate: is it really tautomerization the mechanism responsible for 5BrU mutagenicity ? Chapter 9 presents our results and discussion.

5-bromouracil is not the only halo-substituted of uracil that works as a mutagen and may be used to model the radiosensitizer action; in fact, *5-fluorouracil* is another example [80]. Moreover, uracil itself is a mutagen. Following the results obtained for 5BrU, we have tested the same mechanisms on other halo-derivatives (5-clorouracil, 5ClU, and 5-fluorouracil, 5FU) and uracil (U) (Figure 1.13) and compared the results with 5BrU. The results are collected in Chapter 10.



(a) U (b) 5FU (c) 5ClU (d) 5BrU

Figure 1.13: Chemical structures of: (a) uracil (U); (b) 5-fluorouracil (5FU); (c) 5-clorouracil (5ClU); (d) 5-bromouracil (5BrU).

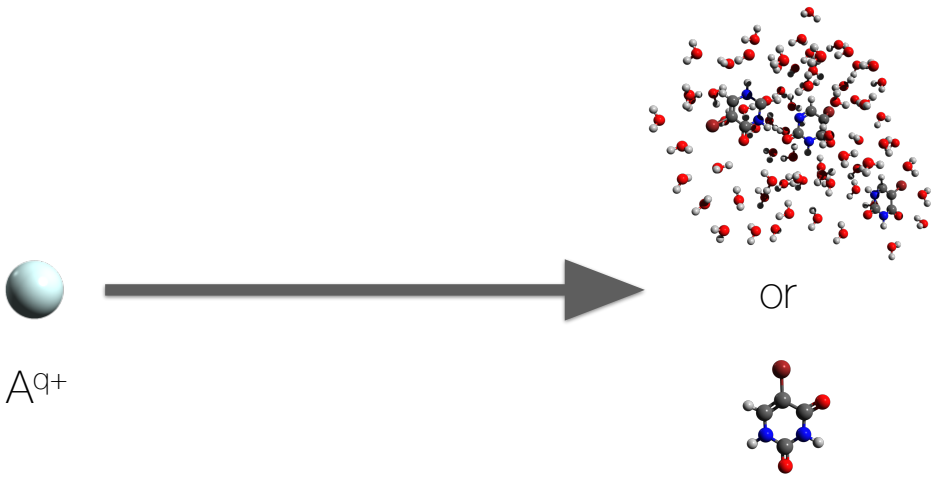
Chapter 2

Related Experimental Details

The experimental work included in this thesis (for the purpose of completeness) was performed by our collaborators in two different facilities. One is part of the ARIBE beam line at the Ganil facility in Caen University, France [81]. Chapters 6 and 9 discuss experiments performed with this technique and the corresponding theoretical calculations. The second is the EIS lab [82], part of DESIREE facility in Stockholm University, Sweden [83, 84]. The experiments performed in this facility are treated in Chapters 7 and 8, together with the related computational results.

The type of collisions generated is different in the two instruments. At the AIRIBE facility, the collisions involve an ionized atom accelerated toward a molecular target (either cluster or isolated molecule) (Figure 2.1a). The EIS lab apparatus uses charged molecular projectiles, which collide with atomic targets (Figure 2.1b). In both cases the collision energy is high enough to induce fragmentation in single or multiple steps, depending on the impact parameters.

The biggest implication of the difference in mass of the projectile ions is the marked difference in their velocities. As a matter of fact, even if the laboratory reference frame is of the order of keV for both the instruments, a heavier projectile will have a lower velocity than a lighter one. Taking into account this, a useful quantity for comparing results from these two different types of experiments is the collision (kinetic) energy in the reference frame



(a) AIRIBE facility collision



(b) EIS lab collision

Figure 2.1: Scheme of collision for: (a) AIRIBE facility and (b) EIS lab.

of the combined center-of-mass of the projectile and the target. This value is frequently referred to as the *center-of-mass energy* (E_{CoM}), which for a projectile kinetic energy in the lab reference frame of E_{lab} is given by the 2.1.

$$E_{CoM} = \frac{m_{target}}{m_{target} + m_{projectile}} \times E_{lab} \tag{2.1}$$

Equation 2.1 tells us that in the case of a lighter projectile (= charged species) we will be much closer to the E_{lab}, meaning we are able to reach higher collision energies in the AIRIBE case (10 - 200 keV) than in the EIS lab case (20 eV - 1 keV).

2.1 AIRIBE: collisions with atomic ions

technique The multiply charged ions are produced in an *Electron Cyclotron Resonance* (ECR) source (Figure 2.2) , which is characterized by a magnetic field of 1.4 T and a 1.2 T hexapole field. This magnetic field is used for the electron confinement. A tunable radio frequency (RF) electric field is applied that accelerates the electrons until their motion is in phase with this field. A flux of the neutral atomic projectile gas continuously passes through the ECR source. The mentioned atoms are ionized in the source by multiple electron impacts. The charge state distribution of the ions created may be selected by varying the electron energy and the time the ions spend in the source. In order to reduce the degradation of the emittance by *Coulomb explosion* (due to the high current density at the extraction region), a high voltage of 20 kV is used to extract multiply charged ions from the ion source [81]. Afterwards, a mass-per-charge selection of the emerging ions is performed by a bending magnet.

After m/q separation, the selected ions are decelerated from the negative potential V_{mag} to ground. Along this path, the ion beam is vertically and horizontally shaped by an electrical quadrupole triplet in a first diagnostic chamber. At this point, the beam is chopped into pulses approximately 500 ns in length with a repetition rate of a few kHz. Finally, the ion pulse train so generated is guided by a series of magnetic beam guides and lenses to the interaction region of a Wiley-McLaren *time-of-flight* (TOF) mass

spectrometer [85], where it meets the neutral beam. In this system, the TOF has a mass resolution, $m/\Delta m$, of several thousands [81].

A resistively heated oven evaporates the solid or liquid target material (as it is the case for 5-bromouracil). Sometimes the target is in gas phase at room temperature (as 1,3-butadiene), in this situation the sample is just injected directly through a nozzle into the interaction region of the experiment (if the target is going to be an isolated molecule) or into the cluster aggregation source (for molecular cluster target) from a bottle.

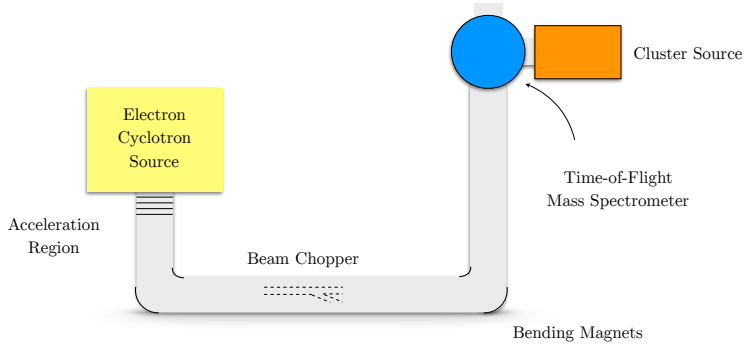


Figure 2.2: Scheme of the entire setup at the AIRIBE facility.

The cluster aggregation source contains a He buffer gas and is cooled with liquid N_2 at 77 K. The cold He atoms will cool down the molecules in gas phase, which condense into weakly bound clusters held together by dispersion forces. It is not possible to control the exact size of the clusters produced in the neutral beam source, in reality a distribution is generated. The intensities of the peaks corresponding to the generated clusters follow a broad log-normal size distribution. The temperature of the oven (T_{oven}) influences the mentioned distribution. T_{oven} is directly proportional to the target pressure (p_{target}) in the gas mixture and, consequently, with the amount of material available for the growth of the cluster. So an increase of the p_{target} leads to a shift of the distribution towards larger sizes and a broadening as well. The carrier gas pressure (p_{He}) and its molecular weight also have some influence on the cluster size, but inferior compared to p_{target}.

When generating water clusters of a particular species, as it was the case for 5-bromouracil-water clusters, a water source at 50 °C is placed on the

entrance canal of the carrier gas. The mixture of He, H_2O and 5BrU will get to the condensation canal, where 5BrU water clusters are actually produced.

After the interaction region, where the clusters are ionized and extracted, they will be accelerated towards the detector (Figure 2.3). At this point, the ionized clusters will traverse a field-free drift region, where cluster ions drift and separate in time. Before reaching the detector, in order to guarantee the detection of heavy ionized clusters, these ones need to be post-accelerated. These cluster ions then will hit a *secondary electron emitter* (conversion plate in Figure 2.3). Finally, the secondary electrons are accelerated and guided to the two-stage *micro-channel-plate* (MCP) assembly by external Helmholtz coils.

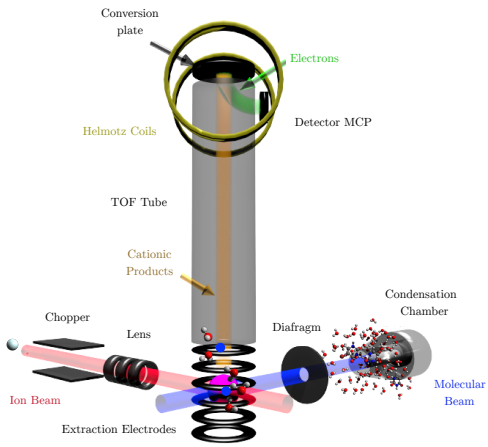


Figure 2.3: Scheme of the Ion crossed beam experiment at the AIRIBE facility.

2.2 EIS lab: collisions with molecular ions

Figure 2.4 presents a schematic representation of the apparatus in Stockholm University [82].

The molecular ions are produced in an *electrospray ionization* (ESI) source. This source consists of a stainless steel needle mounted in air followed by a heated stainless steel capillary as an interface to the vacuum region. A motor-driven syringe provides a constant flow of a sample solution, typically around $1 - 2\ \mu l/min$ to the needle via a fused silica wire [82].

The composition and concentration of the injected solution depends on the particular analyte under investigation. For the TPP, a 20 mM solution in methanol:acetic acid (99:1) was used, the FeTPP chloride salt and the ZnTPP were employed to make a 20 mM solution of methanol:toluene (50:50). In the case of protonated adenine, the solution is water:methanol (50:50 in volume) with 5% acetic acid [86]. The resulting concentration is 0.1 mmol/l in protonated adenine. While for deprotonated adenine, the solvent was methanol:acetonitrile (20% : 80% in volume) with some droplets of ammonium hydroxide.

The needle is set on a potential of typically 3-4 kV with the appropriate polarity (as for all the potentials mentioned hereafter), while the capillary is biased at 100-300 V. An ion funnel provides a collisional focusing effect towards the beam axis and increases the ion transmission through the application of a radio frequency (RF) and direct current (DC) fields with a background gas pressure of $\sim 10^{-1}$ mbar [87]. The high potential of the needle produces a charged aerosol. The small droplets begin to evaporate from the energy gained in collisions with residual gas molecules in the capillary and ion funnel. There is a threshold size, at which the repulsive Coulomb forces from the ions within will overcome the surface tension and the droplets will release isolated charged molecules of both the solvent and the analyte. The mentioned method is a soft ionization method and suitable to treat molecules with low dissociation energies, *i.e.* biomolecules.

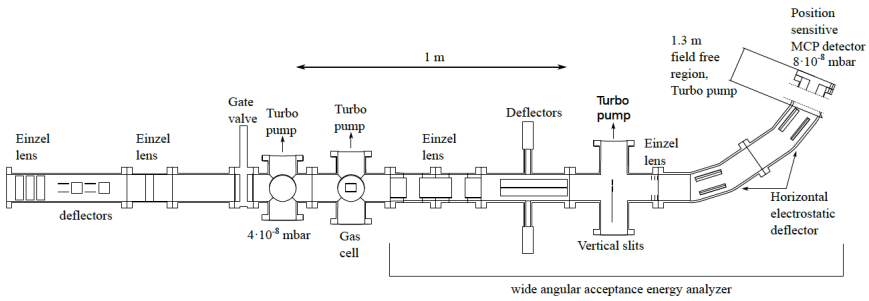


Figure 2.4: Scheme of the EIS lab facility [82].

The ions collected by the ion funnel are guided into the octupole, which is used as an ion beam guide. After passing a second octupole guide, the

ions meet an Extrel CMS quadrupole mass filter with a m/z range of 2-4000 amu/e. A combination of RF and DC voltages determines the stable ion trajectories through the filter for each mass [88]. After the mass filter, a quadrupole deflector deviates the ions to the CID part of the experiment. The latter case is the interesting one in our studies. A series of Einzel lenses are used to steer and focus the ion beam for the passage through a collisional chamber. Here the pressure of the neutral targets is regulated according to the specific experiment needs.

Charged products emerging from the gas cell are analyzed by means of a wide angular acceptance energy analyzer. This section is constituted of pairs of electrostatic deflector plates and a position sensitive MCP detector at the end of the apparatus. During the collision some energy is deposited and fragments are produced, this means that in the case of intact molecules the energy when exiting the collisional chamber will be almost the same as the one they had when they entered the cell. The fragments, though, have lower masses hence lower kinetic energies. The voltage of the plates is scanned to deflect ions with different kinetic energies onto the detector. This means that the ions are analyzed according to their energy-to-charge ratios. As a matter of fact, an energy spectrum is then produced by recording the deflection voltages and the measured positions on the detector.

The large angular acceptance of the deflectors after the gas cell (allowing fragments of the same mass but due to collision processes with different inelasticities to be detected) and the relatively large spread of energies in the incoming ion beam increase the uncertainty on the mass-to-charge ratios measured in Stockholm, compared to the one from Caen.

Part II

Essentials of Theory

Chapter 3

Retracing fragmentation pathways from electronic structure calculations

3.1 Potential Energy Surface Sampling

A good part of the present work consisted in exploring potential energy surfaces (PESs) for the molecules under investigation. Every molecule we can imagine exists in an infinite number of possible geometries (the atoms positions in space), at each of which corresponds a value of the total energy. The geometry of a set of atoms can be described by a vector $\boldsymbol{r}$, whose elements are the atoms positions. If we plot in a 3D space each value of the energy corresponding to every possible geometry ($E(\boldsymbol{r})$) of a single molecule in function of a set of geometric parameters like a certain bond length, an angle or a dihedral angle, we will obtain a potential energy surface. So the potential energy surface is nothing more than a convenient geometrical representation of the function $E(\boldsymbol{r})$. As we can expect, considering that a molecule may have various bond lengths, angles and dihedral angles, the potential energy surface might be extremely large. On the other hand, depending on the system under investigation, it is possible to reduce the sampled area of the potential energy surface.

In our particular case, knowing the composition of the interesting fragments to investigate, we have been exploring pathways suspected to lead

to the desired products. To this end, the numerical search for minima and maxima of the PES, respectively corresponding to stable structures and transition states, has been carried out with the aid of relaxed scans of the molecular coordinates. These scans consisted of a step by step increasing of specific bond lengths, at each step the rest of the atoms are allowed to relax and the corresponding minimum energy E(r) value is calculated. Note that for each value of the energy an optimization of all geometrical paramenters except one has to be done. This particular procedure allows one to explore sections of the PES and makes the transition state search easier. On the other side, also *Synchronous Transit-Guided Quasi-Newton* method [89] (in both the QST2 and QST3 variants) has been used to find transition states. Once the desired transition state was found an *Intrinsic Reaction Coordinate* (IRC)[90] calculation has been performed. Here, starting from the transition state, the paths leading to the two connected minima are followed, such to obtain the particular energy profile for the specific pathway. At this point of our discussion, it is convenient to give some insights on the numerical methods used to perform the mentioned PES sampling.

All the PES calculations performed in this work were carried out with the gaussian09 package [91].

3.2 Solving the electronic problem

3.2.1 A (super)brief history of quantum theory

Beyond the deterministic classical mechanics, where one can calculate the future (and past) positions and velocities of all particles at any other time, a number of physical observations forced physics community to elaborate the more general modern quantum theory. Planck showed in 1900 that electromagnetic radiation was emitted and absorbed from a black body in discrete quanta, each having energy proportional to the frequency of radiation. In 1904, Einstein invoked these quanta to explain the photo-electric effect. All of this implies that, under certain circumstances, one must interpret electromagnetic waves as being made of particles. In 1924 de Broglie asserted that matter also had this dual nature: particles can be "wavey". Even if the foundation of the new quantum theory is attributed to both Erwin Schrödinger and Werner Heisenberg, we, quantum chemists,

grow fond in the time-dependent Schrödinger equation (TDSE)(3.1). The TDSE is a linear partial differential equation, which suitably describes the time evolution of the wave function associated to a physical system.

$$\hat{H}(\boldsymbol{r}, \boldsymbol{R}, t)\Psi(\boldsymbol{r}, \boldsymbol{R}, t) = i\hbar\frac{\partial\Psi(\boldsymbol{r}, \boldsymbol{R}, t)}{\partial t} \tag{3.1}$$

$\hat{\mathrm{H}}$ is the hamiltonian operator, which contains all the relevant forces acting on the particles of the system, specifically nuclei and electrons. Ψ is the wave function of the many-body system (in our case molecule, cluster etc). It is the complex object which contains all the information regarding the system. $\boldsymbol{r}$ and $\boldsymbol{R}$ are the vectors that include, respectively, electronic and nuclear positions, whereas t corresponds to time. Solving the TDSE exactly for a polielectronic system is actually impossible. At this point, a series of approximations may be used to numerically solve equation 3.1. First of all, at this point of our dissertation we do not need to discuss the TDSE, so, for the sake of convenience, we can refer to the simpler time independent Schrödinger equation (TISE) (equation 3.2), which is obtained from the previous when the potential acting on electrons and nuclei does not depend on time. The solution of equation 3.2 yields the discrete (*quantized*) values (or *eigenvalues*) of energy E and for each E, it corresponds a wavefunction Ψ. When multiplied by its conjugate complex, Ψ delivers a probability density. Quantum mechanics is thus not deterministic, but *probabilistic*. We enter now the reign of probability.

$$\hat{H}(\boldsymbol{r}, \boldsymbol{R})\Psi(\boldsymbol{r}, \boldsymbol{R}) = E\Psi(\boldsymbol{r}, \boldsymbol{R}) \tag{3.2}$$

All our problems start and end with equations 3.1 and 3.2. All our efforts are devoted to find good approximations and methods to find accurate enough solutions to our chemical questions. To get to the point where we are able to successfully solve our first quantum chemical calculation, we need to move forward and introduce the first approximation we need: the Born-Oppenheimer approximation [92].

3.2.2 The Born-Oppenheimer approximation

The Born-Oppenheimer approximation starts from the consideration that the nuclei masses are more than a thousand times bigger than the electrons

ones. Thus, we can assume that the electrons are going to move faster than the atoms. So fast that the nuclei can be considered fixed with respect to the electronic movement. Therefore, nuclear coordinates may act as parameters in equation 3.2 (the semi colon separates parameters from variables). This allows us to express Ψ as a product of a nuclear and an electronic wave function (3.3).

$$|\Psi(\boldsymbol{r}, \boldsymbol{R})\rangle = |\psi_k(\boldsymbol{r}; R)\rangle \, |\chi_N(\boldsymbol{R}))\rangle \tag{3.3}$$

Equivalently, given the non-relativistic Hamiltonian to be as in 3.4, neglecting nuclei kinetic energy ($\hat{T}_N(\boldsymbol{R})$), we obtain an efficient Hamiltonian ($\hat{H}_{eff}(\boldsymbol{r}; R)$), which may be used to write an electronic Schrödinger equation (3.5)

$$\hat{H}(\boldsymbol{r}, \boldsymbol{R}) = \hat{T}_N(\boldsymbol{R}) + \hat{T}_e(\boldsymbol{r}) + \hat{V}_{eN}(\boldsymbol{r}; R) + \hat{V}_{NN}(R) + \hat{V}_{ee}(\boldsymbol{r}) \tag{3.4}$$

$$\begin{aligned} \hat{H}_{eff}(\boldsymbol{r}; R)\psi_k(\boldsymbol{r}; R) = \\ \{\hat{T}_e(\boldsymbol{r}) + \hat{V}_{eN}(\boldsymbol{r}; R) + \hat{V}_{NN}(R) + \hat{V}_{ee}(\boldsymbol{r})\}\psi_k(\boldsymbol{r}; R) = \\ E_k(R)\psi_k(\boldsymbol{r}; R) \end{aligned} \tag{3.5}$$

as for the nuclear Schrödinger equation, considering the 3.4 and the expession for $\Psi(\boldsymbol{r}, \boldsymbol{R})$ in 3.3, we may rewrite the TISE as in 3.6

$$\begin{aligned} \{\hat{T}_N(\boldsymbol{R}) + \hat{T}_e(\boldsymbol{r}) + \hat{V}_{eN}(\boldsymbol{r}; R) + \hat{V}_{NN}(\boldsymbol{R}) + \hat{V}_{ee}(\boldsymbol{r})\}\psi_k(\boldsymbol{r}; R)\chi_N(\boldsymbol{R}) = \\ \{\hat{T}(\boldsymbol{R}) + E_k(R)\}\psi_k(\boldsymbol{r}; R)\chi_N(\boldsymbol{R}) = E\psi_k(\boldsymbol{r}; R)\chi_N(\boldsymbol{R}) \end{aligned} \tag{3.6}$$

Further developing the 3.6

$$\hat{T}(\boldsymbol{R})\psi_k(\boldsymbol{r}; R)\chi_N(\boldsymbol{R}) + E_k(R)\psi_k(\boldsymbol{r}; R)\chi_N(\boldsymbol{R}) = E\psi_k(\boldsymbol{r}; R)\chi_N(\boldsymbol{R}) \tag{3.7}$$

Given the nuclear kinetic energy as

$$\hat{T}(\boldsymbol{R}) = -\sum_{\alpha=1}^{N} \frac{\hbar}{2M_\alpha} {\nabla_\alpha}^2 \tag{3.8}$$

Where α corresponds to the N nuclei of the system, we will obtain, following the rules for the second derivatives, for $\hat{T}(\boldsymbol{R})\psi_k(\boldsymbol{r};R)\chi_N(\boldsymbol{R})$ the 3.9

$$\begin{aligned} \hat{T}(\boldsymbol{R})\psi_k(\boldsymbol{r};R)\chi_N(\boldsymbol{R}) = &-\sum_{\alpha=1}^{N} \frac{\hbar}{2M_\alpha} \psi_k(\boldsymbol{r};R){\nabla_\alpha}^2\chi_N(\boldsymbol{R}) \\ -\sum_{\alpha=1}^{N} \frac{\hbar}{M_\alpha} \nabla_\alpha\psi_k(\boldsymbol{r};R)\nabla_\alpha\chi_N(\boldsymbol{R}) &- \sum_{\alpha=1}^{N} \frac{\hbar}{2M_\alpha} {\nabla_\alpha}^2\psi_k(\boldsymbol{r};R)\chi_N(\boldsymbol{R}) \end{aligned} \tag{3.9}$$

In the BO approximation, the second and the third terms of the 3.9 are zero. This means it is assumed that the electronic wavefunction does not change with the position of the nuclei. As a matter of fact, the subindex "k" denotes the electronic level to which corresponds the electronic wavefunction. However, the number of electronic states is not unique for every molecule; in fact k can run from a discrete number to infinite. Linear combinations of these states can be also proposed as a solution of the TISE leading to *coupling* between states. In the *adiabatic approximation*, the electrons are constrained in one k electronic state, thus the problem of coupling between states is eliminated since all coupling terms go to zero. Born and Oppenheimer themselves asserted that electrons do not undergo transitions between states exclusively under *appropriate* conditions [92]. The resulting nuclear TISE is

$$\{\hat{T}(\boldsymbol{R}) + E_k(R)\}\chi_N(\boldsymbol{R}) = E\chi_N(\boldsymbol{R}) \tag{3.10}$$

Solving the electronic TISE for different nuclear geometries (3.5), we obtain a series of values for $E_k(R)$, from the perspective of the nuclear TISE (3.10), the $E_k(\boldsymbol{r})$ act as an electronic potential (referred hereafter as the Potential Energy Surface or PES) in which the nuclei move.

3.2.3 Relying on the unknown: Density Functional Theory

> *"What a strange and complicated beast a wave function is. [...]The wave function itself is essentially uninterpretable – it is an inscrutable oracle that returns valuably accurate answers when questioned by quantum mechanical operators, but it offers little by way of sparking intuition. One may be forgiven for stepping back from the towering edifice of molecular orbital theory and asking, shouldn't things be simpler?"*
>
> C.J. Cramer, "Essentials of Computational Chemistry: Theories and Models"

This was the first thing I read about *Density Functional Theory* (DFT). I always thought this was a good way to justify why we got to the Density Functional Theory.

Historically, till the rigorous foundation of DFT in 1964 [93], the solution to the electronic problem came from the so called *Molecular Orbital* (MO) family of *ab initio* approaches, directly founded on the *Hartree-Fock* (HF) theory. This family of methods still exists and has gone way over the HF approximation of an electron moving in the average field generated by all the other $n-1$ electrons in the system. As a matter of fact, if HF fails in calculating that relevant $\sim 1\%$ of the total energy of a system actually useful to describe all the interesting chemical reactions and physical-chemical properties (the *correlation energy*), modern approaches recover this component via perturbation theory (*e.g.* MP2) or variational schemes (*e.g.* CI). These methods (usually referred to as *post-HF*) can achieve accuracies that are directly comparable to experiments. Nevertheless, these wavefunction-based methods derive all the properties of a system from an approximated wavefunction depending on 4N coordinates (N=number of electrons in the system, 3 spatial coordinates and 1 spin coordinate for each electron), which translates in very demanding calculations, from where the fear of the "bestial" wavefunction. So again: "shouldn't things be simpler?" and in particular: can they ?

The existence of Density Functional Theory means a yes.

Density functional Theory Foundation

"Simpler" in DFT means determining the energy of a chemical system using a physical *observable*, *i.e.* the electron density. The first achievement towards

DFT formalism was the *uniform electron gas*, derived by Thomas and Fermi in 1927, where the kinetic energy is derived as a function of an homogeneous electronic density (*e.g.* from a metal). In spite of this, as mentioned, we will have to wait till 1964 for Hohenberg and Kohn to found the theory. The Hohenberg − Kohn theorems are, in fact, crucial to demonstrate that the electron density is the physical observable suitable to rescue us from the complexity of the MO approaches.

Hohenberg − Kohn Existence Theorem The first obstacle to get over was to demonstrate that the ground state density determines the external potential (in a molecule this is the electron attraction to the nuclei) and thus the Hamiltonian and the wave function. Hohenberg and Kohn proved that *the external potential* $v(\boldsymbol{r})$*, and hence the total energy, is a unique functional of the electron density* $\rho(\boldsymbol{r})$, through a *reductio ad absurdum*, as follows:

Given the exact non-degenerated ground state electron density of an N-electron system

$$\rho(\boldsymbol{r}) = N \int |\psi_0(r, r_2, ...r_N)|^2 d\boldsymbol{r}_2 ... d\boldsymbol{r}_N \tag{3.11}$$

This is associated with the corresponding non-degenerated ground state of N particles. If we assume that the same electron density can be obtained from two external potentials (v_1 and v_2) which differ for more than one constant, that means that we will have two many-electrons hamiltonians, $\hat{H}_1$ and $\hat{H}_2$, with two ground state wave functions, ψ_1 and ψ_2. $\hat{H}_1$ and $\hat{H}_2$ are defined as

$$\hat{H}_1 = \hat{T} + \hat{V} + v_1 \tag{3.12}$$

$$\hat{H}_2 = \hat{T} + \hat{V} + v_2 \tag{3.13}$$

The variational principle states that the expectation value of the Hamiltonian of the trial wave function $\tilde{\psi}$ (E) is always greater or equal to the ground state one (E_0) (eq 3.14).

$$\left\langle \tilde{\psi} \middle| \hat{H} \middle| \tilde{\psi} \right\rangle = E \geq E_0 = \langle \psi_0 | \hat{H} | \psi_0 \rangle \tag{3.14}$$

As a consequence, the expected value of the energy of ψ_1 with $\hat{H}_2$ is

$$E_2 \leq \langle\psi_1|\hat{H}_2|\psi_1\rangle = \langle\psi_1|\hat{H}_1|\psi_1\rangle + \langle\psi_1|\hat{H}_2 - \hat{H}_1|\psi_1\rangle \tag{3.15}$$

and similarly for ψ_2 with $\hat{H}_1$

$$E_1 \leq \langle\psi_2|\hat{H}_1|\psi_2\rangle = \langle\psi_2|\hat{H}_2|\psi_2\rangle + \langle\psi_2|\hat{H}_1 - \hat{H}_2|\psi_2\rangle \tag{3.16}$$

Substituting 3.12 and 3.13 in 3.15 and 3.16 and adding up 3.15 and 3.16, one obtains $E_1 + E_2 < E_2 + E_1$ which is an absurd. So the initial assumption is wrong. In other words, the ground state density uniquely determines one external potential, and then one hamiltonian and one wave function.

It is worth to point out that the theorem is valid for non-degenerated ground states N-representable and v-representable, hence if the density corresponds to the ground state of an external potential. This means that the electron density has to be a positive function defined in all space, and its integral has to be equal to the number of electrons in the system.

Hohenberg − Kohn Variational Theorem Hohenberg and Kohn also provided some indication on how to predict the density expression of a system, as they demonstrated that density obeys a variational principle as wave function within molecular orbital theory does, hence *the density that minimises the total energy is the exact ground state density.* The derivation starts from the previous theorem. In fact, if the exact density (ρ) and the trial density ($\tilde{\rho}$) univocally define their own external potential (respectively v and $\tilde{v}$), hamiltonian and wave function (respectively ψ_0 and $\tilde{\psi}$), then, calculating the energy with the exact Hamiltonian ($\hat{H}$) and according to the variational principle, we can write

$$E_0 = \langle\psi_0|\hat{H}|\psi_0\rangle \leq \left\langle\tilde{\psi}\middle|\hat{H}\middle|\tilde{\psi}\right\rangle = \tilde{E} \tag{3.17}$$

This tells us that the energy for a trial density ($\tilde{E}$) will be always greater than or equal to the exact ground state energy (E_0), so looking for a density that minimises the energy, we will find the exact density of the ground state.

According to what discussed so far, the great convenience of the Hohenberg − Kohn theorems consists in the fact that a problem of 4N variables as in wave function-based methods can be solved with a real-space function

with only 3 variables (x, y, z), independently of the number of the electrons of the system. The beast looks tamed, my friends, but is it really ?

What just argued would be true if we would know the relationship between energy and electron density. Considering that this is not the case, an alternative method to solve the electronic problem is needed.

The Kohn-Sham equations and the exchange-correlation problem
In 1965 Kohn and Sham [94] proposed to solve the DFT problem considering the kinetic energy of a fictitious system (denoted by s) of N non-interacting electrons, moving in an external potential, $v_s(\boldsymbol{r})$, with the same ground-state total density as the real, interacting system. The Hamiltonian for the non interacting system in the external potential $v_s(\boldsymbol{r})$ is given by the 3.18.

$$\hat{H_s} = \sum_{i=1}^{N} \hat{h_i}(\boldsymbol{r}) = \sum_{i=1}^{N} -\frac{1}{2}{\nabla_i}^2 + \sum_{i=1}^{N} v_s(\boldsymbol{r}) \tag{3.18}$$

Being $\hat{h_i}(\boldsymbol{r})$ the i^{th} one-electron operator.

The ground-state wave function associated to the non-interacting system is expressed as a single Slater determinant of the following form

$$|\psi_s(\boldsymbol{r})\rangle = \frac{1}{\sqrt{N!}}|\phi_1(\boldsymbol{r})\bar{\phi_1}(\boldsymbol{r})\phi_2(\boldsymbol{r})\bar{\phi_2}(\boldsymbol{r})...\phi_N(\boldsymbol{r})\bar{\phi_N}(\boldsymbol{r})| \tag{3.19}$$

Each $|\phi_i(\boldsymbol{r})\rangle$ represents a Kohn-Sham spin-orbital. As the molecular orbitals of the HF method, the $\phi_i(\boldsymbol{r})$ are the solutions to the single determinant problem :

$$\big[-\frac{1}{2}{\nabla_i}^2 + v_s(\boldsymbol{r})\big]\phi_i(\boldsymbol{r}) = \epsilon_i\phi_i(\boldsymbol{r}) \tag{3.20}$$

Please, note that with this approach the computational advantage gained in DFT formulation with respect to HF one is practically lost. For this system, the exact density and the kinetic energy are as in, respectively, 3.21 and 3.22

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N_{occ}} |\phi_i(\boldsymbol{r})|^2 \tag{3.21}$$

$$T_s[\rho(\boldsymbol{r})] = \sum_{i=1}^{N_{occ}} \langle\phi_i|-\frac{1}{2}{\nabla_i}^2|\phi_i\rangle \tag{3.22}$$

As a result, the total energy for this system is

$$E_v[\rho(\boldsymbol{r})] = \sum_{i=1}^{N_{occ}} \epsilon_i = T_s[\rho(\boldsymbol{r})] + \int \rho(\boldsymbol{r})v(\boldsymbol{r})d\boldsymbol{r} \tag{3.23}$$

Which is just the sum of the kinetic and potential energy for the system. In the non-interacting system the latter is basically the attraction between the electron density and the external potential of the nuclei, and will be more conveniently called $V_{ne}[\rho(\boldsymbol{r})]$.

If we now switch to the real, interacting system, we will need to introduce a few more ingredients to the energy formulation. Such that the equation for the real energy looks like the 3.24

$$E[\rho(\boldsymbol{r})] = T_s[\rho(\boldsymbol{r})] + V_{ne}[\rho(\boldsymbol{r})] + J_{ee}[\rho(\boldsymbol{r})] + E_{XC}[\rho(\boldsymbol{r})] \tag{3.24}$$

The new ingredients are the coulombic electron-electron repulsion, ($J_{ee}[\rho(\boldsymbol{r})]$), and $E_{XC}[\rho(\boldsymbol{r})]$, which is the *exhange-correlation functional*. Let us now give some meaning to this exchange-correlation function in the context where we are now. We may express the $E_{XC}[\rho(\boldsymbol{r})]$ as

$$\begin{aligned} E_{XC}[\rho(\boldsymbol{r})] = (T[\rho(\boldsymbol{r})] - T_s[\rho(\boldsymbol{r})]) + (V_{ee}[\rho(\boldsymbol{r})] - J_{ee}[\rho(\boldsymbol{r})]) = \\ = \Delta T[\rho(\boldsymbol{r})] + \Delta V_{ee}[\rho(\boldsymbol{r})] \end{aligned} \tag{3.25}$$

The $\Delta T[\rho(\boldsymbol{r})]$ term contains the difference between the kinetic energies of the real and the fictitious system; whereas the $\Delta V_{ee}[\rho(\boldsymbol{r})]$ adds up the difference between the quantum-mechanical electron-electron repulsion and the classical, coulombic one. Note that this exchange−correlation is a subtle feature, which cannot be explained in classical terms. An easy way to understand the concept might be to consider the energy of interelectronic repulsion for two electrons of wavefuntion ψ_a and ϕ_b with same spin. Skipping the mathematical derivation of this formula, we will have

$$V_{ab}[\rho(\boldsymbol{r})] = J_{ab}[\rho(\boldsymbol{r})] - K_{ab}[\rho(\boldsymbol{r})] \tag{3.26}$$

Here the two electrons occupy two different orbitals, a and b. In such a case, very common in atoms and molecules, the classical Coulomb repulsion between the electron clouds in orbitals a and b is reduced by $K_{ab}[\rho(\boldsymbol{r})]$. This is a consequence of the Pauli principle and reflects the reduced probability of finding two electrons of the same spin close to one another − a so-called "*Fermi hole*" is said to surround each electron [95]. So far, DFT formulation contains no approximations. The only requirement is to know E_{XC} as a function of ρ. Unfortunately, the exact form of E_{XC} term is not known and, even if the Hohenberg and Kohn theorems proved that a functional of the density must exist, it does not provide any guidance about its form. So, DFT is an exact theory, but the relevant Kohn−Sham equations must be solved approximately because a key operator has an unknown form. Thus, the E_{XC} term containing all unknown features of the density functional has become the target of a fervent interest in the chemistry community, which has given birth to different exchange-correlation functional models. Despite the necessity of approximating the functionals, DFT provides much more accurate results than HF at the same computational cost. On the other hand, it is worth to mention that in DFT there is no way to systematically improve the functional, hence the solutions cannot be improved by going to bigger basis sets or higher correlation levels, as it is the case for the wavefunction methods. Therefore, the success of the calculation dramatically depends on the choice of the functional.

Exchange-correlation Functionals Before discussing the nature of various functionals, we will express the functional dependence of E_{XC} on the electron density as an interaction between the electron density and an energy density ϵ_{XC} that is also dependent on the electron density

$$E_{XC}[\rho(\boldsymbol{r})] = \int \rho(\boldsymbol{r})\epsilon_{XC}[\rho(\boldsymbol{r})]d\boldsymbol{r} \tag{3.27}$$

The families of functionals not employed in this work will exclusively be mentioned. The simplest formulation of the energy density, ϵ_{XC}, comes from the *Local Density Approximation* (LDA) [96], where the electronic

density is considered locally as that of a uniform electron gas, *i.e.* is a slowly varying function. Clearly, improvements over the LDA approach must consider a non-uniform electron density. A step in this direction is to make the exchange and correlation energies dependent not only on the electron density, but also on derivatives of the density, which is to say its variation in space. In *Generalized Gradient Approximation* (GGA) methods [97], the corrected energy densities are constructed from the LDA ones, plus a correction term (3.28)

$$\epsilon_{XC}^{GGA}[\rho(\boldsymbol{r})] = \epsilon_{XC}^{LDA}[\rho(\boldsymbol{r})] + \Delta\epsilon_{XC}\left[\frac{|\nabla\rho(\boldsymbol{r})|}{\rho^{4/3}(\boldsymbol{r})}\right] \tag{3.28}$$

Note that the dependence of the correction term (in brackets) is on the dimensionless reduced gradient, not the absolute gradient and refers to the local density inhomogeneity. For instance, it is large for large values of the gradient density $\nabla\rho(\boldsymbol{r})$, such as on surfaces, and also for small values of the density $\rho(\boldsymbol{r})$, such as in areas that are far from nuclei. Nevertheless, GGA functionals present some limitations. On top of others, physical and chemical processes are controlled by energy differences (rather than absolute energies), if absolute energies are improved in GGA with respect to LDA, the relative ones are not really [98]. The best-known examples are atomic ionization potentials (IP)[98], electron affinities (EA) [98], molecular atomization energies [99–101] and the cohesive energies of solids [102, 103]. In these energy differences the major contribution to the total energies due to the core electrons usually cancels out (as the core electrons do not really participate in the process of interest). The percentage error in energy differences is therefore primarily determined by the description of the valence states, for which xc-effects play a much more prominent role than for the core electrons [98]. It is possible to analyze the accuracy of xc-functionals on a local level, for which the corresponding potential provides an unambiguous measure. Considering that for a few atoms the exact exchange and correlations potentials are known [104], when we compare the LDA and GGA performance, we can conclude that if the exchange potential has improved for GGA with respect to LDA functionals, there is no improvement in the description of the correlation potential.

Due to this and other limitations, *meta-GGA* were designed. Here the GGA set of density variables is extended with the Laplacian of the density, $\nabla^2\rho(\boldsymbol{r})$ [105, 106], which is the key quantity in the second order gradient correction for E_{XC}. Besides $\nabla^2\rho(\boldsymbol{r})$, also the kinetic energy density (t_s) [107, 108] has been used for the representation of E_{XC}, as it is the case for the functional used in this work, M06-2X [109].

Apart from these families, another family of functionals is worth to be mentioned: the *hybrid functionals* [110]. This class of functionals is based on the adiabatic connection, according to which the E_{XC} can be formulated as in the 3.29.

$$E_{XC}^{Hyb} = (1-a)E_{XC}^{DFT} + aE_{XC}^{HF} \tag{3.29}$$

Where we can see that E_{XC} can be expressed with a "mix" of HF exact exchange and E_{XC} from a suitable DFT functional. Directly from this consideration, the "multi-purpose" functional B3LYP [110, 111] (3.30) came to us

$$E_{XC}^{B3LYP} = (1-a)E_{XC}^{LSDA} + aE_{XC}^{HF} + b\Delta E_x^B + (1-c)E_c^{LSDA} + cE_c^{LYP} \tag{3.30}$$

In 3.30 a, b, and c are the same ones used in the B3PW91 functional [112, 113] (0.20, 0.72, and 0.81, respectively). LSDA stands for *Local Spin Density Approximation* [114]. The local spin density approximation is suited to treat the open shell systems or molecules next to dissociation. While in the LDA two paired electrons with opposite spin have the same spatial Kohn−Sham orbital $|\phi(\boldsymbol{r})\rangle$, in the LSDA they occupy different spatial orbitals: $|\phi_\alpha(\boldsymbol{r})\rangle$ and $|\phi_\beta(\boldsymbol{r})\rangle$ for spin up and spin down respectively. Thus α and β densities are separately minimized. LYP [111] is a GGA functional where the Laplacian of ρ can be rewritten as the sum of simple gradients of the density and the KS kinetic energy density (relying on the fact that normalizable KS states can always be chosen to be real). LYP then approximate the KS kinetic energy density by its second order gradient expansion (GE), so that the Laplacian of ρ is expressed as a density functional. This serendipitian functional became very popular for its "overall-good" performances in molecular approaches, and it has been used in this work in several contexts.

Expanding wavefunctions At this point we have found a way to mathematically express the total energy of the system, thus, we need to find a suitable way to describe the electron density too.

The total electronic density can be described as the sum of the square module of all the Khon-Sham molecular orbitals of the system.

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} |\phi_i(\boldsymbol{r})|^2 \tag{3.31}$$

At the same time, each of this molecular may be expanded as a linear combination of atomic orbitals (LCAO approach)

$$|\phi_i(\boldsymbol{r})\rangle = \sum_{n=1}^{m} c_{ni} \, |f_n(\boldsymbol{r})\rangle \tag{3.32}$$

The atomic orbitals may, in turn, be represented as a linear combination of basis functions. There are two types of basis functions commonly in use: the *Slater Type Orbitals* (STO) and the *Gaussian Type Orbitals* (GTO), also called *Gaussian primitives*. STO and GTOs have, respectively, the following analytical expression in polar coordinates

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{n-1}e^{-\zeta r} \tag{3.33}$$

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{2n-2-l}e^{-\zeta r^2} \tag{3.34}$$

N is a normalization constant and $Y_{l,m}$ is a spheric harmonic function. n, l and m are, respectively, the principal quantum number, the angular moment and its z component. GTO provides a worse description in representing the proper behaviour in the proximity of the nucleus and too far from it, as in the first case it has a zero slope and in the second it drops too rapidly, such that the "tail" of the wave function is poorly represented. To avoid this problem, a linear combination of several GTO (called *Contracted Gaussian Functions* or CGF) is used to describe an AO.

A collection of basis function is called *basis set*. A *minimum basis set* contains only one CGF per AO. In addition, just a number of AOs up to the to valence shell of a particular system are included. A more accurate description would include a higher number of basis functions, from there the

double zeta and *triple zeta* basis sets, which double and triple the number of CGT per AO. When the mentioned expansion is done on the valence shell only (the core AO are usually well described with a single CGT), then we obtain the *split-valence* family of basis functions.

A large number of basis sets is available nowadays. In this work, the split-valence basis sets by Pople et al., 6-31++G(d,p) and 6-311++G(d,p) [115, 116] have been especially used (see Chapters 8, 9, 10). Here, two different type of basis functions are used for the core and valence shells, such that we have more flexibility. In particular, a CGF of six primitive GTO describes core atomic orbitals, while valence orbitals split in two functions formed by 3 and 1 primitives in the 6-31++G(d,p) case; or three functions formed by 3 and 1 primitives in the 6-311++G(d,p) case. On top of this, diffuse functions are added to the heavy atoms and hydrogen atoms ("++"). Finally, polarization functions, *i.e.* functions with higher angular momentum, are also included: one p set on each hydrogen and one d set on each heavy atom.

The *Split Valence Polarized* (SVP) basis set has also been used. It is a [3s2p] contraction of a (7s4p) set of primitive functions.

Finally, the *correlation consistent* (cc) polarized Valence Double Zeta basis (CC-pVDZ) has been used in chapters 6 and 7. This basis set is geared towards recovering the correlation energy of the valence electrons. Functions that contribute similar amounts of correlation energy are included at the same stage, independently of the function type. Moreover, they are designed such that the energy error from the *sp*-basis should be comparable with (or at least not exceed) the correlation error arising from the incomplete polarization space, and the *sp*-basis therefore also increases as the polarization space is extended. Last but not least, the *s*- and *p*-basis set exponents are optimized at the HF level for the atoms, while the polarization exponents are optimized at the CISD level (configuration interaction limited to single and double excitations), and the primitive functions are contracted by a general contraction scheme using natural orbital coefficients.

Solutions to the Khon-Sham Equations We have now arrived to the point where we describe the routine to solve the relevant equations and how

we get our results. Here we will exclusively describe the diagonalization technique.

From the second HK theorem we can extract a condition for an electronic density to minimize the energy functional:

$$\partial E_v[\rho] = \partial T_s[\rho] + \int d\boldsymbol{r}\left(v(\boldsymbol{r}) + \frac{\partial V_{ee}[\rho]}{\partial \rho} + v_{XC}(\boldsymbol{r})\right)\partial\rho(\boldsymbol{r}) = 0 \tag{3.35}$$

Considering $\frac{\partial V_{ee}[\rho]}{\partial \rho}$ as equal to $v_J(\boldsymbol{r})$, the Hartree potential, and expressing the $\partial T_s[\rho]$ in terms of the external potential of the non-interacting system we will obtain

$$\partial E_v[\rho] = -\int d\boldsymbol{r} v_s(\boldsymbol{r})\partial\rho(\boldsymbol{r}) + \int d\boldsymbol{r}(v(\boldsymbol{r}) + v_J(\boldsymbol{r}) + v_{XC}(\boldsymbol{r}))\partial\rho(\boldsymbol{r}) = 0 \tag{3.36}$$

The condition for n(r) to be the minimum of Ev[n] is given by the 3.37, substituting $v_s(\boldsymbol{r}) = v_{KS}(\boldsymbol{r})$

$$v_{KS}(\boldsymbol{r}) = v(\boldsymbol{r}) + v_J(\boldsymbol{r}) + v_{XC}(\boldsymbol{r}) \tag{3.37}$$

where $v_{KS}(\boldsymbol{r})$ is the so-called Kohn-Sham potential or effective potential.

Given this, the Schrödinger equation for the set of non-interacting electrons is

$$-\frac{1}{2}\nabla_i^2 |\phi_n(\boldsymbol{r})\rangle + v_{KS} |\phi_n(\boldsymbol{r})\rangle = \epsilon_i |\phi_n(\boldsymbol{r})\rangle \tag{3.38}$$

and

$$\hat{H}^{KS} = -\frac{1}{2}\nabla_i^2 + v_{KS} \tag{3.39}$$

is the *Kohn−Sham operator*. i runs for electrons. This set of integro-differential equations is known as *Kohn−Sham equations*.

The self-consistent procedure used in the diagonalization method consists of:

1. Choosing a basis set, an initial molecular geometry and an exchange-correlation functional. At this pont a guess density $\tilde{\rho}$ is constructed and from this the initial $v_{KS}(\boldsymbol{r})$ and the 3.38.

2. The electronic orbitals are expanded into a basis set and the i 3.38 equations can be written as the matrix

$$\boldsymbol{H}^{KS}\boldsymbol{C} = \boldsymbol{S}\boldsymbol{C}\boldsymbol{\epsilon} \tag{3.40}$$

3. The diagonalization of the 3.40 gives a new ρ and a new minimal energy, so the new ρ is used as a new guess to repeat the scheme till in two consecutive SCF cycles the energy difference between the two cycles is lower than an established threshold.

As a final remark, it is useful to end this section underlining some well known limitations of the DFT:

- Excited states cannot be properly determined;
- Dispersion forces are not taken into account;
- Self-Interaction Error (SIE), in DFT it comes from the approximate forms of the exchange term in classical DFT functionals, which does not cancel exactly the self-interaction Coulomb term that includes the unphysical interaction of an electron with itself. For this reason, LDA/GGA functionals tend to overdelocalize electrons.

Chapter 4

From static(s) to dynamic(s)

We are now about to chart the seas of Molecular Dynamics (MD). This alternative approach to chemical and physical problems has found a lot of success among the scientific community nowadays. We shall consider the reasons of this success.

In order to do so, we might point out some of the intrinsic characteristics of single point optimisation that make it problematic. First of all, performing a successful single point calculation, no matter if the search is for a minimum or a maximum, very often requires providing to the program an initial guess, which is close enough to the critic point needed. Especially for reactions that are not dominated by a few (internal) reaction variables, it may be difficult to generate such a guess. In many cases, the convergence radius is small, *i.e.* the starting geometry must be (very) close to the saddle point in order to converge [114]. For Hessian-based methods the calculation of the second derivative methods might be computationally expensive, as the handling of the Hessian matrix itself for large systems may become problematic. Furthermore, systems with many soft vibrational modes are often problematic, as the resulting low Hessian eigenvalues interfere with the negative curvature along the reaction vector [114]. What's the advantage of the dynamics then ?

Basically, the smaller the molecule we are treating the more we can stick to the image of two structures (the reactant and the product) connected by a maximum (the transition state). Although for large complex systems this idyllic picture blurs. In chemistry, this blurring translates as a situation

where we have lots of local minima very very close in energy to the global minimum. We refer to this situation as a *combinatorial explosion*. In such a case, any experimentally observed property at room temperature will be a Boltzmann average over many individual conformations, and so would be the reaction rate, thus the reaction kinetics. Based on what we have already discussed, it seems that a systematic sampling of the potential energy surface urges, even though stronger tools are available these days to perform this sampling, such an operation might be problematic even for medium size systems. Not to mention that the PES sampling is always limited to an area of the PES, which should reasonably coincide with what we expect *a priori* from our system. Here is where *molecular dynamics* may be of great help.

By the name of molecular dynamics goes a class of methods that induces a perturbation into the initial geometry to make it pass to another local minimum, and so on for the new minimum. The perturbation is given by the total energy set for the system, which is divided into potential and kinetic energy, thus the molecules are enabled to overcome barriers not superior to the total energy minus the potential energy. As a consequence, a reasonable choice of the mentioned total energy is essential to the good performance of the dynamics. Finally, a dynamic simulation may be initiated in the transition state region and the trajectory followed in both directions, solving the enigma of choosing the interesting part of the PES. Nevertheless, this method does not guarantee to find the global minimum of the system, but, most probably, will find many local minima very close to the global minima. Thus, collecting a significative statistics on the system under analysis, one should be able, in principle, to unveil the mechanisms behind the observable behaviour of the mentioned system. It is a natural consequence of this brief discussion that such a tool can best serve the bio-world, where there are large molecules, but not only, as it results particularly useful for liquids, solids, nanoparticles and more.

MD methods may be divided in classical, *ab initio* and quantum ones. The classical ones are based on Newton equations of motion and model potentials. Here we will discuss the *Force Field* (FF) MD, in particular the reactive FFs, as one of the tools used in this work. Another important set of simulations is the *ab-initio* MD, for which quantum description is used for electrons while the nuclei are kept as classical particles. Finally, the

Quantum MD simulations aim to a more complex quantum description of electrons and nuclei. Figure 4.1 shows a graph where time scale is related to the length scale of the system, so that if the *ab initio* methods are located in the origin of the axes the other, less accurate methods follow. The higher the level of theory, the lower the treatable size of the system becomes.

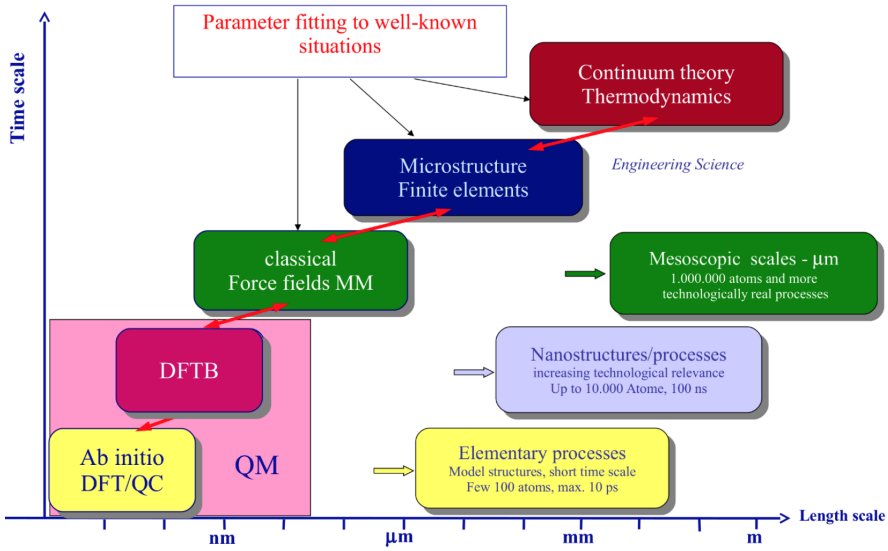


Figure 4.1: Time scale to length scale classification of the available computational methods for molecular dynamics.

4.1 Classical Molecular Dynamics

Considering Newton's equations of motion, it is possible to write the force f_i acting on a molecule i of mass m_i is given by

$$m_i \frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = \boldsymbol{f_i} \tag{4.1}$$

For the purpose of numerically solve this equation we would make a Taylor expansion out of the second order derivative of the 4.1

$$x(t+h) = x(t) + h\frac{\mathrm{d}x(t)}{\mathrm{d}t} + h^2\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} + h^3\frac{\mathrm{d}^3x(t)}{\mathrm{d}t^3} + \ldots \tag{4.2}$$

Equation 4.2 implies that x at time $(t + h)$ can be expressed as the sum of x itself, the first-order differential, the second-order differential, and so on, multiplied by a constant for each term. Same can be done for the time $(t - h)$. Truncating the expansion at the second order, summing up the two and solving the resulting equation, we get for the second derivative

$$\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} = \frac{x(t+h) - 2x(t) + x(t-h)}{h^2} \tag{4.3}$$

4.3 goes by *central difference approximation*, which rearranged gives

$$x_i(t+h) = 2x_i(t) + x_i(t-h) + \frac{h^2}{m_i} f_{xi}(t) \tag{4.4}$$

In a molecule with N atoms, 3N of these equations are solved at each step. This constitutes the base of the *Verlet method*, which is nowadays still pretty popular, and from the 4.3, the velocity can be extrapolated. Now that we have an idea of the equation of motion used, we need to move one step further and discuss the potential.

4.1.1 Tersoff potentials' big family

Tersoff potentials are part of the family of the FFMD methods, where the nuclei evolve according to an empirical potential designed as close as possible to the real system. More in general the force field potential can be expressed as

$$U(\boldsymbol{R}) = U_{bond}(\boldsymbol{R}) + U_{angle}(\boldsymbol{R}) + U_{dihedral}(\boldsymbol{R}) + U_{nonbond}(\boldsymbol{R}) \tag{4.5}$$

Where $U_{bond}(\boldsymbol{R})$ depends on the oscillations about the equilibrium bond length, $U_{angle}(\boldsymbol{R})$ on the oscillations of three atoms about an equilibrium angle, $U_{dihedral}(\boldsymbol{R})$ on the torsional rotation of 4 atoms around a central bond and $U_{nonbond}(\boldsymbol{R})$ on the so-called non-bonded energy terms, usually build up as a combination of an electrostatic term and a *Lennard-Jones* one.

Nevertheless one down-side of most of these potentials may be identified in the fact that they are intended for modeling physical processes, not chemical reactions. From here the necessity of some reactive FF potential that is able to empirically model changes in covalent bonding, such as

Tersoff-type potential [117]. In fact, here many-body terms reflecting the local coordination environment of each atom are used to modify the strength of more conventional pairwise terms. With this approach, individual atoms are not constrained to remain attached to specific neighbors, or to maintain a particular hybridization state or coordination number.

Tersoff potential original formulation

The Tersoff potential as it was originally formulated for the sole calibration of Si is expressed by the 4.6

$$E = \sum_i E_i = \frac{1}{2} \sum_{i \neq j} V_{ij} \tag{4.6}$$

Where E is the total energy of the system, E_i is the site energy for site i, whereas V_{ij} is the interaction energy between the atoms i and j (4.7) .

$$V_{ij} = f_c(r_{ij})[Ae^{(-\lambda_1 r_{ij})} - B_{ij}e^{(-\lambda_2 r_{ij})}] \tag{4.7}$$

From 4.7 Vij depends on the distance between the closer neighbour atoms i and j, r_{ij}, through a cutoff function, $f_c(r_{ij})$, that restricts the range of the potential. The first term of the 4.7 is repulsive and the second term is interpreted as representing bonding. In this optic, the B_{ij} includes the bond order and the dependency upon the local environment. In this formulation all deviations from a simple pair potential are ascribed to the dependence of B_{ij} upon the local environment. So B_{ij} has the following expression

$$B_{ij} = B_0 e^{(-z_{ij}/b)} \tag{4.8}$$

Here z_{ij} is a weighted measure of the number of bonds competing with the bond ij, and b determines how rapidly the bond strength falls off with increasing effective coordination. More specifically z_{ij} is better expressed as

$$z_{ij} = \sum_{k \neq i,j} \left[\frac{w(r_{ik})}{w(r_{ij})}\right]^n [c + e^{(-dcos\theta_{ijk})}]^{-1} \tag{4.9}$$

Considering that w is the "bare" bonding potential and k a non-neighbour atom to i and j, the first term of the 4.9 determines the weight of the closer neighbours with respect to the others in the interaction potential; whereas

the second term gives the dependance on the bond angle. Inspecting the 4.8 and 4.9, one might spot that such formulation lacks the consideration of dispersion and non-bonded repulsion, which makes the potential pretty poor to describe systems with significant intermolecular interactions.

Clearly, since this first formulation many attempts have been done to try to include these ingredients in the Tersoff formulation and yet maintaining the reactive connotation of the model. Among them, we will discuss one that has been used in this work, the *adaptive intermolecular reactive empirical bond-order* (AIREBO) potential [118]. The AIREBO potential is an improved version of the REBO potential [119], basically a Brenner formulation of the original Tersoff potential, still lacking the already mentioned ingredients.

What's special about AIREBO force field ?

In the AIREBO potential the dispersion and intermolecular repulsion interactions are modeled with a $Lennard - Jones$ (LJ) 2-6 potential. The insertion of this potential is a menace to the survival of the reactive aspect of the potential, as two atoms should overcome the barrier imposed by this non-bonding interactions in order to start interact. The LJ barrier is known to be hard, hence it is necessary to diminish in certain circumstances the mentioned wall. This results in the choice of three criteria to establish whether and at what distance to switch off the LJ interaction:

- The distance separating a given couple of atoms;
- The strength of any bonding interaction between them;
- The network of bonds connecting them.

and in the following mathematical expression for the LJ interaction between i and j

$$E_{ij}^{LJ} = S(t_r(r_{ij}))S(t_b(b_{ij}))C_{ij}V_{ij}^{LJ}(r_{ij}) + [1 - S(t_r(r_{ij}))]C_{ij}V_{ij}^{LJ}(r_{ij}) = \\ = A + B \tag{4.10}$$

In the 4.10 three different functions are used to smoothly switch off the LJ potential barrier according to the mentioned three criteria. The first one

takes into account the distance r_{ij} through a universal switching function $S(t)$, which value is unity when $t < 0$ and zero when $t > 0$. Please, note that here t_r and t_b do not refer to time, but to a scaling function, which depends on the distance between i and j according to the 4.11

$$t_r(r_{ij}) = \frac{r_{ij} - rij^{LJmin}}{rij^{LJmax} - rij^{LJmin}} \tag{4.11}$$

where rij^{LJmin} and rij^{LJmax} represent the distance at which the LJ interactions are minimal and maximal, respectively. In this way when $r_{ij} > r_{ij}^{LJmax}$, $S(t_r(r_{ij}))$ will be zero and the pair distance has no impact on the LJ interaction. On the other hand, when $r_{ij} < r_{ij}^{LJmax}$, then $S(t_r(r_{ij})) \neq 0$ and the LJ interactions are contingent on the values of the other two switching functions. The switching region, $[r_{ij}^{LJmin}, r_{ij}^{LJmax}]$, was chosen such that the minimum of the LJ potential well remains unperturbed from the unswitched interaction, and such that there is no artificial reaction barrier due to LJ repulsions at short distances when the LJ interactions are switched off.

The bonding switch is given by the $S(t_b(b_{ij}))$ term in the equation 4.11. In this case the scaling function has the form

$$t_b(b_{ij}) = \frac{b_{ij} - bij^{min}}{bij^{max} - bij^{min}} \tag{4.12}$$

Where b_{ij} is the bond order between atom i and atom j. When the bond order is large, covalent bonding will occur between the two given atoms, so $t_b > 1$, $S(t_b(b_{ij})) = 0$, as the repulsive LJ interactions will disappear ($A = 0$ in 4.10). Whereas for low values of the bond order, when covalent bonding is not likely to happen, $t_b < 1$ or < 0, so that $S(t_b(b_{ij})) \neq 0$ and the repulsive LJ will be included at a variable degree depending on b_{ij} and then t_b. If b_{ij} is low enough, $S(t_b(b_{ij})) = 0$ and the LJ normal barrier will be intact. The bond order considered in the intermolecular interaction part of the AIREBO potential differs from the Tersoff/REBO one, as it does not refer to the intramolecular distance, but the intermolecular one.

The connectivity switch, C_{ij}, introduces the mathematical dependence on the third criterion. C_{ij} is defined as

$$C_{ij} = 1 - max\{w_{ij}(r_{ij}), w_{ik}(r_{ik})w_{kj}(r_{kj}), \forall k \\ w_{ik}(r_{ik})w_{kl}(r_{kl})w_{kl}(r_{kl})w_{lj}(r_{lj})w_{lj}(r_{lj}), \forall k, l\} \tag{4.13}$$

where w is a bond-weight term which is used to implement a smooth transition from bonded to non-bonded interactions as a molecule dissociates. Mathematically w is a bond-based switching function as the ones already analized that varies smoothly from one to zero over a covalent bonding region. Thus, the 4.13 is designed such that when i and j are neighbours, or are connected by one or two intermediate neighbours, Cij is zero and there is no LJ interaction between the atoms.

The other addition of AIREBO potentials with respect to the REBO potential is a term dependent on the dihedral angles. As the original potential was designed for solid networks, this absence did not negatively affect the model, although when treating hydrocarbons bigger than methane this dependence should be taken into account. The torsional potential is usually a cosine power series of the dihedral angle ω. Such that over the 360° range, some potential barriers exist (Figure 4.2), for our potential mentioned barriers must change as the molecule undergoes a chemical reaction. Therefore, it is preferable if the symmetry of the torsional potential arises naturally from the environment rather than being built into the parameter set.

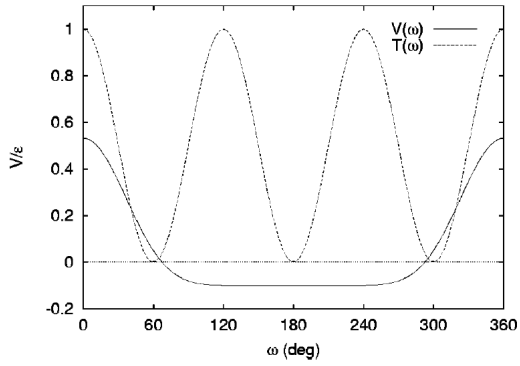


Figure 4.2: Conventional (T(ω)) and AIREBO (V(ω)) torsional potentials in comparison from [118].

This is accomplished in AIREBO using a torsional potential with a single minimum 4.14, where the constants are optimized for sp^3 carbons. When implementing the potential for all the dihedral angles in the system, we get the complete AIREBO torsional potential 4.15

$$V(\omega) = \epsilon\left[\frac{256}{405}cos^{10}\left(\frac{\omega}{2}\right) - \frac{1}{10}\right] \tag{4.14}$$

$$E_{tors}(\omega) = \frac{1}{2}\sum_i \sum_{j\neq i} \sum_i \sum_{k\neq i,j} \sum_{l\neq i,j,k} w_{ij}(r_{ij}) w_{jk}(r_{jk}) w_{kl}(r_{kl}) V^{tors}(\omega_{ijkl}) \tag{4.15}$$

The use of bond weights ensures that the torsional energy associated with a given dihedral angle will be removed smoothly as any of the constituents bonds are broken.

Finally, given the difference in energy scales between LJ and covalent bonding interactions (10^{-2} kcal/mol and 10^2 kcal/mol), the LJ contributions can be parametrized independently from the other existing REBO components.

4.2 Ziegler-Biersack-Littmark Potential

When describing collisions between projectile and target, as to include the nuclear scattering component, we have been using *Ziegler-Biersack-Littmark* (ZBL) potential [2]. The electronic clouds of the colliding particles at very short distance behave as point charges repelling each other. Although when the distance between the two increases, the screening by the electrons decreases the effective strength of the scattering potential. From this, it follows that we can write the interaction potential between two colliding particle as a Coulomb screened potential of general form

$$V(r) = \frac{Z_1 Z_2}{r} f(x) \tag{4.16}$$

where Z_1 and Z_2 are the nuclear charges of the species involved in the impact, and $f(x)$ is the *screening function*. In the particular case of the ZBL potential the screening function has the following mathematical expression

$$f_{ZBL}(x) = 0.1818e^{-3.2x} + 0.5099e^{-0.9423x} + 0.2802e^{-0.40299x} + \\ +0.02817e^{-0.2016x} \tag{4.17}$$

4.3 Modeling collisions with atom/ion projectiles

In our collision simulation, the interaction projectile-target is described by the ZBL potential; whereas the processes happening inside the cluster are modeled with AIREBO if the target is a molecule or a cluster made of exclusively C and H atoms. REBO potentials are used if the target contains atomic species other than just C and H atoms. In order to set up the collisions, it is necessary to choose a virtual box where the simulation will take place, this box shall be big enough to contain the target. Once the x, y and z dimensions of the box are established, with the purpose of collecting statistically meaningful data we need to randomize the impact parameter. This is accomplished, firstly, positioning the cluster/molecule center of mass at the origin of a three-dimensional coordinate system and randomly rotating the target around this point. Then the projectile atom is fired towards the cluster, from a position where the x and y values are randomly assigned and the y is fixed (Figure 4.3).

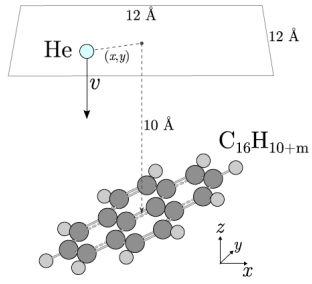


Figure 4.3: Scheme of the collision simulation with classical MD as developed for PAHs [120].

All the calculations involving Tersoff and AIREBO potentials were performed with the LAMMPS package [121].

Chapter 5

We don't do (just) classics

The main drawback of the use of the AIREBO potential comes from the non-consideration of the charges in the systems under investigation and then the impossibility to compute them during the calculations in an explicit way. Such a situation implies that we need to move to more advanced methods to consider the charge. At this point of our discussion, we shall introduce another method that has been used in this work: the *density functional tight-binding* (DFTB) method [122].

Born in the Slater-Koster original formulation for treating periodic solids [123], nowadays the standard tight-binding (TB) scheme works by expanding eigenstates of a Hamiltonian in an orthogonalized basis of atomic-like orbitals and representing the exact many-body Hamilton operator with a parametrized Hamiltonian matrix, where the matrix elements are fitted to the band structure of a suitable reference system [124] and it applies to a larger variety of systems, including: discrete molecules, thanks to Froyen and Harrison's inclusion of the interatomic distance dependence of the matrix elements [125, 126]; and surfaces due to the work of Chadi et al [127]. For the way it is built, the TB limits the transferability of a given set of parameters, hence the need of improved schemes based on density functional theory [128]. In order to understand where we are now, it is necessary to clarify that this scheme sits between classical methods and density functional theory methods (Figure 4.1). According to Figure 4.1, the DFTB methods are able to treat systems up to 10000 atoms with a simulation time of up to 100 ns.

Even if the reactive force field potentials (*i.e.* Tersoff and AIREBO in the case of this work) have introduced a way to take into account the structure and the bond pattern changes happening in the course of the simulation (4), they still suffer the serious drawback of being based on "predefined potentials". Thus, when the variety of the atoms and molecules increases in a system, myriad of different interatomic interactions have to be parametrized. From this and for the sake of higher accuracy, the need of a more solid method to compute the forces acting on the nuclei, as the *ab initio* molecular dynamics methods. So, because some of our systems were small enough to be treated with an higher level of theory, we have been exploring the possibilities of the *Born-Oppenheimer Molecular Dynamics* (BOMD) [129, 130] and the *Atom Centered Density Matrix Propagation (ADMP)* [131–133].

5.1 Pills of DFTB Formalism

In order to build the DFTB formalism, we need to write the total energy as a Taylor expansion of the KS density functional total energy (ignoring all the terms after the second order one) around a properly chosen reference density, ρ_0

$$E_{tot}[\rho_0, \delta\rho] = E_{bs}[\rho_0] + E_{rep}[\rho_0] + E^2[\rho_0, \delta\rho^2] \tag{5.1}$$

ρ_0 corresponds to the sum of compressed atomic densities, meaning that it corresponds to a superposition of KS atomic orbitals, where the basis is restricted to the sole valence shells of the atoms. Such restriction requires an orthogonalization to the core orbitals of the atoms. In this context, the total density of a chemical system, ρ, is the reference density perturbed by some density fluctuation, $\delta\rho$. In the 5.1 the E_{bs} is the sum over the occupied orbital energies derived from the diagonalization of the electronic Hamiltonian, E_{rep} is a short-range repulsive two-particle interaction and E^2 is a second order expansion of the energy in the fluctuation $\delta\rho$. The 5.1 is a general expression for the so-called DFTB2 or *self consistent charge DFTB* (SCC-DFTB), which differentiates from the original one (DFTB1) for including a self-consistent scheme and then allowing charge transfer to happen.

Having the orthogonalized AOs, it is possible to write a general eigenvalue problem of the form

$$\sum_{\nu} c_{\nu i}(H^0_{\mu v} - \epsilon_i S_{\mu\nu}) = 0 \tag{5.2}$$

with the Hamiltonian matrix elements $H^0_{\mu v}$ and the overlap matrix elements $S_{\mu\nu}$ and where ν belongs to atom b and μ to atom a.

The effective KS potential V_{eff} may be written also as a superposition of atomic-like potentials, $V_{eff} \sim V^c_{eff}$. To this effective potential a core correction is added such that it would become a *pseudo-potential*.

Neglecting the three-centre terms and pseudo-potential contributions leads to the two-centre approximation for the Hamilton matrix elements

$$H^0_{\mu v} = (\phi_\mu| -\frac{1}{2}\nabla^2 + V^a_{eff} + V^b_{eff}|\phi_\nu) \tag{5.3}$$

where $|\phi_\nu)$ represents this valence-shell orthogonalized atomic orbital. The approximations performed lead to the same structure of the secular equations as in (non-orthogonal) TB schemes or the extended Hückel method; however all matrix elements are calculated within DFT, because the AO basis set ϕ_μ is computed by solving the DFT-KS equations for atoms [122].

Now, with AO basis and initial density determined, the KS equations can be solved leading to the first Taylor term expansion for the total energy (5.4)

$$E^1 = \sum_i n_i \epsilon_i = E_{bs}[\rho_0] \tag{5.4}$$

The second order correction, E^2, is obtained considering the density fluctuation, $\delta\rho$, as a sum of atomic density fluctuations, $\delta\rho_a$. Once here, each of these atomic density fluctuations is evaluated assuming an exponential decay for the charge density. Hence, only the Hartree integral is analytically evaluated, leading to a γ_{ab} function and E^2 in 5.1 reads as

$$E^2(\tau_a, \tau_b, R_{ab}) = \frac{1}{2}\sum_{ab, a\neq b} \Delta q_a \Delta q_b \gamma_{ab}(\tau_a, \tau_b, R_{ab}) = E_2[\rho_0, \delta\rho^2] \tag{5.5}$$

The Hartree term therefore describes the interaction of the charge density fluctuations $\delta\rho_a$ and $\delta\rho_b$, which reduces to a Coulomb interaction of partial charges Δq_a and Δq_b for large distances, *i.e.* γ_{ab} approaches $\frac{1}{R_{ab}}$ for large distances.

For short distances, one has to deal with the electron-electron interaction within one atom, and here the exchange-correlation contributions become important. In this case, E^2 describes the electron-electron interaction within one atom, and it becomes proportional to Δq_a^2, so γ_{ab} describes the effective on-site electron-electron interaction. The exponent τ_a is chosen such that the on-site value of the γ-function properly describes the atomic chemical hardness (or alternatively the Hubbard parameter as calculated from DFT, the second derivative of the energy with respect to the charge density) and thus implicitly takes into account the exchange-correlation contribution to the second-order term. Then, the Hartree energy can be evaluated analytically, as an effective interaction between atomic charges, computed using the Mulliken scheme. Finally for $R_{ab} = 0$, the function γ_{aa} is simply the Hubbard parameter.

As a last step, the repulsive term in 5.1, $E_{rep}[\rho_0]$, is obtained as

$$E_{rep} = E_{tot} - (E^1 + E^2) \tag{5.6}$$

Even if the second order formalism is usually 5-10 times slower than the first order one, the SCC method is necessary for systems exhibiting a delicate charge balance. In addition to this, a DFTB3 formalism exists. The third-order terms within SCC-DFTB can be rather seen as a robust way to introduce the charge dependence capturing some deficiencies of problematic approximations within the second-order formalism, namely, the small size of the pseudo-AO basis as well as the very simplified density fluctuation scheme.

5.2 Fundamentals of *ab initio* Molecular Dynamics

The use of *ab initio* molecular dynamics has become more and more popular in recent years. Figure 5.1 shows the exponential increase in the number of

papers published containing the phrase "*ab initio* molecular dynamics" and "Born-Oppenheimer molecular dynamics".

Table 1

	1970-1980	1980-1990	1990-2000	2000-2010	2010-2018
AIMD	57	185	2790	12100	18900
BOMD	3	10	63	883	2620

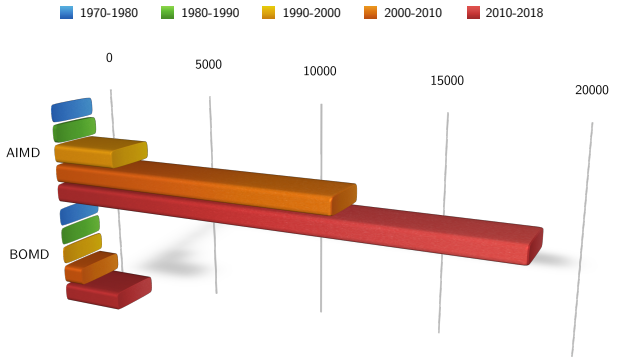


Figure 5.1: Statistics as obtained from Google Scholar on the number of papers containing the phrase "*ab initio* molecular dynamics" and "Born-Oppenheimer molecular dynamics" published in the decades: 1970-1980 (blue bar), 1980-1990 (green bar), 1990-2000 (yellow bar), 2000-2010 (orange bar). In addition, also the papers published in the interval 2010-2018 (red bar) were included.

Despite the already mentioned advance it constitutes in the computational chemistry field and its popularity, one may easily imagine that a cost has to be paid to use *ab initio* molecular dynamics. In fact, for example, the correlation lengths and relaxation times that are accessible are much smaller than what is affordable via standard molecular dynamics. It is worth noticing a not very well recognised feature of classical MD, someone calls "the experimental aspect of playing with the potential" [134]. The absence of the mentioned "experimental" connotation of the potential in *ab initio* molecular dynamics, on one side, makes less intuitive and straightforward tracing back the properties of a given system to a simple physical picture or mechanism. Nevertheless, on the other side, new phenomena, which were not foreseen before starting the simulation, can simply happen if necessary. This gives *ab initio* molecular dynamics a truly *predictive power*.

Ab initio molecular dynamics can also be viewed from the field of classical trajectory calculations [135–137]. This approach, as it was designed in gas

phase molecular dynamics, consisted in, first of all, building up a *global* potential energy surface from electronic structure calculations. The second step reconstructs the dynamical evolution of the nuclei on the previously builded PES. The nuclear dynamics is generated by means of classical mechanics, quantum mechanics or semi/quasi-classical approximations of various sorts. If classical dynamics is employed, it is natural to consider that the PES construction is the limiting step of the dynamics. As a matter of fact, there are $3N-6$ internal degrees of freedom that span the global potential energy surface of an unconstrained N-body system. If, for simplicity, we decide to sample just 10 discretization points per coordinate, this implies that the electronic structure calculations needed in order to map such a global potential energy surface are, roughly, of the order of 10^{3N-6}. Therefore, the computational workload for the first step grows roughly like $\sim 10^N$ with increasing system size, making very difficult to treat even relatively small systems.

Let us consider the requirements of *ab initio* molecular dynamics instead. A useful trajectory consists of about 10^M molecular dynamics steps, *i.e.* 10^M electronic structure calculations are needed to generate one trajectory. On a second level, it is assumed that 10^n independent trajectories need to be collected in order to average over different initial conditions (such that the statistics may be significative), so 10^{M+n} *ab initio* molecular dynamics steps are required in total. The final assumption is that each single-point electronic structure calculation needed to devise the global potential energy surface and one *ab initio* molecular dynamics time step requires roughly the same amount of cpu time. Taking into account this truly simplistic order of magnitude estimate, the advantage of *ab initio* molecular dynamics vs. calculations relying on the computation of a global potential energy surface amounts to about $10^{3N-6-M-n}$. The crucial difference with respect to global PES sampling methods is that, for a given statistical accuracy (that is for M and n fixed and independent on N) and for a given electronic structure method, the computational advantage of "on-the-fly" approaches grows like $\sim 10^N$ with system size.

Even though, because of the already mentioned growing popularity of these methods, considerable progress has been achieved in trajectory calculations (by carefully selecting the discretization points and reducing

their number, choosing sophisticated representations and internal coordinates, exploiting symmetry etc.), the scaling of $\sim 10^N$ with the number of nuclei remains a problem. Alternative strategies to face this issue consist, for instance, in reducing the number of active degrees of freedom by constraining certain internal coordinates, representing less important ones by a (harmonic) bath or friction, or building up the global potential energy surface in terms of few-body fragments. All these approaches, however, invoke approximations beyond the ones of the electronic structure method itself.

As a final remark, it is worth to notice that the computational advantage of the "on-the-fly" approaches diminish as more and more trajectories are needed for a given (small) system. For instance, extensive averaging over many different initial conditions is required in order to calculate quantitatively scattering or reactive cross sections. Summarizing this discussion, it can be concluded that *ab initio* molecular dynamics, rather than a static electronic structure PES sampling, is the method of choice to investigate the chemical reaction pathways of medium-sized and "chemically complex" systems.

5.2.1 Deriving Nuclear Classical Dynamics

We will now derive the nuclear classical molecular dynamics starting from the non-relativistic time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t}(\{\boldsymbol{r_i}\},\{\boldsymbol{R_I}\};t) = \hat{H}\Psi(\{\boldsymbol{r_i}\},\{\boldsymbol{R_I}\};t) \tag{5.7}$$

The standard Hamiltonian is given as

$$\hat{H} = -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \sum_{i<j}\frac{e^2}{|\{\boldsymbol{r_i}\}-\boldsymbol{r_j}|} - \sum_{I,i}\frac{e^2 Z_I}{|\{\boldsymbol{R_I}\}-\{\boldsymbol{r_i}\}|} + \\ + \sum_{I<J}\frac{e^2 Z_I Z_J}{|\{\boldsymbol{R_I}\}-\boldsymbol{R_J}|} \tag{5.8}$$

Where the first and second terms are the electron and nuclear kinetic energies. Electron-electron, nuclear-electron and nuclear-nuclear Coulombic interactions follow. We are now going to follow the Tully's demonstration

[136, 137]. At this point we need to separate the nuclear and the electronic parts of the wavefunction as follows

$$\Psi(\{\boldsymbol{r_i}\},\{\boldsymbol{R_I}\};t) \approx \psi(\{\boldsymbol{r_i}\};t)\chi(\{\boldsymbol{R_I}\};t)exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt'\tilde{E}_e(t')\right] \tag{5.9}$$

Here the nuclear and electronic wavefunctions are separately normalized to unity at every instant of time. Please, notice that equation 5.9, which is a product *ansantz* differs from the one used in BO approximation by a convenient face factor $\tilde{E}_e$, which may be expressed as

$$\tilde{E}_e(t') = \int d\boldsymbol{r}\boldsymbol{R}\psi*(\{\boldsymbol{r_i}\};t)\chi*(\{\boldsymbol{R_I}\};t)\hat{H}_e\psi(\{\boldsymbol{r_i}\};t)\chi(\{\boldsymbol{R_I}\};t) \tag{5.10}$$

It is worth to mention that this approximation is called a *one-determinant* or *single-configuration ansatz* for the total wavefunction, which at the end must lead to a mean-field description of the coupled dynamics.

Inserting the separation ansatz (equation 5.9) into equations 5.7 and 5.8 gives (after multiplying from the left by $\langle\psi|$ and $\langle\chi|$ and imposing energy conservation $d < \hat{H} > /dt = 0$) the following mathematical relations

$$i\hbar\frac{\partial\psi}{\partial t} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \left[\int d\boldsymbol{R}\chi*(\{\boldsymbol{R_I}\};t)V_{n-e}(\{\boldsymbol{r_i}\},\{\boldsymbol{R_I}\})\chi(\{\boldsymbol{R_I}\};t)\right]\psi \tag{5.11}$$

$$i\hbar\frac{\partial\chi}{\partial t} = -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + \left[\int d\boldsymbol{r}\psi*(\{\boldsymbol{r_i}\};t)\hat{H}_e\psi(\{\boldsymbol{r_i}\};t)\right]\chi \tag{5.12}$$

These two coupled equations define the basis of the Dirac's *time-dependent self-consistent field* (TDSCF) method [162]. Both electrons and nuclei move quantum-mechanically in time-dependent effective potentials (or self-consistently obtained average fields) obtained from appropriate averages (quantum mechanical expectation values $< ... >$) over the other class of degrees of freedom (by using the nuclear and electronic wavefunctions,

respectively). The product is then a *mean-field* description of the coupled nuclear-electronic quantum dynamics.

We need now to approximate the nuclei as classical point particles. To this end, we may rewrite the nuclear wavefunction as

$$\chi(\{\boldsymbol{R_I}\};t) = A(\{\boldsymbol{R_I}\};t)exp\big[iS(\{\boldsymbol{R_I}\};t)/\hbar\big] \tag{5.13}$$

where A is an amplitude factor and S is a phase, both considered to be real and $A > 0$. After transforming the nuclear wavefunction in Eq. (7) accordingly and after separating the real and imaginary parts, the TDSCF equation for the nuclei become

$$\frac{\partial S}{\partial t} + \sum_I \frac{1}{2M_I}(\nabla_I S)^2 + \int d\boldsymbol{r}\psi * \hat{H}_e\psi = \hbar^2 \sum_I \frac{1}{2M_I}\frac{(\nabla_I^2 A)}{A} \tag{5.14}$$

$$\frac{\partial A}{\partial t} + \sum_I \frac{1}{M_I}(\nabla_I A)(\nabla_I S) + \sum_I \frac{1}{2M_I}A(\nabla_I^2 S) = 0 \tag{5.15}$$

as re-expressed in terms of the new variables A and S. This is known as the *quantum fluid dynamical representation*. The relation for A (equation 5.15) can be rewritten as a continuity equation [163]. This continuity equation is independent of and ensures locally the conservation of the particle probability $|\chi|^2$ associated to the nuclei in the presence of a flux.

More important for the purpose of the present discussion is the relation for S (equation 5.14). This equation contains one term that depends on $\hbar$, a contribution that vanishes in the classical limit ($\hbar \to 0$). The resulting equation

$$\frac{\partial S}{\partial t} + \sum_I \frac{1}{2M_I}(\nabla_I S)^2 + \int d\boldsymbol{r}\psi * \hat{H}_e\psi = 0 \tag{5.16}$$

is isomorphic with the equations of motion in the Hamilton−Jacobi formulation [138]

$$\frac{\partial S}{\partial t} + \hat{H}(\{\boldsymbol{R_I}\}, \nabla_I S)^2) = 0 \tag{5.17}$$

of classical mechanics with the classical Hamilton function

$$\hat{H}(\{\boldsymbol{R_I}\},\{\boldsymbol{P_I}\} = T(\{\boldsymbol{P_I}\}) + V(\{\boldsymbol{R_I}\}) \tag{5.18}$$

defined in terms of (generalized) coordinates $\{\boldsymbol{R_I}\}$ and their conjugate momenta $\{\boldsymbol{P_I}\}$. With the help of the connecting transformation

$$\boldsymbol{P_I} \equiv \nabla_I S \tag{5.19}$$

the Newtonian equation of motion reads as

$$M_I \ddot{\boldsymbol{R}}_{\boldsymbol{I}}(t) = -\nabla_I \int d\boldsymbol{r}\psi * \hat{H}_e\psi = -\nabla_I V_e^E(\{\boldsymbol{R_I}\}) \tag{5.20}$$

The consequence of this is that the nuclei move according to classical mechanics in an effective potential V_e^E due to the electrons. This potential is a function of only the nuclear positions at time t as a result of averaging $\hat{H}_e$ over the electronic degrees of freedom, *i.e.* computing its quantum expectation value $\langle\psi|\hat{H}_e|\psi\rangle$, while keeping the nuclear positions fixed at their instantaneous values $\{\boldsymbol{R_I}\}$.

However, the nuclear wavefunction still occurs in the TDSCF equation for the electronic degrees of freedom and has to be replaced by the positions of the nuclei for consistency. In this case the classical reduction can be achieved simply by replacing the nuclear density $|\chi|^2$ in equation 5.12 the limit $\hbar \to 0$ by a product of delta functions $\prod_I \delta(\boldsymbol{R_I} - \boldsymbol{R_I}(t))$ centered at the instantaneous positions $\{\boldsymbol{R_I}\}$ of the classical nuclei as given by equation 5.20.

5.2.2 Born-Oppenheimer and Atom-centered Density Matrix Propagation Molecular Dynamics

One of the approaches to include the electronic structure in molecular dynamics simulations consists in straightforwardly solving the static electronic structure problem in each molecular dynamics step given the set of fixed nuclear positions at that instance of time [129, 130]. *Ab initio* may be carried out at both ground and excited states BO potential energy surface. The BOMD dynamics in this thesis were performed with turbomole [139] package. As already mentioned, being an *ab initio* method, the nuclei propagation is performed through classical molecular dynamics, hence here the time

dependence is a consequence of nuclear motion and it is not intrinsic to the dynamics. The resulting Born-Oppenheimer molecular dynamics method is defined by

$$M_I \ddot{\boldsymbol{R}}_{\boldsymbol{I}}(t) = -\nabla_I \min \psi_0 \langle \Psi_0 | H_e | \Psi_0 \rangle \tag{5.21}$$

$$E_0 \Psi_0 = H_e \Psi_0 \tag{5.22}$$

for the electronic ground state. The minimum of $\langle H_e \rangle$ has to be reached at each step of the dynamics, differently from what happens with ADMP [131, 132]. The biggest advantages coming from the use of ADMP lagrangian have to be identified in the amount of time saved for performing the dynamics with respect to the BOMD one, on one side; and, on the other side, the relatively small number of atom centered functions, compared to the plane waves' one (Car-Parrinello molecular dynamics, CPMD), needed to describe a given molecular system with the same accuracy. Given the density matrix $\mathbf{P}$, the propagation of the nuclei and the density in the orthonormal basis is

$$M\ddot{\boldsymbol{R}}(t) = -\frac{\partial E(\boldsymbol{R}, \boldsymbol{P})}{\partial \boldsymbol{P}}\bigg|_{\boldsymbol{P}} \tag{5.23}$$

$$\mu\ddot{\boldsymbol{P}}(t) = -\left[\frac{\partial E(\boldsymbol{R}, \boldsymbol{P})}{\partial \boldsymbol{R}}\bigg|_{\boldsymbol{P}} + \Lambda\boldsymbol{P} + \boldsymbol{P}\Lambda - \Lambda\right] \tag{5.24}$$

with Λ corresponding to the lagrangian multipliers.

5.2.3 Main drawbacks of treating fragmentation at DFT level

Fractional charges

For the way DFT is formulated, there is no functional so far that is able to describe the integer nature of the electrons. This arises the problem of assigning the charges to the fragments when unphysical fractional charges are distributed over the system.

Basis set superposition error

The impossibility of using a complete basis sets for the calculations inevitably gives birth to errors on the absolute energies calculated. As we are interested

in relative energies, our goal would be to keep the error constant on different calculations. A first move to achieve this is certainly to perform all the calculations with a same basis set. Unfortunately this is often not enough. Infact, the quality of the basis set is not the same at all geometries, owing to the fact that the electron density around one nucleus may be described by functions centred at another nucleus. This is especially troublesome when calculating small effects, such as energies of van der Waals complexes and hydrogen bonds. Consider for example the hydrogen bond between two water molecules. The same happens in fragments interacting through H-bonds, here basis functions from one molecule can help compensate for the basis set incompleteness on the other molecule, and vice versa. This results in the artificial decrease of energy of the system, known as *Basis set superposition error* (BSSE) [114]. For intramolecular cases, however, it is difficult to define a unique procedure for estimating the BSSE, and it is almost always ignored.

Part III

Results & Discussion

Chapter 6

Ion-induced molecular growth in clusters of small hydrocarbon chains

Carbon forms the basis of the majority of the molecular species that so far have been identified in space [7]. A large portion of the interstellar carbon - perhaps more than 10 percent - is expected to be contained in some of the largest molecules identified in space, such as the Polycyclic Aromatic Hydrocarbons (PAHs) and the fullerenes [140, 141]. As already mentioned in section 1.2, despite the ubiquity of PAHs and fullerenes in space, the mechanisms for formation and growth of aromatic systems in these environments remain largely unknown [142]. Several experimental and theoretical studies have been already exploring possible formation pathways. Some bottom-up processes were proposed where carbon rings are formed through the sequential addition of small hydrocarbon building blocks such as CH_3 and C_2H_2 [143–151]. However, the low densities and fragility of the reaction intermediates, likely to photodissociate before they may grow further in the harsh environment of the ISM, limit the applicability of this model [152] Instead, top-down models are often used to predict the formation of fullerene cages in the ISM through the decay of larger carbon-based complexes [153], *e.g.* large PAHs, when these are processed by energetic photons [141, 154, 155], *e.g.* by the intense UV radiation fields in planetary nebulae, or particles

[156], *e.g.* by energetic atoms/ions in supernova shockwaves. The formation pathways of these large PAHs are also largely unknown.

Recently it has been shown that keV ions colliding with loosely bound clusters of PAHs or fullerenes can induce molecular growth within these clusters [1, 157–160]. The mechanism of this growth has been discussed in section 1.2. Here we investigate if impulse driven (knockout) reactions like those observed in PAH and fullerene clusters can drive the growth of small hydrocarbon chains into larger structures and to form aromatic molecules from non-aromatic ones.

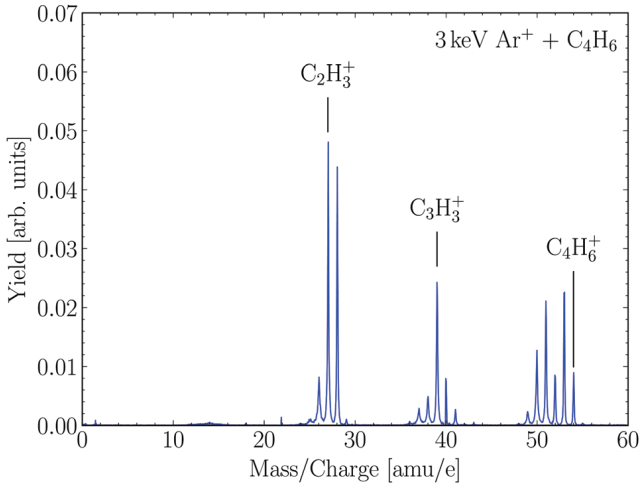


Figure 6.1: Mass spectrum of isolated C_4H_6 molecule in collision with 3 keV Ar^+. [142].

The small hydrocarbon under investigations is 1,3-butadiene (C_4H_6). The experiments shown in this work were performed at the AIRIBE facility in Caen University (see section 2.1). Figure 6.1 shows a mass spectrum for 3 keV Ar^+ ions colliding with isolated butadiene molecules. This is realized by operating the cluster source at room temperature and without the buffer gas. The fragment mass spectrum exhibits strong peaks from H-loss (without C-loss) resulting in fragments down to C_4H^+ at 49 amu per e. We observe an odd-even effect in the number of H atoms for the $C_4H_x^+$ fragments, where the odd-numbered fragments (which have closed electronic shells) being significantly more abundant than those with even numbers of H atoms. Other notable features are the strong peaks due to $C_2H_3^+$,

$C_2H_4^+$, and $C_3H_3^+$ fragments. This last one is the main product observed in photodissociation [161] and electron impact [162] experiments with 1,3-butadiene. Nevertheless, collision induced dissociation produces a wider range of fragments than the one from the photodissociation experiments [161]. Moreover we see smaller fragments, *e.g.* $C_2H_3^+$, $C_2H_4^+$, than in the electron impact experiments [162]. These differences could be caused by the molecules having higher excitation energies in our experiments, but could also be influenced by knockout-driven fragmentation.

6.1 Collisions with butadiene clusters

The 1,3-butadiene molecules condense into weakly bound clusters, when cooling the cluster aggregation source to 77 K. Fig. 6.2 shows the mass spectrum of intact, singly ionized clusters ranging in sizes of up to at least 17 molecules from collisions between these neutral clusters and the 3 keV Ar^+ projectiles. These charged clusters mainly come from large distance electron capture by the projectile ions, which leaves the charged clusters with low internal energies. Due to the low binding energy of the butadiene clusters (190 meV per molecule, according to our MD simulations of large clusters, 120 meV per molecule in DFT calculations of dimers [163]), even the low excitation energies involved when the clusters are ionized, together with other processes such as ion induced dipole interactions within the charged clusters, could be enough to induce some cluster dissociation. This dissociation primarily leads to the evaporation of intact molecular monomers from the charged clusters.

Some interesting features may be observed if we zoom in the region from a little below the intact molecular cation ($\sim$ 50 amu per e) up to a mass slightly above that of the molecular dimer (130 amu per e) of the mass spectrum in Fig. 6.2 (left panel of Fig. 6.3). Here, we see that a wide range of products with mass-to-charge ratios greater than $C_4H_6^+$ (54 amu per e) are detected. The products consist of molecules containing between 5 and 9 C atoms, each with a different number of H atoms. The most prominent peak in this spectrum corresponds to the protonated butadiene cation, $C_4H_7^+$. The $C_5H_x^+$ group stands out with only $C_5H_7^+$ giving a strong signal. $C_6H_x^+$ and $C_7H_x^+$ clusters present broader distributions in the number of H atoms.

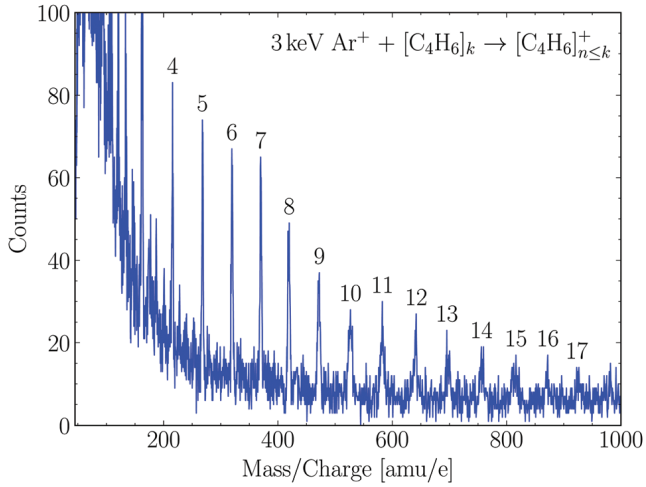


Figure 6.2: Mass spectrum for $[C_4H_6]_k$ in collision with 3 keV Ar^+. The mass per charge distribution shows the presence of intact, singly charged butadiene clusters, $[C_4H_6]^+_{n\leq k}$. The numbers above the peaks indicate the number, n, of intact molecules in a given cluster size [142].

[163]

However, it is worth notice that molecules with odd numbers of H atoms are significantly more abundant than those with even numbers. The only exception to this is the $C_6H_8^+$ peak, which is comparable to the $C_6H_9^+$ peak in intensity. Among the $C_8H_x^+$ peaks, a signal at 108 amu per e is detected, even if it has been labelled in Figure 6.3 as $C_8H_{12}^+$, it could also be due to intact, singly charged van der Waals dimers $[C_4H_6]_2^+$. The products with 9 C atoms do not show the odd-even effect for the number of H atoms.

The right panel of Figure 6.3 displays the computational mass spectrum from our MD simulations (see section 4.3) of 3 keV Ar atoms colliding with $[C_4H_6]_{50}$ clusters. The details regarding the simulations are presented in section 4. At a first qualitative look, we can conclude that the simulations agree with the experimental results, as there are many different growth products formed in the collisions. However, the simulations predict much broader distributions in the number of H atoms in products with any given number of C atoms than what we observe in the experiments (especially for the products with 7 or less C atoms). As already observed in pyrene clusters [158], the main growth mechanism involves the knockout of reactive atoms

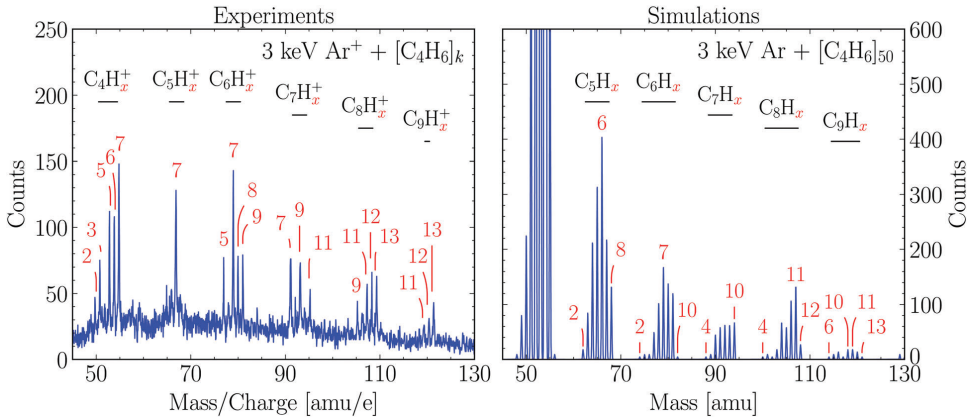


Figure 6.3: Left panel: collisional mass spectrum of neutral butadiene cluster 3 keV Ar^+ projectiles, showing the mass range from below the intact $C_4H_6^+$ to products consisting of up to 9 C atoms. Right panel: corresponding mass spectrum from classical MD simulations. Clusters of peaks are grouped by the number of C atoms and each peak is labelled (in red) by the number of H atoms it contains. In the simulated mass spectrum, just the lowest mass, highest mass, and most abundant growth products for each number of C atoms are labelled. [142].

or fragments from individual molecules, such fragments rapidly form bonds with neighbouring molecules on sub-picosecond timescales. Figure 6.3 shows a marked difference between the intensity of the peaks in the C_4H_x group in the simulated and experimental spectrum. The main reason is related to the fact that the experimental mass spectrum just detects charged fragments; whereas simulations take into account all reaction products, including *e.g.* evaporation of neutral C_4H_x due to cluster dissociation.

The different timescales between experiments and simulations (hundreds of μs vs a few tens of ps) also considerably influences the mass distributions obtained. In fact, on the experimental timescales the reaction dynamics is likely to be followed by secondary fragmentation, where the newly formed molecules undergo additional fragmentation. On the other hand, considering the simulation length and the model used, this is not expected to be the case for the simulated mass spectrum. In particular, secondary fragmentation plays a much more important role here than in the case of molecular growth inside pyrene clusters. The computational and experimental results are much closer to each other in the case with pyrene clusters [158]. Additionally, in the butadiene case one should consider the role of charge, which is not included in

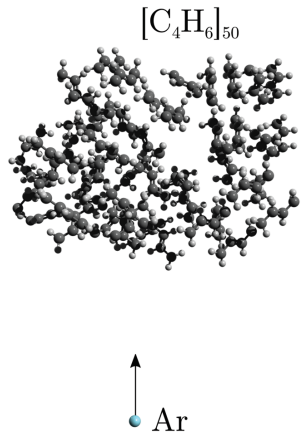


Figure 6.4: Schematic representation of butadiene cluster collision with an Ar atom[142]. $[C_4H_6]_{50}$ cluster optimized with AIREBO potential [142].

the present classical simulations. For larger molecules with many delocalized electrons, like PAHs and fullerenes, the reactivity and stability is largely the same regardless of whether the molecule is neutral or a singly-charged cation [159]. This is likely not the case for smaller molecules, like those that we detect here. The fact that the strongest features of the experimental mass spectrum ($C_4H_7^+$, $C_5H_7^+$, $C_6H_5^+$, $C_6H_7^+$, $C_6H_9^+$, $C_7H_7^+$, $C_7H_9^+$, $C_7H_{11}^+$ etc.) all have closed-shell electronic structures does indeed suggest that the charge of the molecules affects their stability. Despite the differences between the experimental and theoretical results, both are consistent with prompt knockout of individual atoms acting as the main mechanism responsible for driving the first steps of the bond-forming reactions. Nevertheless, other processes, *e.g.* heating of the clusters through electronic scattering or through the formation of radicals by ionization, may also play a role in the growth processes we observe. In fact, photoionization of ethylene (C_2H_4) clusters by synchrotron radiation has already been shown to induce molecular growth in those systems [164].

The presence of "magic" peaks (dominant peaks) in our experimental results suggests that only the most stable molecules produced in the collisions survive long enough to be detected on the experimental time scale of tens of microseconds. From our experiments, it is not possible to obtain a structure

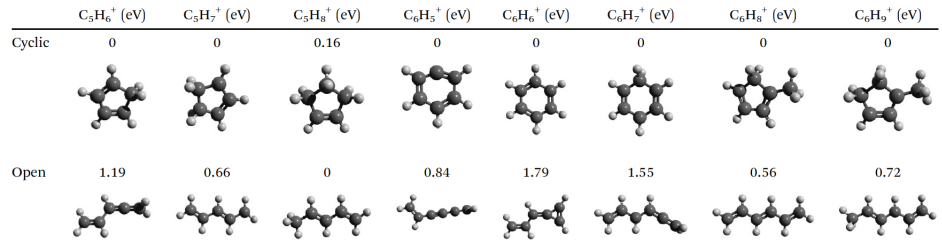


Figure 6.5: Lowest energy cyclic and open isomers for $C_5H_x^+$, where x = 6, 7, 8, and $C_6H_x^+$, where x = 5, 6, 7, 8, 9, with zero-point corrected energies relative to the global minima for each size at the B3LYP/CC-pVDZ level of theory (see text)[142].

for the mass/charge detected. At this point, Density Functional Theory (DFT) (section 3.2.3) calculations of all possible isomers for a given cluster C_xH_x were performed in order to determine the most stable structures. Due to the large number of possible isomers for each species, *e.g.* C_7H_6 alone has 1230 unique isomers, we have limited our search to $C_5H_x^+$, where x = 6, 7, 8, and $C_6H_x^+$, where x = 5, 6, 7, 8, 9. Among all the possible isomers, we show in Figure 6.5 just the lowest energy open (linear or branched carbon backbone) and cyclic (only 5 or 6 elements rings were considered) structures, minimized at the B3LYP/CC-pVDZ level of theory. The minimized rings analyzed are more stable than the lowest energy open structures in all cases, except the one of $C_5H_8^+$. We see that the energy difference between open and cyclic structures varies significantly, for example it is nearly 2 eV in the case of $C_6H_6^+$. According to our classical MD simulations most growth products are created when small knocked out fragments form new covalent bonds with intact C_4H_6 molecules. Products formed in this way will initially have open structures, and most probably high potential energies with respect to the cyclic isomers. During its journey towards the detector, this excited molecule has time to relax to the most stable geometry (the cyclic one according to our calculations). In this moment, the extra energy to release will be converted into vibrational energy. This, together with the energy deposited in the collisions, could lead to secondary fragmentation. Phenomena like this could help explain why some products are less abundant than other similarly sized ones, e.g. $C_6H_6^+$ to $C_6H_5^+$. It is worth mentioning that the energy excess may be released as photons through radiative cooling, an effect which has also not been included in our simulations.

We may get further information using the DFT energies as in [165] and determining the δ parameter (equation 6.1) for each product.

$$\delta = \frac{[E(C_nH^+_{x+1}) + E(C_nH^+_{x-1})]}{2} - E(C_nH_x) \tag{6.1}$$

E($C_nH^+_x$) is the total electronic energy of the most stable isomer of the $C_nH^+_x$ ion. This term allows us to compute the stability (in terms of the enthalpy of formation) of a molecule relative to the mean of its nearest neighbours in size, *i.e.* molecules with one less H atom and one additional H atom. As a consequence, positive values of δ indicates that a molecular ion is (relatively) more stable than its neighbours and, in the same way, ions that are less stable than their neighbours will have negative values of δ.

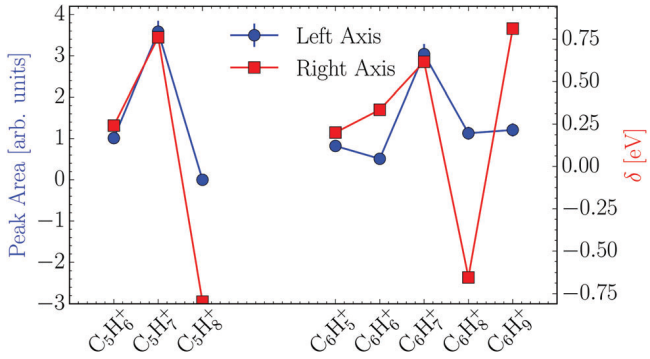


Figure 6.6: Left vertical scale: Integrated peak intensities (blue circles) of $C_5H^+_x$ and $C_6H^+_x$ reaction products following 3 keV $Ar^+ + [C_4H_6]_k$ collisions. Right vertical scale: Relative stabilities of $C_5H^+_x$ and $C_6H^+_x$ as defined by equation 6.1 in the text (red squares).[142].

In Figure 6.6 we plotted the calculated δ-values for reaction products with 5 and 6 C atoms (in red), together with the integrated areas of the respective peaks in the experimental mass spectrum (in blue). The oscillations in the δ-values between different molecules agree with the experimental odd-even effects found in the abundances of the reaction products. This qualitative agreement supports the hypothesis that an initially broad distribution of hot reaction products as the one we got from our simulations may evolve in distributions with "magic" features (as seen in the experiments) through losses of H atoms. It is important to point out that, since δ only compares the stability of a molecule with those nearest in size, it is a parameter to

measure the relative stability of a molecule against the loss of single H atoms. Nevertheless, the hot molecules in a decaying cluster could may lose not only individual H atoms, but also one or more C atoms or H_2 molecules. This means that a complete theoretical description of the secondary fragmentation involved in our system that reshapes the primary products would have to consider all possible statistical fragmentation channels (see *e.g.* ref. [166] and [167]).

Chapter 7

Knockout driven fragmentation of porphyrins

In section 1.4.1 we have introduced the context of a study on tetraphenylporphyrin (TPP) and its iron complex (FeTPP and ZnTPP). As already remarked, this type of study is particularly useful for understanding the effect ionizing radiation may have on this class of molecules, considering that many enzymes [168, 169], chromophores [23] and proteins essentials to life [21, 22] contain porphyrin structures.

The experiments related to this work were conducted in the EIS lab at Stockholm university, so they involved the collision between singly charged TPP and an atom of He. The details regarding this facility and the type of experiments performed may be found in section 2.2.

As for the calculations, the collisions TPP-He were modelled as presented in section 4.3. Ordinary Tersoff potential was used, as the AIREBO is not parametrized for atoms other than C and H. Calculations with Density functional theory (DFT) at the B3LYP/CC-pVDZ level of theory allowed us to obtain the adiabatic dissociation energies for some of the dominant fragmentation channels of TPP and FeTPP molecules. The computational part of this work has been carried out with the aid of the Gaussian 09 software package [91] and LAMMPS package [121].

7.1 He/Ne collisions with porphyrin isolated molecules

Figure 7.1 shows the mass spectra due to $(TPP+H)^+ + He$ (top panel), $FeTPP^+ + He$ (middle panel) and $ZnTPP^+/(ZnTPP+H)^+ + He$ (bottom panel) collisions at center-of-mass energy (E_{CoM}) = 50 eV. The fragmentation yields in each panel are normalised to their respective total fragmentation cross section. The latter have been measured separately by means of the beam attenuation method. The three spectra display similar features with the dominant peak (off scale in Figure 7.1) corresponding to the intact molecule.

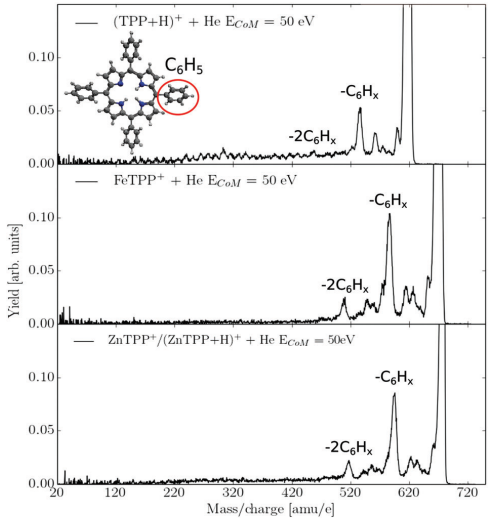


Figure 7.1: Collision-induced dissociation spectra for (from top to bottom): $(TPP+H)^+ + He$, $FeTPP^+ + He$ and $ZnTPP^+/(ZnTPP+H)^+ + He$ collisions at a center-of-mass energy of 50 eV. The labels $-C_6H_x$ and $-2C_6H_x$ indicate, respectively, losses of one or two phenyl groups from the cation.

The second most prominent peak has been identified as the loss of a phenyl group (C_6H_5, indicated by the red circle in Figure 7.1), which most likely is accompanied by the loss of an additional hydrogen atom [28]. In the case of the MTPP molecules (M= metal, here Fe or Zn, middle and bottom spectra in Figure 7.1), the absence of small fragments (with masses lower than 500 amu per e) show that the molecules are more stable with than without the metal. Zooming in between the primary and the single phenyl-loss peaks (Figure 7.2), some interesting features are found.

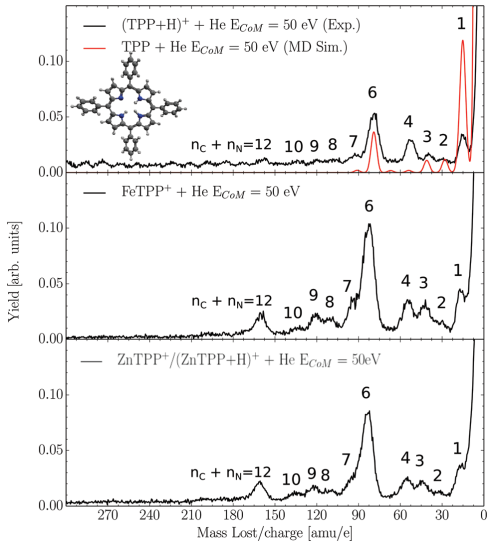
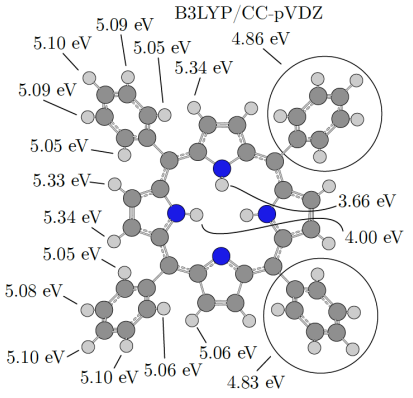
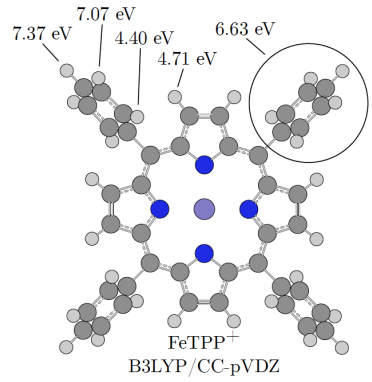


Figure 7.2: Zoom-ins of CID spectra for (from top to bottom): $(TPP+H)^+ + He$ (in black), $FeTPP^+ + He$ and $ZnTPP^+/(ZnTPP+H)^+ + He$ collisions at a center-of-mass energy of 50 eV. The labels indicate the number of carbon atoms and nitrogen atoms (n_C+n_N) lost from the porphyrin ions. The simulated spectra, as obtained from the classical MD, is shown in red, top panel.



(a) $(TPP+H)^+$



(b) $FeTPP^+$

Figure 7.3: An explicative scheme collecting all the dissociation energies calculated for: (a) $(TPP+H)^+$ and (b) $FeTPP^+$.

The peaks in Figure 7.2 are labelled by the number of heavy atoms (C and/or N) that are lost from the parent ion. These peaks are separated by masses corresponding to a single C or N atom, typically accompanied by the loss of a small number of H atoms. The mass spectra in Figure 7.2 are similar to those measured earlier for collisions between $(TPP + H)^+$ and He at a center-of-mass energy of 278 eV [170] and for collisions with protonated protoporphyrin IX and Fe(III)-heme [171, 172]. There as well [170–172], peaks corresponding to the loss of one or several heavy atoms clearly appear, but the actual fragmentation mechanism was not discussed. The fragment distributions in Figure 7.2 and those reported in ref. [170–172] are markedly different from those observed in other experiments [24, 29, 31]. Fragments from the loss of the phenyl groups were seen following 30 keV O^{3+} and 70 eV electron impact on neutral FeTPPCl [24, 31]. In the latter case, fragmentation due to loss of 1-4 heavy atom was also observed but only with intensities that are orders of magnitude lower than in the present experiments. Gozet et al. [29] studied the fragmentation of $FeTPP^+$ in low-energy multiple collisions with N_2. In this case, the most intense fragmentation peaks observed are the ones corresponding to the loss of one or two phenyl groups and the loss of one or several H atoms. According to our electronic structure calculations these channels are the ones with the lowest dissociation energy associated (Figure 7.3). In particular, the dissociation energies for the loss of H atoms bonded to C atoms are between 5.0 eV and 5.3 eV in the case of $(TPP^+H)+$ (Figure 7.3a); whereas, the dissociation of H atoms connected to N atoms in the centre of the TPP ring is around 4 eV, and 4.8 eV for phenyl loss. As for FeTPP+ (Figure fig:tpp-3-b), dissociation energies range between 4.4 eV and 7.4 eV for H atoms bond to C atoms and the dissociation energy is 6.6 eV for phenyl loss.

On the other hand, the losses of one or several heavy atoms (with or without hydrogen loss) are associated with significantly higher dissociation energies (except for the previously mentioned phenyl groups). These channels were not observed in the low energy collisions with N_2 [29], but are clearly seen in our mass spectra (see Fig. 7.2). This suggests that the spectra recorded in ref. 17 are dominated by statistical fragmentation processes, while we have significant contributions from knockout processes in the present work. Another important support to this conclusion comes from the MD

simulations. In Figure 7.2 top panel, we show a comparison between the experimental mass spectrum (in black) and classical MD simulation for direct atom knockouts (in red) in collision between protonated TPP and He at a center-of-mass energy of 50 eV. Because the TPP and MTPP experimental spectra display similar features, we may use the TPP MD results to guide the interpretation of the experimental results for all three porphyrins (see Figure 7.2). The experimental spectrum is normalised to the measured absolute total fragmentation cross section, while the simulated one is normalised to the absolute total cross section for prompt single- and multiple-atom knockouts from the simulations. This value is calculated as the ratio between the number of events that lead to the knockout and the number of total trajectories, which is then multiplied by the area through which the projectiles passes in the trajectories considered. The equation of heavy atom knockout calculation used is shown in equation 7.1

$$\sigma_{KO}^{MD}(C,N) = \frac{n_{traj}^{KO}}{n_{traj}^{TOT}} A_{eff} \times \frac{4}{3} \tag{7.1}$$

The multiplication by the factor 4/3 comes from previous studies, as for PAHs it has been shown that experimental knockout cross sections are approximately 4/3 larger than the calculated cross sections for the same energy range [173]. Neither of these cross sections include losses of H-atoms (see Table 7.1). Comparing the experimental and simulated (heavy atom knockout) spectra, we may conclude that they partially display the same fragmentation channels, although the branching ratios are different. One peculiar difference is the very intense peak corresponding to one heavy atom with respect to the four atoms loss, opposite to what is observed in the experiment. This mismatch may be explained by the difference in timescale between the MD simulations and the experiments. The simulations only consider prompt fragmentation by nuclear stopping processes and following secondary fragmentation on the picosecond timescale. The experimental timescale is much longer (hundreds of microseconds) such that delayed statistical fragmentation and secondary fragmentation processes following knockout may have time to occur to much larger extents. The experimental results suggest that there is a preference for the loss of three heavy atoms following single-atom knockouts.

Table 7.1: Experimental absolute total fragmentation cross sections for 80 eV $(TPP + H)^+ + He$ and 110 eV $(TPP + H)^+ + Ne$ collisions (in units of $10^{-15} cm^2$) and MD simulations of the corresponding knockout (KO) cross sections. The experimental results do not include the cross sections for single and multiple H-losses. The MD results are separated for the knockout of heavy atoms, $\sigma_{KO}^{MD}(C,N)$, and H atoms, $\sigma_{KO}^{MD}(H)$.

Target	E_{CM} [eV]	σ_{TOT}^{Exp}	$\sigma_{KO}^{MD}(C,N)$	$\sigma_{KO}^{MD}(H)$
He	80	4.1 ± 0.5	1.69 ± 0.03	2.72 ± 0.04
Ne	110	7.3 ± 0.9	5.02 ± 0.05	2.19 ± 0.03

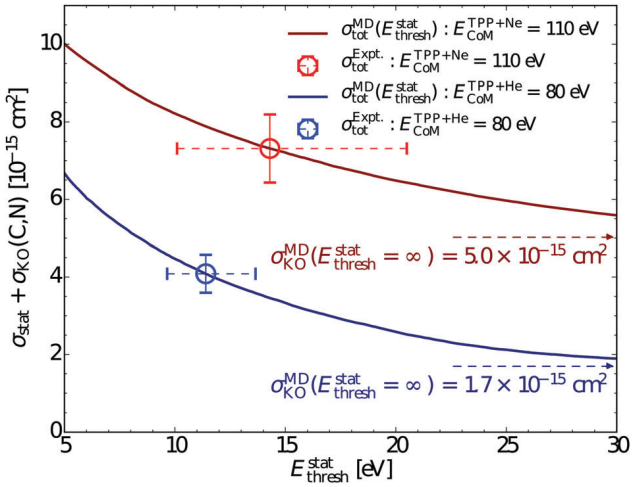


Figure 7.4: The calculated absolute total fragmentation cross sections (excluding H-loss) for TPP in collisions with He at center-of-mass energies of 80 eV (blue solid line) and Ne at 110 eV (red solid line) as a function of the assumed value of the threshold energy for statistical fragmentation. The circles are the experimental absolute fragmentation cross sections. The solid error bars represent the systematic uncertainties in the measured fragmentation cross section, which have been translated to uncertainties in the threshold energy (dashed error bars). The asymptotic values in the inset (σ_{KO}^{MD}) are total knockout cross sections obtained from the MD simulations (dashed lines), excluding H-loss, which are about 40% and 70% of the absolute total cross section for He and Ne, respectively, under the present experimental conditions.

By comparing the simulated cross sections for depositing different amounts of energy in the TPP molecules and the measured total fragmentation cross section, we may determine a semi-empirical threshold energy for statistical fragmentation. The curves in Figure 7.3 show the sums of calculated statistical and knockout cross sections as a function of the value of an assumed threshold energy for statistical fragmentation, E^{stat}_{thresh}. The data points are the present experimental values for the absolute total fragmentation cross sections (see Table 7.1). Based on the experimental uncertainties and the statistical uncertainty of the simulations, E^{stat}_{thresh} is $11.4(^{+2.2}_{-1.8})$ eV for collisions with He, and $14.3(^{+6.2}_{-4.2})$ eV for collisions with Ne. A weighted average of $E^{stat}_{thresh} = 12.8 \pm 1.4$ eV gives the energy required to induce statistical fragmentation on the experimental (microsecond) timescale. This threshold energy is significantly higher than the dissociation energy, which is similar to what has been observed for PAHs [174] and fullerenes [175]. This reflects large heat capacities for these systems that protect them from prompt decay even when the excitation energy is larger than the dissociation energy. The asymptotic value of the calculated absolute total fragmentation cross section for infinite E^{stat}_{thresh} corresponds to the total knockout cross section, $\sigma^{MD}_{KO}(C,N)$ (dashed lines in Figure 7.3) for the MD simulations. For He atoms colliding with TPP at 80 eV, $\sigma^{MD}_{KO}(C,N) = 1.7 \times 10^{-15} cm^2$, and for Ne, $\sigma^{MD}_{KO} = 5.0 \times 10^{-15} cm^2$ at 110 eV. Relying on these numbers, we find that $\sim 40\%(\sim 70\%)$ of the total fragmentation cross section of TPP in collision with He (and Ne) is due to non-statistical, *i.e.* knockout, fragmentation. These numbers are comparable to those for PAHs under similar conditions [7]. The values of the knockout cross sections are governed by the energies required to permanently displace (remove) an individual atom from the molecule, which depend on the atom type and its chemical environment within the molecule. Here, we determine this molecular property (displacement energy) through MD simulations without using a specific projectile trajectory but displacing the atom in randomly generated straight line directions. In Figure 7.5, we show the distributions of displacement energies for prompt knockout of a single N (in blue), C (in green) or H (in red) atom. The mean values of these broad energy distributions are 24.7 eV for knockout of a N atom, 30.6 eV in the case of C and 6.7 eV for an H atom. For the PAH molecule coronene ($C_{24}H_{12}$), the mean value of the

displacement energy calculated using the same method for removing a C atom along any trajectory is slightly higher (35.4 eV) [1] than in the case of porphyrin molecules. This and the similarities in the dissociation energies [176] for (M)TPP and PAH cations, explain why the branching ratios between statistical and non-statistical fragmentation (knockout fragmentation) are comparable for the PAHs and the (M)TPP systems.

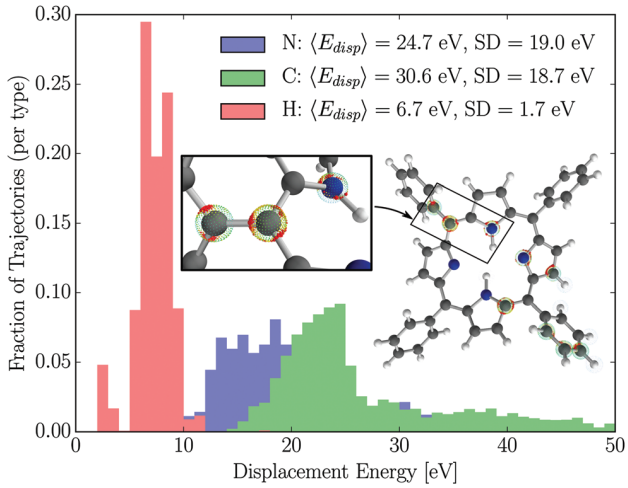


Figure 7.5: Distributions of displacement energies for prompt knockout of a single N (in blue), C (in green) or H (in red) atom from TPP molecules obtained from the present classical MD simulations using the Tersoff force field. The reported values correspond to the mean values of the displacement energies for each type of atom in the molecule (SD is the standard deviation of each distribution). The pictures represent an example of the displacement energies for two different carbons and one nitrogen position in the TPP molecule. Red dots in the inset show directions with the higher displacement energies and blue dots shows those with the lower displacement energies.

Chapter 8

Dissociative Decay of Protonated and Deprotonated Adenine

Section 1.4.2 presented adenine as the best candidate, among nucleobases, to survive the ISM condition [39]. Its biological role as one of the DNA basis is very well known. The combination of experimental and theoretical studies presented in this chapter focused on the dissociative behaviour of protonated and deprotonated adenine when exposed to energetic He projectiles.

The experiments were carried out using the single-pass experimental setup (EIS-LABoratory) [82], at DESIREE infrastructure [83, 84] in Stockholm. Here the molecular ions of adenine (protonated or deprotonated) are accelerated at the desired center of mass energy (E_{CoM}) to collide with neutral He atoms. The pressure inside the He chamber is regulated to ensure single collision conditions. For further details regarding this facility and the experimental conditions the reader is redirected to section 2.2 of this thesis.

The computational modeling of these systems involved the simulation of collision sequences with a combination of Tersoff and ZBL potentials, as described in section 4.3. LAMMPS package [121] was used to run these trajectories. Overall, a 100000 trajectories were accumulated per E_{CoM}. According to a well-implemented procedure [173], we calculated the heavy atom (carbon and nitrogen) knockout cross section (chapter 7). As for the

energy deposited by the projectile on the molecule, it was calculated as the difference in the kinetic energy of the projectile before and after the collision.

Density functional theory (DFT) at the B3LYP/6-311++g(d,p) level of theory was employed for the minimum and transition state (TS) search, aimed to reconstruct part of the potential energy surface of the molecules under investigation. In addition ab initio molecular dynamics (AIMD) simulations on the isolated molecule were carried out, using the atom-centered Density Matrix Propagation method (ADMP) [133], employing the B3LYP [110, 111] functional together with the 6-31++G(d,p) basis set [115, 116]. Each trajectory had a timestep (Δt) of 0.5 fs, a fictitious mass (μ) of 0.1 amu and a maximum simulation time (t_{max}) of 4.0 ps. In order to mimic the effect of the helium atom depositing energy on the protonated and deprotonated adenine, we have collected statistics for different initial nuclear kinetic energies (NKE, here also called internal energy), specifically 10, 15, 20, 25, 30 eV, and chosen the energy value which would give the right amount of fragmentation at a maximum t of 4 ps. Considering the large amount of energy deposited in the system, we have have performed a full convergence at each step in order to preserve the adiabaticity of the system and to avoid the excessive diminishing of the Δt employed, which would have increased the computational time and cost of each trajectory.

8.1 Fragmentation cross-sections

Figure 8.1 shows the experimental values of the absolute total fragmentation cross section of protonated adenine (black squares) and deprotonated adenine (red triangles) as a function of the center-of-mass collision energy. The experimental cross sections are similar for both species and display a weak collision energy dependence, suggesting a comparable stability between the deprotonated and protonated adenine molecule.

Previous studies [1, 37, 177] have presented evidence that sub-keV E_{CoM} collisions may lead to a single or multiple atoms knockout. Nevertheless, if the energy deposited on the system is high enough, even after the knockout event, a secondary fragmentation may take place. This excitation process, which mainly involves the electrons of the system, is suspected to be the dominant effect for small molecules such as adenine [177]. As a matter of

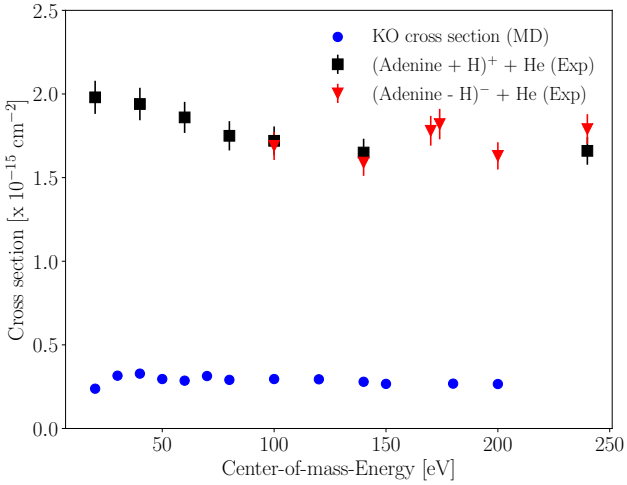


Figure 8.1: Absolute total fragmentation cross section of protonated adenine (black squares), deprotonated adenine (red triangles) and the heavy atom knock out (KO) cross section from molecular dynamics simulation (green dots) in function of the center-of-mas energy.

fact, already for systems bigger than adenine, such as anthracene ($C_{14}H_{10}$), acridine ($C_{13}H_9N$) or phenazine ($C_{12}H_8N_2$), it has been shown that their radical cations in collision with 110 eV He have a heavy atoms (C, N) knockout cross section which is just about 20% of the total fragmentation cross section [173].

The corresponding heavy atom knockout cross section for neutral adenine from the classical molecular dynamics simulation is represented in Figure 8.1 (blue dots). In table 8.1 the total fragmentation cross section values (σ_{TOT}^{Exp}) are reported together with the corresponding calculated heavy atoms knockout fragmentation cross section ($\sigma_{KO}^{MD}(C,N)$). The $\sigma_{KO}^{MD}(C,N)$) contributes to about 10 % of the total fragmentation cross in the present collision energy range, which is consistent with the results for anthracene, acridine and phenazine.

The simulated cross sections, $\sigma_{KO}^{MD}(C,N)$, suggests that statistical fragmentation is the dominant decay pathway. From the MD-simulations we can extract the energy deposited for collision events where the molecules survive on the simulation timescales. Depending on the amount of energy being

Table 8.1: Experimental absolute total fragmentation cross sections (in units of $10^{-15} cm^2$) for $[Ad+H]^+$ and $[Ad-H]^-$ in collision with He atoms at different E_{CoM} values and the corresponding heavy atom knockout (KO) cross sections as calculated from MD simulations ($\sigma_{KO}^{MD}(C,N)$).

E_{CoM} [eV]	$\sigma_{TOT}^{Exp}([Ad+H]^+)$	$\sigma_{TOT}^{Exp}([Ad-H]^-)$	$\sigma_{KO}^{MD}(C,N)$
100	1.72	1.69	0.296
140	1.65	1.59	0.279
240	1.66	1.79	0.266

deposited (*i.e.* internal energy), the molecules may then undergo statistical fragmentation on much longer (experimental) timescales. Figure 8.2 shows the distribution of energy deposited in the range 3 - 20 eV when the molecule interacts with the He atom at center-of-mass energy 40 (purple), 120 (red) and 240 (green) eV. The reported values refer to the percentage of collision events for which the molecule is left intact on the timescale of the simulation, which will be called *inactive* trajectories.

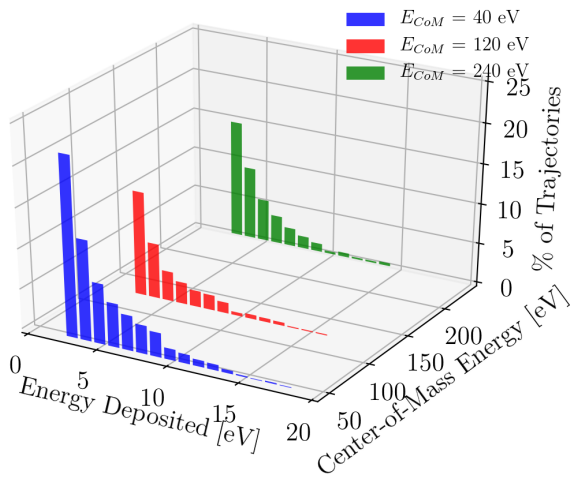


Figure 8.2: Distribution of energy deposited in the inactive trajectories for 40 (purple), 120 (red) and 240 (green) eV of center-of-mass energy for the He projectile. A trajectory is defined inactive when the molecule is left intact after the interaction with the He atom.

The present internal energy distributions are broad, extending up to 16 eV for 120 eV, to 14eV for 240 eV and to 18 eV for 40 eV. In order to get insights for statistical fragmentation processes we have been performed *ab initio* molecular dynamics simulation with the ADMP method at different NKE. Considering the timescale difference between experiments and theory (hundreds of microseconds vs a few units of ps), it was necessary to use higher initial internal energy in order to see fragmentation processes at computationally affordable maximum simulation times. For this reason, considering our tests at different NKEs, we used 20 eV and 15 eV as internal energies in our ADMP simulations of, respectively, internally heated protonated and deprotonated adenine molecules.

8.2 Protonated adenine stability

As already mentioned, protonated adenine has been more extensively studied in the literature compared to deprotonated adenine. Among them, some involving an extended computational study on the molecule isomers stability [178, 179], possible dissociative pathways for HCN neutral loss and NH_3 [179] and the electronic absorption spectrum [180]. These studies suggest the most thermodinamically stable structure of the $[Ad+H]^+$ to be the one in Figure 8.3, and specifically to be the one generated by the electrospray ionization method [179]. Isomerizations to structures protonated at the remaining N in the ring structure are all endoergic in bulk water, water clusters, and the gas phase, and these isomers are calculated to amount to $< 3\%$ at room temperature [179]. Nonetheless, isomerizations may occur if enough energy is deposited into the system, specifically for our case, after the interaction with a projectile. The isomerization barriers reported in [179] for $[Ad+H]^+$ are comparable to the ones in [39] for $[Ad-H]^-$, being the maximum barrier energy for proton transfer around 3.6 eV. Some trajectories may overcome this threshold, in fact some of the ADMP trajectories pass through an isomerization, as the thermodynamical pathways presented in both [179] and [39].

Figure 8.4 shows a comparison between the experimental mass spectrum for collisions $[Ad+H]^+$ - He at 240 eV (top panel in black) and the mass distribution obtained by the ADMP trajectories (bottom panel in blue) with

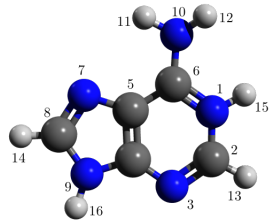


Figure 8.3: Minimum energy structure of $[Ad + H]^+$ most stable isomer as calculated at B3LYP/6-311++g(d,p) level of theory. All the atoms of the molecule have been numbered.

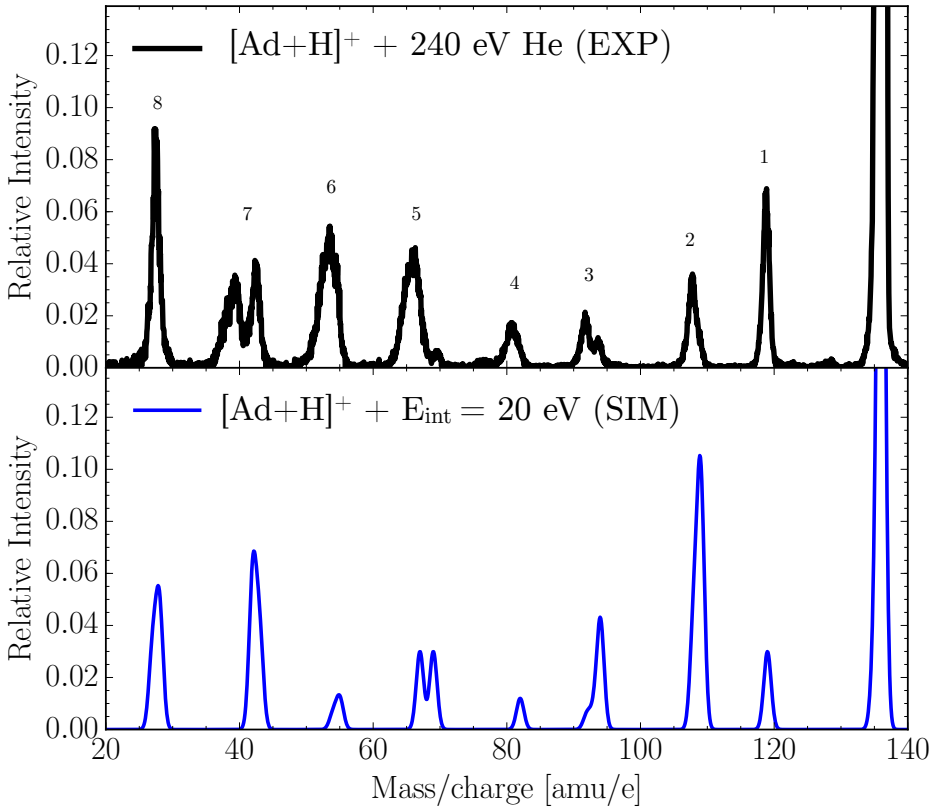


Figure 8.4: Experimental mass spectrum (top panel, black) for $[Ad+H]^+ + He$ at E_{CoM} = 240 eV and simulated mass distribution (bottom panel, blue) from the ADMP trajectories at NKE (= E_{int}) = 20 eV. The different clusters of peaks have been numbered.

initial nuclear kinetic energy of 20 eV. Table 8.2 reports the m/z detected and the masses of the fragments obtained from the dynamics. Most of the positions of the different groups of peaks presented coincide, except for one case. The number 7 is a double peak in the experiments and a single peak in the simulations. The m/z differerence between the two peaks in group 7 in the experiments has been assigned as 2 a.u./e. The presence of this double peak may be due to secondary fragmentation processes.

Table 8.2: All the masses detected are grouped per cluster in which they have been found. In parenthesis the corresponding formula or the neutral loss is evidenced.

Fragment Peak	m/z [a.u./e] (SIM)
1	119 ($-NH_3$)
2	109 ($-HCN$)
	108 ($-HCNH$)
3	94 ($-HNCNH$)
	92 ($-H_4CN_2$)
4	82 ($-2HCN$)
5	69 ($H_3C_2N_3^+$)
	67 ($H_3C_3N_2^+$)
6	55 ($-3HCN$)
	54 (82 $-H_2CN$)
7	43 ($H_3CN_2^+$)
	42 (H_2CN_2)
8	28 (H_2CN^+)
	27 (HCN^+)

The branching ratios of the experimental and computed peaks does not correspond for a several reasons. The considerable timescale difference between experiments and simulations is the main cause of such a discrepancy. As a matter of fact, the fragments generated may be subject of further fragmentations not observable on the present t_{max}. Moreover, the statistics accumulated, as already mentioned, is not suitable for a significative quantitative analysis of the mass distribution obtained. In fact, to be able to compare the branching ratio one should run a higher number of trajectories at all different internal energies obtained in Figure 8.2 for hundreds of

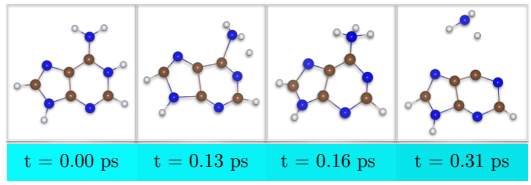
μseconds. In addition, around 10 % of the fragments may come from further fragmentation of molecules which have already suffered C, N, H knockout.

Examples of the pathways obtained with the ADMP dynamics for the $[Ad+H]^+$ are reported in Figure 8.5. The masses of the fragments obtained in these sequences coincide with the experimental cases. This makes the pathways reported possible for the formation of the corresponding fragments in the present experimental mass spectrum.

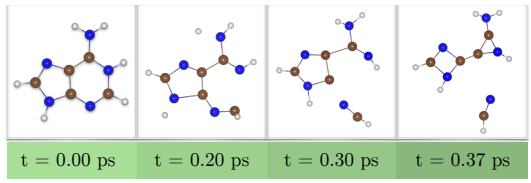
The dynamics unveil the presence of the fragments at 109 a.u./e, 82 a.u./e and 55 a.u./e, which correspond to the neutral loss of HCN, HCN + HNC and 2HCN + HNC, respectively. These very same species were already detected in previous CID experiments [50], where multiple collisions take place, with resulting higher energy deposited in the system with respect to our single collision case. Moreover, through the use of isotopic substitution, Nelson et al. could identify which one of the C atoms is lost as HCN from the system [50] in the case of the formation of the 109 a.u./e species. The results from our dynamics are in agreement with what observed in this study. The dynamics also provide a possible pathway for the reaction (path 109 a.u./e in Figure 8.6), which would proceed through the N1 - C2 bond cleavage to open the pyrimidinic ring, followed by the C4 - N3 bond breaking which releases HCN and leaves an isomerized structure with two rings for the $C_4H_5N_4^+$ (109 a.u./e) species. On the other end, the 82 a.u./e and 55 a.u./e formation pathways (in Figure 8.6) see the rupture of the smaller ring first. In particular, the C8 - N9 bond is broken. For the 82 a.u./e case, HCN is lost as a consequence of the C5 - N7 cleavage, then HNC loss follows the isomerization of the 6-elements ring to a 5-elements ring and the C5 - C4 bond breaking. As for the 55 a.u./e species, the smaller ring opening is followed by the pyrimidinic ring opening through the C4 - N3 rupture. The N1 - C2 cleavage allows the HCN loss. A second HCN is released when C5 - N7 is broken. Finally, from the isomerized $HNCC(NH_2)CNH^+$ fragment, an HNC molecule is loss.

The 94 a.u./e species comes from the neutral loss of the $HNCNH$ molecule. This mechanism involves the breaking of the bond N1-C2, followed by the C5-C6 bond cleavage which happens at the same time of the H15-transfer from N10 to C5. The product formed is a protonated isomer of

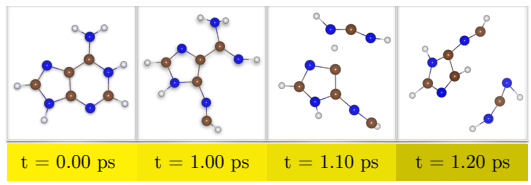
$C_5H_6N_5^+ \rightarrow C_5H_3N_4^+$ (119 a.u./e) + NH_3



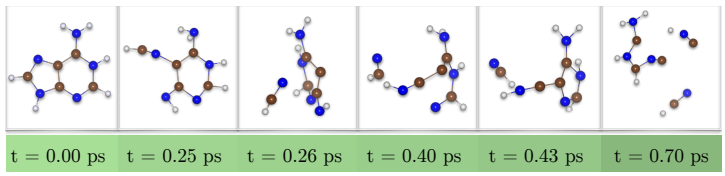
$C_5H_6N_5^+ \rightarrow C_4H_5N_4^+$ (109 a.u./e) + HCN



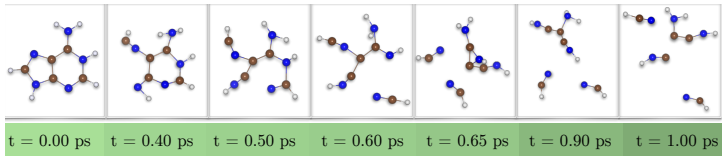
$C_5H_6N_5^+ \rightarrow C_4H_4N_3^+$ (94 a.u./e) + HNCNH



$C_5H_6N_5^+ \rightarrow C_3H_4N_3^+$ (82 a.u./e) + HCN + HNC



$C_5H_6N_5^+ \rightarrow C_2H_3N_2^+$ (55 a.u./e) + 2HCN + HNC



$C_5H_6N_5^+ \rightarrow HCN^+$ (27 a.u./e) + HNC + HNCH + $C_2H_2N_2$

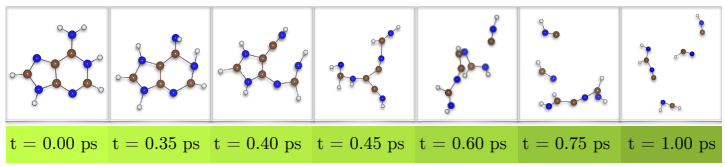


Figure 8.5: Snapshots from selected ADMP trajectories. The pathways may be responsible for the formation of the positive fragments (from top to bottom): 119 a.u./e (- NH_3) , 109 a.u./e (-HCN), 94 a.u./e (-HNCNH), 82 a.u./e ($-2HCN$), 55 a.u./e ($C_2N_2H_3^+$), and 27 a.u./e (HCN^+).

the product of the $HNCNH$ neutral loss for the deprotonated isomer [39], although the formation mechanisms proposed in the two cases differ.

As for the NH_3 loss, the dynamics propose a pathway that goes through the proton transfer from the pyrimidine ring, and subsequent neutral NH_3 release. In this case, the experiments did not detect a peak corresponding to the NH_2 loss. This is in agreement with the PES shown in Figure 8.6 show. For the protonated adenine the NH_3 loss mechanism in turquoise in Figure 8.6 is not competing with the others (in red and blue). As a matter of fact, both the NH_2 neutral loss (red pathway in Figure 8.6) and the NH_3 neutral loss proceeding through the H-transfer from the 5-elements ring (blue pathway in Figure 8.6) have a maximum barrier which is higher than 5.0 eV, so most of the trajectories would not deposit enough energy to overcome these barriers, whereas the turquoise pathway, with 3.93 eV as the highest barrier, would be the dominant one. In particular, this pathway has barriers considerably lower than the other two (by 1.30 and 2.73 eV, respectively for the red and blue pathways in Figure 8.6), which is consistent with the fact that in the dynamics this is also the dominant one.

The most intense peak detected in the experiments would correspond to a combination of HCN^+ (27 a.u./e) and H_2CN^+ (28 a.u./e). The former originated in a process shown by the last series of snapshots in Figure 8.5. Here, first of all, H11 is transferred to N7, then C6 - N1 is broken and the molecule is broken down to HCN^+, HNC, $HNCH$ and $C_2H_2N_2$.

8.3 Deprotonated adenine stability

The existence of different isomeric forms leads to a large variety of possible fragmentation pathways. Cole et al. studied some of the possible isomeric structures of $[Ad-H]^-$ and their isomerization barriers [39]. Among the possibilities explored the most stable isomer turned up to be the one presented in Figure 8.7. This considered, we took this structure as the input geometry for the ADMP trajectories. The 100 trajectories explored are not sufficient to draw quantitative conclusions, but they contribute to elucidate the fragmentation pathways responsible for the fragments observed in the experimental mass spectrum and propose a structure for them.

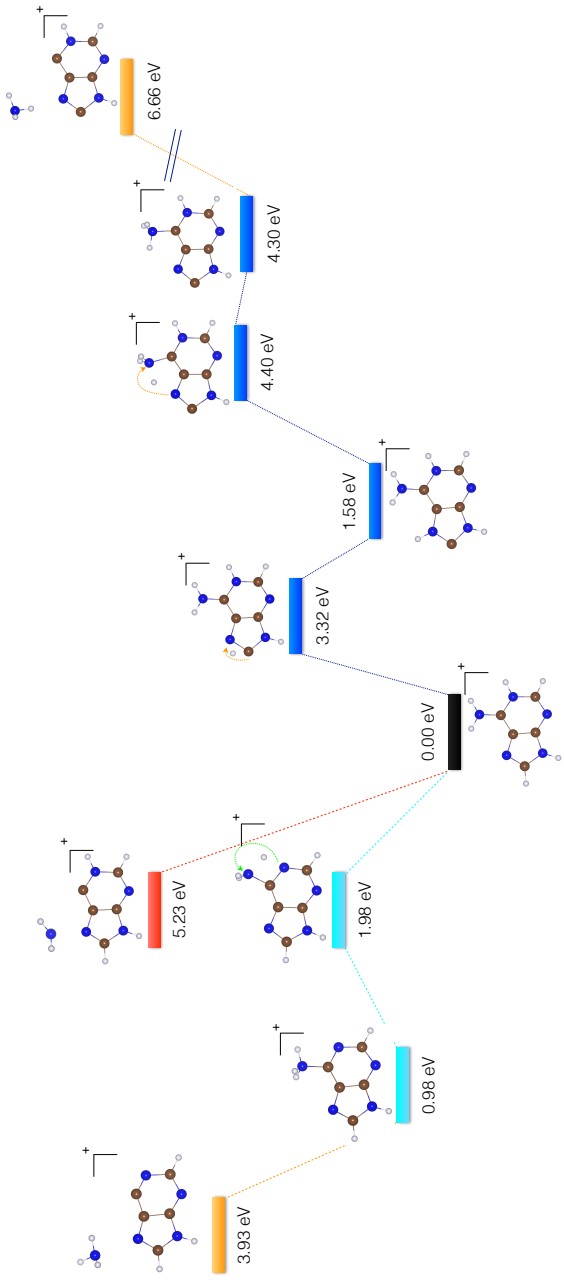


Figure 8.6: Dissociation pathways for NH_2 neutral loss (pathway in red), and NH_3 neutral loss. The NH_2 loss does not present a TS. There are two possibilities for the loss of NH_3. One option involves the isomerization through the progressive proton transfer of the hydrogen proceeding from the 6-elements ring and final NH_3 release (turquoise pathway). The second option sees the proton on the 5-elements ring transfer to the NH_2 and the following NH_3 release (blue pathway). Both NH_3 losses proceed with no TS. The energies reported correspond to the total electronic energy + the zero point energy (ZPE) correction.

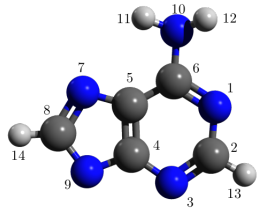


Figure 8.7: Minimum energy structure of $[Ad-H]^-$ most stable isomer as calculated at B3LYP/6-311++g(d,p) level of theory. The nomenclature attributed to the atoms in the molecule is shown.

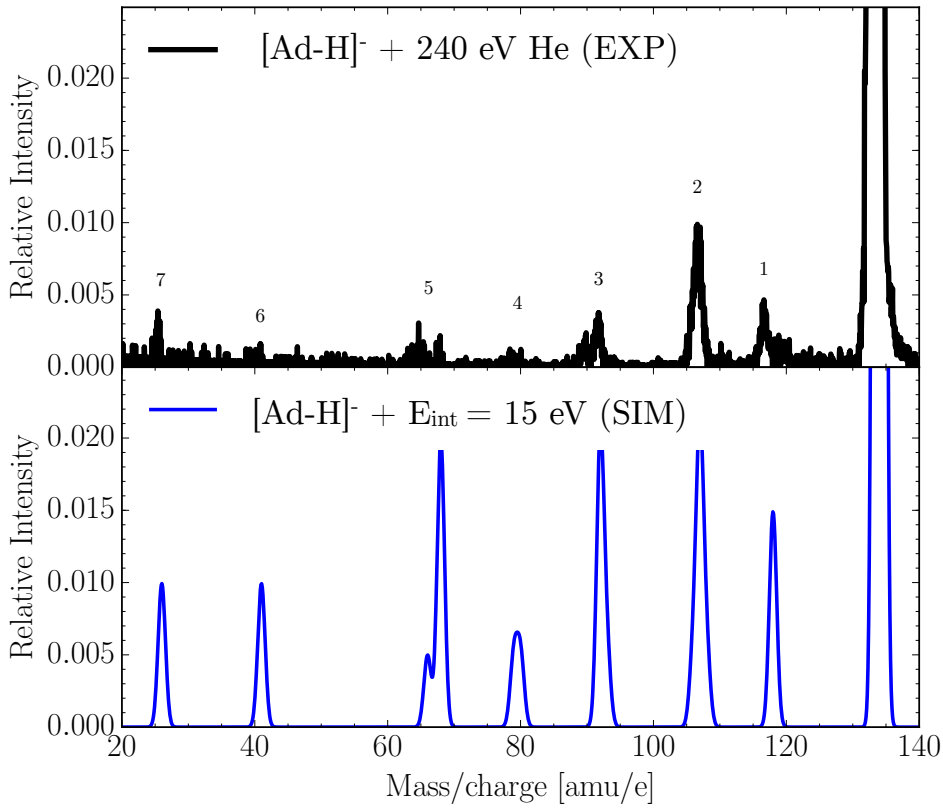


Figure 8.8: Experimental mass spectrum (top panel, black) for $[Ad-H]^- + He$ at E_{CoM} = 240 eV and simulated mass distribution (bottom panel, blue) from the ADMP trajectories at NKE (= E_{int}) = 15 eV. All the clusters of peaks have been numbered.

Following the same scheme as the one presented for the deprotonated adenine, we show a comparison between the experimental mass spectrum for $[Ad-H]^-$ in collision with He atom (top panel in black, Figure 8.8) at 240 eV of E_{CoM} and the simulated mass spectrum from the ADMP dynamics (bottom panel in blue, Figure 8.8) at NKE (E_{int}) = 15 eV. The same peaks appear in both spectra, suggesting that we can use the results from the simulation to assign the different peaks and to gain insights on the fragmentation dynamics. The peak assignments are summarized in Table 8.3. The peaks corresponding to the HCN loss (m/z = 107) and NCHNH (m/z =92) [48] /HNCNH (m/z = 92) [39], have been already identified in works by other groups. Other peaks may be assigned considering precedent publications on electron impact and [49, 53] and collision induced dissociation [50, 52] or photoionization [51] experiments on protonated and radical cationic adenine.

Table 8.3: All the m/z collected in the simulations are reported with the corresponding fragment peak detected. In parenthesis the corresponding formula or the neutral loss is evidenced.

Fragment Peak	m/z [a.u./e] (SIM)
1	118 ($-NH_2$)
2	108 ($-CN$)
	107 ($-HCN$)
	106 ($-H_2CN$)
3	93 ($-HCN_2$)
	92 ($-HNCNH$)
4	80 ($-2HCN$)
	79 (107 $-H_2CN$)
5	68 ($H_2C_2N_3^-$)
	66 ($H_2C_3N_2^-$)
6	41 (HCN_2^-)
7	26 (CN^-)

Despite the similarities in the peaks position, the branching ratio differ very much for reasons already mentioned for the protonated case.

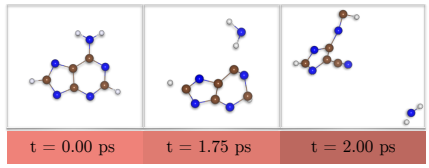
The pathways suggested by the dynamics performed are collected in Figure 8.9. According to the computational results, the 107 and 92 a.u./e correspond to the loss of HCN and HNCNH as in [39]. The dynamic formation pathway displayed here for 107 a.u./e differs from the statical pathway proposed in [39], which goes through the formation of a different isomer. More specifically, HCN is released as a consequence of a H11-transfer from N10 to N7, the N7 - C8 and C4 - N9 bond breakings. Finally, the pyrimidinic ring isomerizes to a branched 5-elements ring. This is not in contrast with what proposed by Cole et al., it just introduces a new fragmentation pathway which leads to a same m/z fragment, but to a different molecular structure. This pathway may be in competition with the one already discovered or, considering the present NKE, accessible just at higher collisional energies. As a matter of fact, the present center-of-mass energies are higher than the ones in [39], where just 107 and 92 a.u./e are detected. Same considerations should be applied to the case of the 92 a.u. formation pathway shown in Figure 8.9. Here, an isomerization of the initial species, followed by N1 - C6 cleavage, leads to the $HNCNH$ neutral loss.

The 79 a.u./e and 26 a.u./e formation pathways are also proposed by the dynamics. The former comes from a breaking down and isomerization of the initial molecule, which ultimately leads to the loss of H_2CN and HCN. The latter is the result of CN^- formation from both rings opening. The resulting neutral species later on will lose neutral HNC.

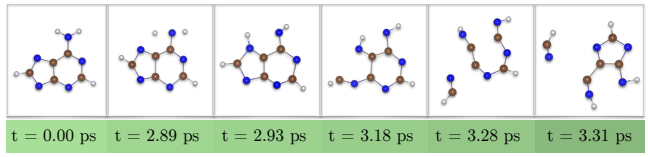
In addition, the 68 a.u./e fragment comes from the H12-transfer to N1 and following N1-C6 bond cleavage. The next steps are the C5-C4 and C8-N7 bond breakings to form the $C_2H_2N_3^-$ fragment from the neutral loss of $C_3H_2N_2$. Whereas, the 66 a.u./e fragment is formed when H11 is transferred to N7. This induces the concerted release of HCN (H14-C8-N9) from the 5-elements ring.

If we focus on the two peaks at 117/118 a.u./e in the experimental mass spectrum, this find a correspondence in the dynamics at 15 eV initial nuclear kinetic energy with a fragment of mass 107 a.u. (Figure 8.9). The mentioned fragment is the result of the neutral loss of NH_2 radical. The barriers for this loss and the possible losses of neutral NH_3 from the $[Ad-H]^-$ most stable isomer considered so far have been found with the aim of relaxed scans of coordinates and TS searches.

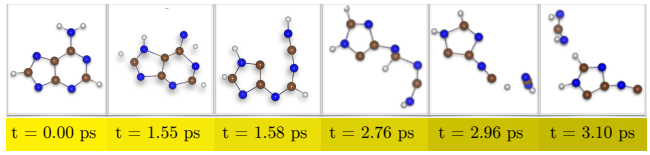
$C_5H_4N_5^- \rightarrow C_5H_2N_4^-$ (118 a.u./e) + NH_2



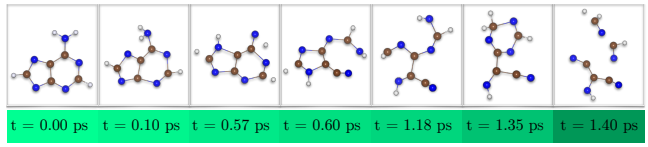
$C_5H_4N_5^- \rightarrow C_4H_3N_4^-$ (107 a.u./e) + HCN



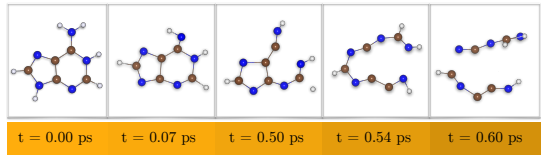
$C_5H_4N_5^- \rightarrow C_4H_2N_3^-$ (92 a.u./e) + HNCNH



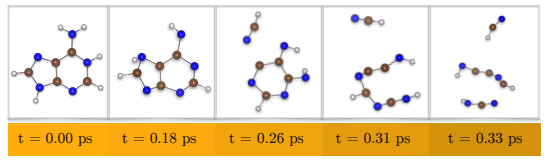
$C_5H_4N_5^- \rightarrow C_3HN_3^-$ (79 a.u./e) + HCN + H_2CN



$C_5H_4N_5^- \rightarrow C_2H_2N_3^-$ (68 a.u./e) + $C_3H_2N_2$



$C_5H_4N_5^- \rightarrow C_3H_2N_2^-$ (66 a.u./e) + HCN + HNCN



$C_5H_4N_5^- \rightarrow CN^-$ (26 a.u./e) + HNC + $C_3H_3N_3$

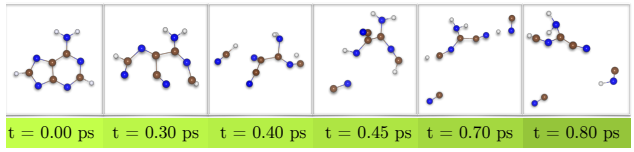


Figure 8.9: Snapshots from selected ADMP trajectories. The pathways shown lead to the formation of the negative fragments (from top to bottom): 118 a.u./e (- NH_2) , 107 a.u./e (-HCN), 92 a.u./e (-HNCNH), 79 a.u./e ($C_3N_3H^-$), 68 a.u./e ($C_2N_3H_2^-$), 66 a.u./e ($C_3N_2H_2^-$), and 26 a.u./e (CN^-).

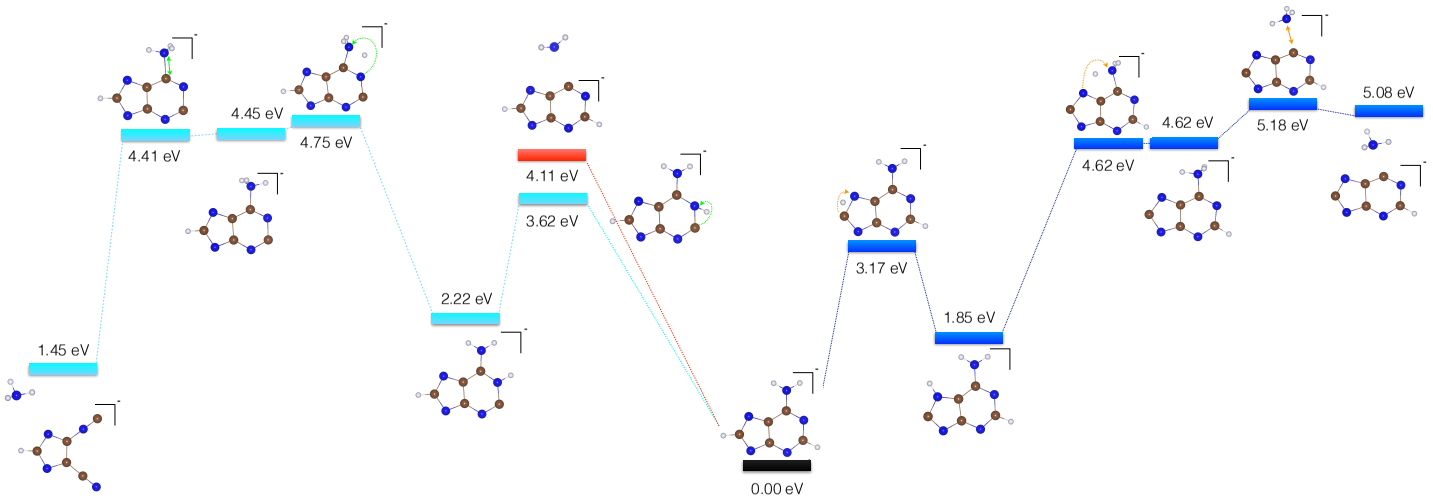


Figure 8.10: Potential energy surface section for NH_2 neutral loss (pathway in red), and NH_3 neutral loss. The NH_2 loss does not present a TS from, as in the case of $[Ad+H]^+$, but the energy of the fragments is 1.12 eV lower for $[Ad-H]^-$. There are two possibilities for the loss of NH_3. One option involves the isomerization through the progressive proton transfer of the hydrogen proceeding from the 6-elements ring and final NH_3 release (turquoise pathway). The second option sees the proton on the 5-elements ring transfer to the NH_2 and consequent NH_3 release (blue pathway). The energies reported correspond to the total electronic energy + the zero point energy (ZPE) correction.

The resulting part of the PES reconstructed has been sketched in Figure 8.10. The NH_2 neutral loss (in red in Figure 8.10) proceeds to the fragments through a pathway without a transition state. On the other side, starting from the ground state of $[Ad-H]^-$, it is possible to lose neutral NH_3 through two different pathways. For one, the proton migrates from the major ring to the NH_2 group and then NH_3 is released (turquoise pathway in Figure 8.10). The second possibility sees the progressive proton transfer from the smaller ring to NH_2 and following NH_3 loss. Looking at the barrier energies for the two processes (NH_2 and NH_3 losses), the NH_2 direct loss is the lowest energy path to follow. Although, the barrier is not very different from the NH_3 loss through proton transfer from the pyrimidine ring (turquoise path), 0.65 eV. Considering the energy deposited distribution in Figure 8.2, for our center-of-mass energies, in many cases the internal energy is high enough to overcome both barriers and for the mechanisms to be competitive. The presence of both peaks in the experiment corroborates this hypothesis. In principle in our simulation the initial kinetic energy of the system is also enough to overcome both barriers, but again the statistics collected might be insufficient and the simulation time too short to observe both processes. It has to be added that among the trajectories collected at NKE = 20 eV, the turquoise pathway for the NH_3 loss (Figure 8.10) turns up.

Chapter 9

Direct evidence of water mediated keto - enol tautomerisation in the mutagen 5-bromouracil?

In section 1.5 we have discussed the role of 5-bromouracil (5BrU, Figure 9.1) as a base analog and a mutagen. Even though the debate on the mechanisms of its mutagenicity remains open [67–74], already Watson and Crick looked at DNA base mispairing due to an occasional tautomerization of a base as the main cause for spontaneous mutations [63]. So far, there is no direct evidence of such an hypothesis. Here we discuss the results of our computational work and, for the sake of completeness, of the experimental work conducted by our collaborators at the AIRIBE facility in Caen University [81].

This type of experiment involves the collisions between a multiply charged ions of C^{4+} from a synchrotron source and a neutral cluster of molecular target at 36 keV in the lab frame. Even though experimental results for single and homogeneous clusters of 5BrU were already compared with the results from its hydrated clusters, we will mainly focus on the computational results that have helped explaining new features of the existing collisional mass spectrum of 5BrU hydrated cluster [181]. These features are suspected to be caused by tautomerization. In section 2.1 we have described how the

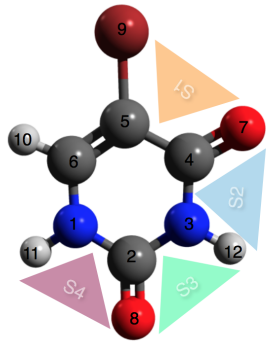


Figure 9.1: 5-bromouracil structure as obtained by minimum optimization at M06-2X/6-311++G(d,p) level of theory [109, 115, 116]. The atoms of the molecule have been numbered, and four possible hydration sites where evidenced. These zones have been named: S1, S2, S3, S4, as in [70].

neutral cluster beam is formed, this particular method does not allow the selection of a specific composition for the neutral cluster, this means we shall rather expect a distribution of cluster composition emerging from the condensation channel. Nevertheless, taking into account equation 2.1, as the mass of the target is negligible with respect to the one of the 5BrU and its clusters with water, we may reasonably assert that the center-of-mass energy is very close to the energy in the lab frame. Further details about the target preparation and experimental conditions may be found in section 2.1.

The computational work for this study was conducted with a combination of dynamic and electronic structure approaches. The dynamic approach consisted of *ab initio* molecular dynamics Born-Oppenheimer type (BOMD), with a $\Delta t = 0.05$ fs and a maximum simulation time (t_{max}) of 1 ps. The dynamics were carried out at M062x/SVP level of theory [109]. The calculations were run with the package Turbomole 7.2 [139]. We have selected two different cluster sizes to work with, $[5BrU(H_2O)_k]^{2+}$ with k = 5 and 9. Considering the size of the cluster and the t_{max} limitations imposed by the method, we have collected statistics at 6.75 and 13.5 eV of initial kinetic energy deposited and doubly charged cluster. The hypothesis that a 2+ charge may be deposited on the cluster is not to be excluded in the experimental conditions, as the C^{4+} ion may deposit multiple charge on the system, as a consequence of a double electron absorption from the cluster.

With the aid of relaxed scans of bond lengths, we have sampled the potential energy surface for the isolated 5-bromouracil molecule in its neutral, ionized and protonated states, together with the species in which 5BrU has lost one H and it is singly charged (denoted as $[5BrU - H]^+$) this form can be seen as the one arising from the lost of one proton in the doubly charged 5BrU. All the transition state and minimum search of this series of calculations were performed at M06-2X/6-311++G(d,p) level of theory [109, 115, 116] and with the Gaussian09 [91] quantum chemistry package.

9.1 C^{4+} collisions with nano-hydrated clusters of 5BrU

The present collision energies are much higher than the ones considered in the previous chapter. If we compare the order of magnitude of the present collision energy (10^5 eV) with the Figure 1.2 for the stopping power discussed in Chapter 1, we may deduct that we are operating in the energy regime of the dominance of electronic stopping, rather than nuclear stopping. As a consequence, we expect statistical dissociation processes to dominate the resulting mass spectrum. The water molecules in the cluster act as a buffer for the 5BrU molecules, so the hydrated cluster is expected to follow a statistical decay, where water molecules and 5BrU monomers are progressively lost. As already pointed out by Castrovilli et al., the presence of very intense peaks in the experimental mass spectrum of the hydrated 5BrU (Figure 9.2, blue line) corresponding to protonated water clusters and the ion $5BrU^+$ (also called M^+) seems to corroborate the hypothesis of the statistical decay dominance [181].

Figure 9.2 shows a comparison between the collision induced mass spectra of 5BrU homogeneous clusters (in red) and 5BrU hydrated clusters (in blue). In the mass spectrum of the hydrated clusters many new features, in the region both below and above the monomer [181], can be observed. Among the differences with the isolated molecule case, in both the homogenous and the hydrated cluster the protonated 5BrU peak appears (MH^+). Moreover, a series of new peaks appear (most evidently) in the region 155-255 a.u./e (Figure 9.2 b). A useful hint for the assignment of these m/z values may be given by the observation that most of the new peaks form series whose

members are displaced by 18 a.u./e, which corresponds to water atomic mass. Hence, this lead to the identification of the hydrated series:

- $[HCNH]H^+(H_2O)_n$;
- $[BrC_2O]H^+(H_2O)_n$;
- $[BrC_2N]H^+(H_2O)_n$;
- $[BrC_2NH]H^+(H_2O)_n$;
- $[BrC_2ONH_2]H^+(H_2O)_n$;
- $[M - HBr]H^+(H_2O)_n$;
- $[M - HNCO]H^+(H_2O)_n$;
- $[M - OH]H^+(H_2O)_n$.

This behaviour was not observed in precedent experiments on hydrated adenine and thymine bases [57, 182–184], although hydrated peaks were observed in previous collision induced fragmentation of uracil nano-hydrated clusters [185]. Here, it is suggested that the detection of low-intensity hydrated fragments, stable on the μs timescale, *i.e.* the time spent in the extraction region of the mass-spectrometer [57, 186], may be related to non-statistical decay. This process would involve the local energy absorption by 5BrU in the cluster, which will induce its ultra-fast fragmentation before the energy deposited is redistributed among the cluster vibrational degrees of freedom and provokes the mentioned statistical events. The presence of Br seems to increase the weight of the non-statistical events, as just one hydrated series has been observed in the uracil case [185] against the at least seven for the 5BrU case.

The mass spectrum of a cluster is characterised by broad peaks due to larger kinetic energy or successive losses in the cluster fragmentation compared to the isolated molecule, and the formation of protonated species [181]. Most of these features display not only broad, but also asymmetric shapes and shoulders, a clear indication that several unresolved fragmentation channels contribute to each peak. A deconvolution analysis was then necessary to formulate a guess on the composition of the experimental peaks.

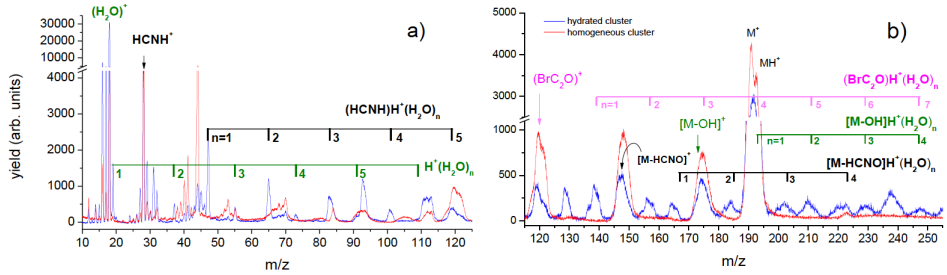


Figure 9.2: The mass spectra of 5BrU homogeneous (red) and hydrated clusters (blue) as obtained from their collisions with a C^{4+} ion beam at 36 keV in the m/z regions: 10-125 (panel a), and 115-255 (panel b). The main series of 'hydrated fragments' identified in the 5BrU hydrated cluster mass spectra are indicated by coloured vertical bars, in particular: $[M-HNCO]H^+(H_2O)_n$ in black, $[BrC_2O]H^+(H_2O)_n$ in pink and $[M-OH]H^+(H_2O)_n$ in green.

This process revealed the presence of unexpected m/z, as the consideration of "only hydrated" species would leave unassigned m/z values, which could be explained just by the inclusion of an extra H atom. In particular the following species were identified:

- $[M-HNCO]H^+(H_2O)_3 + H$ (203/205 a.u./e);
- $[M-HNCO]H^+(H_2O)_4 + H$ (221/223 a.u./e);
- $[BrC_2O]H^+(H_2O)_6 + H$ (229/231 a.u./e);
- $[M-OH]H^+(H_2O)_3 + H$ (229/231 a.u./e).

One possible explanation for the presence of these species may be that they come from an intramolecular keto-enolic isomerization of 5BrU, previous to fragmentation. The fact that these *extra-H* species are detected only in clusters with, at least, three water molecules could be taken as a direct evidence that the keto-enolic tautomerism is favoured by the presence of water.

The explanation of the origin of these peaks, where an extra H is present, has been the main objective of our theoretical calculations, which are presented in the next section. One of the key aspect of the computational work has been to elucidate if keto-enol tautomerization can explain the observed products and if alternative mechanisms can be envisaged.

9.2 Computational modeling of nano-hydrated 5BrU decay

In this intricate problem, the mass spectrum interpretation may strongly benefit from the computational modeling of such a system. The theoretical results, in fact, may provide hints on the structures of the most interesting features of the mass spectrum. In addition, their formation mechanisms may be unveiled.

9.2.1 Molecular dynamics simulations

The cluster formation process introduces some uncertainty on the composition of the hydrated clusters generated. If we consider the mass distribution of charged hydrated clusters reaching the detector (mainly $(M)H^+(H_2O)_2$ / $MH^+(H_2O)_2$) and the 'slow' statistical decay already discussed, it seems reasonable to assume that, even if big clusters of composition $(5BrU)_n(H_2O)_m$ are formed in the condensation channel, they may dissociate in smaller ones of the type $5BrU(H_2O)_k$, where $k = m - x$. If there is an excess of energy deposited, once these simpler clusters are formed, further dissociation processes may take place. This assumption is particularly convenient from the computational point of view, as the cluster composition $5BrU(H_2O)_k$ is easier to treat at high level of theory than the $(5BrU)_n(H_2O)_m$ one. To follow the statistical decay processes induced by the collision, we decided to run BOMD trajectories on doubly charged clusters with 5 and 9 water molecules: $[5BrU(H_2O)_5]^{2+}$ and $[5BrU(H_2O)_9]^{2+}$. For the $[5BrU(H_2O)_5]^{2+}$ cluster, we have considered three different initial geometries, called 4:1, 3:2, 5:0 (Figure 9.3a,b,c). Whereas a single cluster geometry is considered for $[5BrU(H_2O)_9]^{2+}$, called 4:5 (Figure 9.3d). The geometries were selected such that the water molecules are distributed in proximity of the NH to increase the probability of observing water assisted tautomerization. As a matter of fact, given the complexity of the real $(5BrU)_n(H_2O)_m$ system, it would be too ambitious to try to reproduce all the possible events induced by the ionic collisions. Here, with the MD calculations performed, we aim to have a first clue of the main chemical processes that can explain some of the interesting features of the mass spectrum. In this respect, the choice of doubly charged clusters instead of the singly charged ones is also justified

by the necessity of speeding up the simulations. In the next subsection, the PES exploration of the most probably formed singly charged species will be discussed.

Incidentally, including more than one initial geometry in the calculations improves the quality of the statistics collected, as many different distributions for the water molecules in the 5BrU hydration sphere are possible at room temperature. Although, this number is still not sufficient to allow us to make quantitative considerations on the results of the trajectories. For each geometry $\sim$ 100 trajectories were run. The choice of the double charge for the system combined with the the initial nuclear kinetic energy of choice, 6.75 eV, allows us to observe fragmentation pathways on the reasonable time scale of $t_{max} = 1$ ps (time step was fixed in the simulations to $\Delta t = 0.05$ fs). Thus, the computational cost per trajectory is maintained fairly low. Additional dynamics were collected at higher energy (13.5 eV) for $[5BrU(H_2O)_k]^{2+}$ 4:1 and $[5BrU(H_2O)_9]^{2+}$.

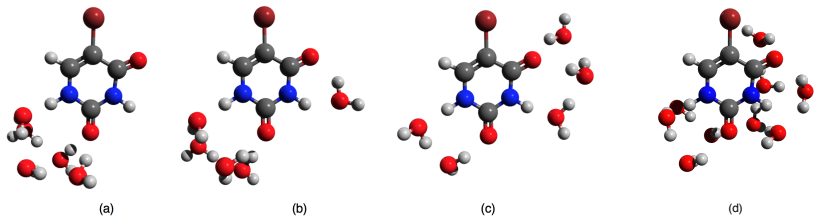


Figure 9.3: Different geometries for the $[5BrU(H_2O)_k]^{2+}$ cluster, with $k = 5$ and 9. The names chosen for the geometries of the $[5BrU(H_2O)_5]^{2+}$ reflect the distribution of the water molecules in the hydration sites of the 5-bromouracil molecule. In particular: (a) has 4 water in one S4 and 1 in S2; (b) has 3 water in S4 and 2 in S2; (c) has all 5 water in S4. The d panel shows the 4:5 $[5BrU(H_2O)_9]^{2+}$ geometry, where 4 molecules are in S4 and 5 in the S2 hydration zone.

Table 9.1 summarizes the results obtained from the dynamics of the $[5BrU(H_2O)_5]^{2+}$ and $[5BrU(H_2O)_9]^{2+}$. From the simulations, 4 different channels could be identified for the $[5BrU(H_2O)_5]^{2+}$ cluster:

- $[5BrU - H](H_2O)_n^+$, the loss of protonated water cluster;
- $[tauto - 5BrU](H_2O)_n^+$, the keto-enol tautomerization;
- $[5BrU + OH](H_2O)_n^+$, the water insertion;

Table 9.1: Statistics from the BOMD dynamics for the 4:1, 3:2 and 5:0 geometries of $[5BrU(H_2O)_5]^{2+}$ and the 4:5 of $[5BrU(H_2O)_9]^{2+}$ (Figure 9.3) at 6.75 eV of initial nuclear kinetic energy. k represents the initial number of water molecules in the system. The pedix n in the products formula is equal to the $(k-x)$ waters coordinated to the 5BrU at t_{max}, after x H_2O molecules left the cluster. The statistics have been grouped in different channels: (a) $[5BrU-H](H_2O)_n^+$ for the protonated water cluster loss; (b) $[tauto-5BrU](H_2O)_n^+$ for the tautomerization process; (c) $[5BrU+OH](H_2O)_n^+$ for the water insertion and (d) for the ring opening.

$[5BrU(H_2O)_k]^{2+}$	**% of trajectories**					
k		**5**			**9**	
Product	**geometry = 4:1**	**4:1**	**2:3**	**5:0**	**4:5**	**4:5**
NKE (eV)	**6.75**	**13.5**	**6.75**	**6.75**	**6.75**	**13.5**
$[5BrU-H](H_2O)_n^+$	93	46	36	94	47	55
$[tauto-5BrU](H_2O)_n^+$	0	0	2	0	0	0
$[5BrU+OH](H_2O)_n^+$	0	0	4	2	53	15
Ring opening	7	53	58	4	0	30

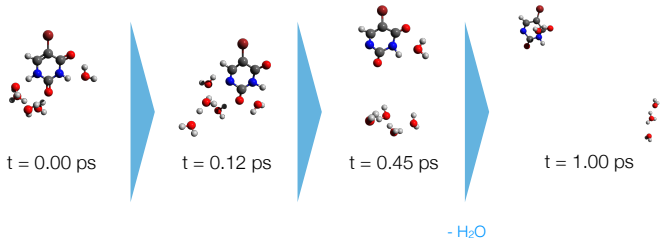
- Ring opening.

Figure 9.4 presents snapshots for a selected trajectory per channel. In the 4:1 and 5:0 clusters at 6.75 eV, the prevalent channel is the loss of a protonated cluster of water. The presence of this channel is consistent with what predicted and confirmed by the series of protonated water clusters peaks in the experimental mass spectrum. The H extracted is always the H11 and the $[5BrU-H]^+$ ring remains intact till t_{max} (Figure 9.4a). The extraction of H11 or H12 may lead to the tautomerization of the bromo-uracil, through water mediated proton transfer to the O8 (Figure 9.4b). The existence of the not observed channel that leads to the enolization of O7 is not to be excluded, as it may appear simply accumulating more statistics. The loss of a protonated water cluster is not dominant in the case of 2:3 cluster, where about half of the trajectories get to a channel generally called "ring opening". All the geometries (and energies) present this channel, where the H12 is extracted and a bond cleavage occurs on the ring. Mainly, the bond breaking occurs at the N1 − C2 bond (Figure 9.4c). Raising the energy results in increasing the occurrence of this channel, as the statistics for 4:1 and 4:5 seem to indicate. No fragments are produced in the time range of the present dynamics.

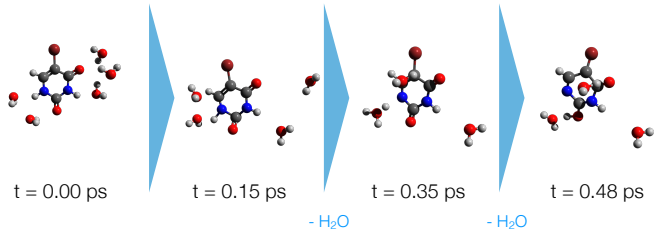
An interesting result is the formation of the $[5BrU+OH](H_2O)_n^+$. Here, once the water cluster has left with H11, one of the water molecules produces a nucleophilic addition to the C5. A new covalent bond is formed between O_w (the water oxygen) and C5, with migration of H on the neighbour O7 (Figure 9.4d).

The series of statistics ($\sim$ 100 trajectories) collected for $[5BrU(H_2O)_9]^{2+}$ (Figure 9.3d) at NKE = 6.75 eV registered just the protonated water cluster loss and the water insertion as channels. Table 9.2 reports some interesting data on the $[5BrU-H](H_2O)_n^+$ and $[5BrU+OH](H_2O)_{n-1}^+$ products (with n = 0 - 9) obtained from these dynamics.

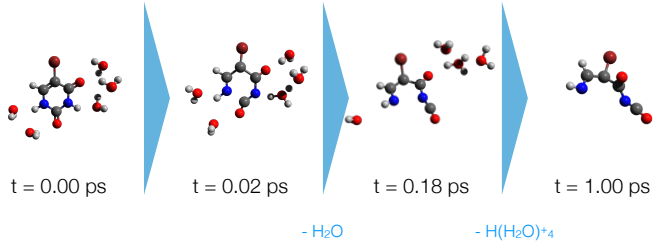
The extra-H hydrated series has a number of water molecules ≥ 3, this seems to suggest there is a threshold number of H_2O in the hydration sphere at which the OH-inserted product are favoured with respect to the H-loss products. Nevertheless, the relative energies calculated clearly indicate that the OH-inserted products are always the most stable ones (ΔE_{el} in Table 9.2). However, it is true that the higher the number of water molecules, the



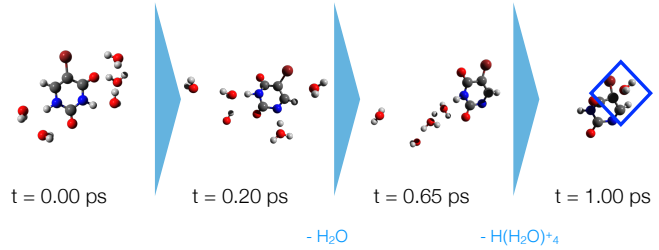
(a) $[5BrU - H](H_2O)_n^+$ formation.



(b) $[tauto - 5BrU](H_2O)_n^+$ formation.



(c) Ring opening with N1-C2 bond cleavage.



(d) $[5BrU + OH](H_2O)_n^+$ formation

Figure 9.4: Snapshots from selected dynamics. An example of each of the different channels observed in the dynamics is reported.

Table 9.2: Statistics from the BOMD dynamics for the of $[5BrU(H_2O)_9]^{2+}$ cluster (Figure 9.3) at 6.75 eV of initial nuclear kinetic energy. (+OH) stands for $[5BrU + OH]^+$ and (-H) for $[5BrU - H]^+$. n is the number of water molecules coordinating the products (# H_2O, which is equal to $k - x$). The (+OH) always has $n - 1$ water around, because one has added up to the 5BrU. The percentage of $[5BrU - H]^+$ and $[5BrU + OH]^+$ products with 1-9 water molecules in the first hydration sphere is reported. ΔE_{el} is the difference between the minimum E_{el} including the zero-point correction energy of the two products ($E_{el}(+OH) - E_{el}(-H)$).

$[5BrU(H_2O)_9]^{2+}$	**% of trajectories**		
n = # H_2O	$[(-H)(n)]^+$	$[(+OH)(n-1)]^+$	ΔE_{el} (eV)
0	10	–	–
1	11	10	1.72
2	6	14	2.09
3	4	13	2.40
4	6	13	2.66
5	4	1	2.66
6	0	2	3.63
7	0	0	–
8	4	0	–
9	2	0	–

higher is the energy gap between 5BrU-H and 5BrU+OH. In addition, the statistics is not only unsuitable to draw quantitative conclusions, but also biased by the initial structure of the cluster and its composition, and - last but not least - by the composition of the cluster after the $H(H_2O)_x^+$ release. It is reasonable to assume, in fact, that initial structures with water closer to the C5 are more likely to suffer nucleophilic addition by water. This is the case for the 3:2 $[5BrU(H_2O)_5]^{2+}$ and $[5BrU(H_2O)_9]^{2+}$ clusters, for example. However, there will be less possibility for the water molecule to be close to the C-Br in the case of 1 water molecule remaining attached to the cluster than in the one of 2, and for 2 less than for 3 and so on. As a final remark, we need to add that the final assumption is reasonable not only statistically, but also physically. In fact, the most hydrophilic sites for the 5BrU molecule are the S2, S3, S4 (Figure 9.1), which are the most polar and the ones with less steric hindrance. This considered, it will be more likely for the water to be located in the surrounding of the C5 when more than 1 water molecule is in the cluster, such that the extra water molecule(s) would

occupy the hydration zones mentioned and a water molecule would be free to attack the C5.

9.2.2 PES sampling

Once obtained the product $[5BrU + OH]^+$ we started digging into the energetics of the process. The effects of depositing charge and extracting an H from the system are explored in Figure 9.5. Here, the results from the intrinsic reaction coordinate (IRC) of the water insertion on the 5-bromouracil are shown. The left column refers to the addition on the neutral intact 5BrU, whereas the right column shows the insertion on the $[5BrU - H]^+$, the one obtained from the dynamics. Top row displays results for the 5BrU with just one water molecule in the cluster; instead, the reactant is surrounded by two extra water molecules in the bottom row case. From this figure, it is possible to appreciate that the presence of extra coordinating water molecules does not stabilise the transition state, thus it does not change the energy barrier or influence in any way the mechanism of reaction. The loss of H and the presence of a positive charge change dramatically the mechanism of the reaction, if we look from left to right in Figure 9.5. As a matter of fact, on the left of Figure 9.5 we are assisting at a nucleophilic substitution reaction. Here, water (the nucleophile) attacks the electrophile site (C5), transfers the H to Br, transforming it in a better leaving group. HBr molecule leaves the complex and an enol with a planar ring is left. On the right, the situation is completely different. The addition of the nucleophile does not result in the H-transfer to Br, but to O7. The binding situation of C5 is then changed and the planarity is lost. In this case, the whole process is a barrier-less process. It has to be added that an effect of the increasing number of water molecules in the hydration zone is registered in relation to the inserted product. In fact, the more water molecules it has, the more it is stabilized with respect to the reactant, the $[5BrU - H]^+$. This effect is visible in Table 9.2.

Finally, the water addition effect was analysed also for the ionized and protonated species. This last one existing as two isomers: one protonated on the O7 (5BrUH+) and one on the O8 (5BrUH+_O8). This investigation is motivated from the fact that the simply charged (5BrU+) and the protonated species (5BrUH+ and 5BrUH+_O8) are the parent ions in the experimental spectrum, then we aimed to discover if they could suffer a water insertion

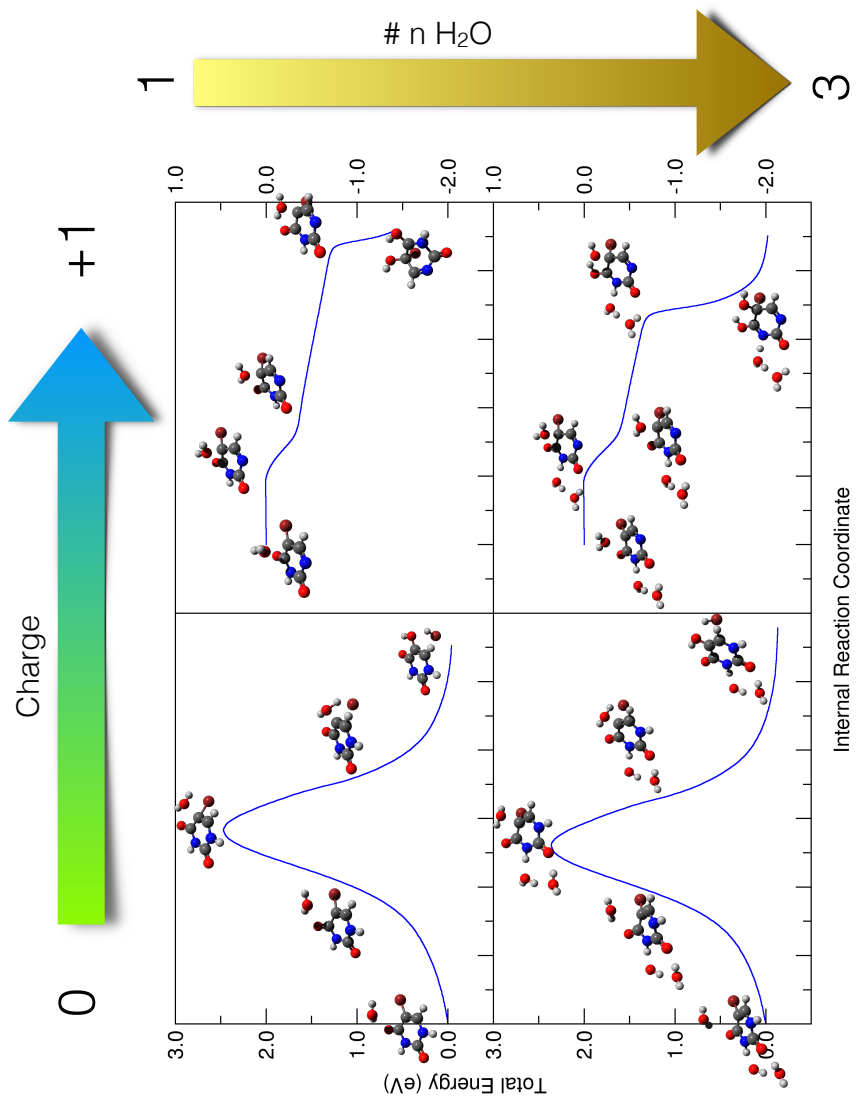


Figure 9.5: IRC results, where the total energy (eV) is reported with respect to the intrinsic reaction coordinate. Top row presents the bromo-uracil with one water inserting; in the bottom row 3 waters surround the 5BrU. The left column refers to the neutral intact 5BrU, whereas the $[5BrU - H]^+$ species is subject to the water attack in the left column.

reaction too. A comparison of the pathways obtained is displayed in Figure 9.6. Interestingly a substitution of Br for water happens with a barrier of 1.12 eV in the $5BrU+$ case. The substituted product appear at mass 129 a.u./e.

A 128 a.u./e fragment corresponding to the neutral loss of HBr may come from the nucleophilic substitution acting on the neutral (barrier of 2.5 eV) and both the protonated species (barriers of 3.46 and 3.04 eV, respectively for 5BrUH+ and 5BrUH+_O8). A series of peaks corresponding to this mass/charge with number of H_2O between 1 and 6 has been detected. It has to be added that 128 a.u. may refer to the cluster $[[M - HBr](H_2O)]^+$, as $[M - HBr]^+$ has been already identified in the monomer and homogenous cluster mass spectra [181], and a 110 a.u./e assigned to $[M - HBr]^+$ has been found in the present mass spectrum too. Thus, in the hydrated cluster mass spectrum the nucleophilic substitution product may either compete with the HBr loss mechanism guessed for the monomer and the homogeneous cluster series [181] or be the original cause of this specific hydrated fragment series 5BrU is in a water environment.

The discovery of this inserted stabilised structure raised the question on if the extra-H peaks identified in the mass spectrum could come from the fragmentation of this species. So by means of bond lengths relaxed scans and TS searches, we have been investigating the potential energy surface of $[5BrU + OH]^+$, aiming to obtain the interesting extra-H peaks described at the end of the previous section. This sampling found the path in Figure 9.7 for the formation of the $BrC_2O + H$ fragment, responsible for one of the extra-H peaks. The remaining two could not be explained by this mechanism.

To try to explain these extra-H peaks and to verify the experimental hypotheses that initial tautomerization of BrU is needed to explain them, we have done a systematic study of different PES. As the hydrated fragment series come from a local depositing of energy on the sole 5-bromouracil molecule, it seems reasonable to try to obtain the path to these fragments from the neutral ($5BrU$), the ionized ($5BrU+$) and the protonated forms of the 5BrU. As a matter of fact, neutral fragments originating from the neutral 5BrU may be detected when in a protonated water cluster. Whereas the simply charged ($5BrU+$) and the protonated species ($5BrUH+$ and

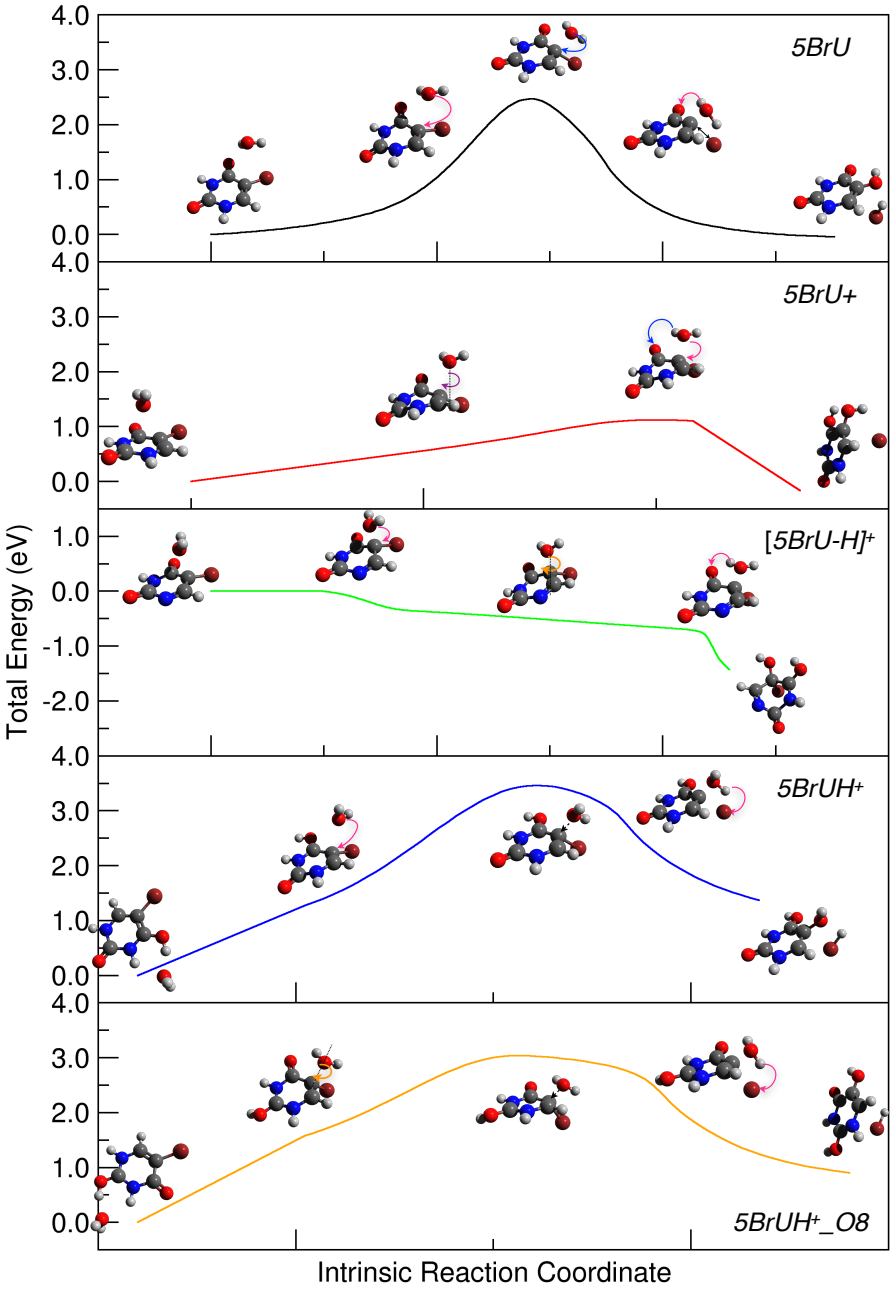


Figure 9.6: IRC profiles for H_2O nucleophilic attack to (from top to bottom): neutral 5BrU, ionized 5BrU, $(5BrU - H)^+$, 5BrUH+, 5BrUH+_O8.

$5BrUH + _O8$) are both parent ions produced by the particular ionizing method. $[5BrU - H]^+$ was the major product from the MD calculations and therefore its contribution was also considered.

Similarly to the case of $[5BrU + OH]^+$, all the bond length scans and TS searches were oriented to find the extra-H products, so the isomers that would not possibly lead to our fragments were excluded from the PES sampling.

In this case the water molecules were not included in the PES sampling as not directly involved in the reaction mechanism. Thus, it was assumed they would have not significantly influenced the bond breaking barriers.

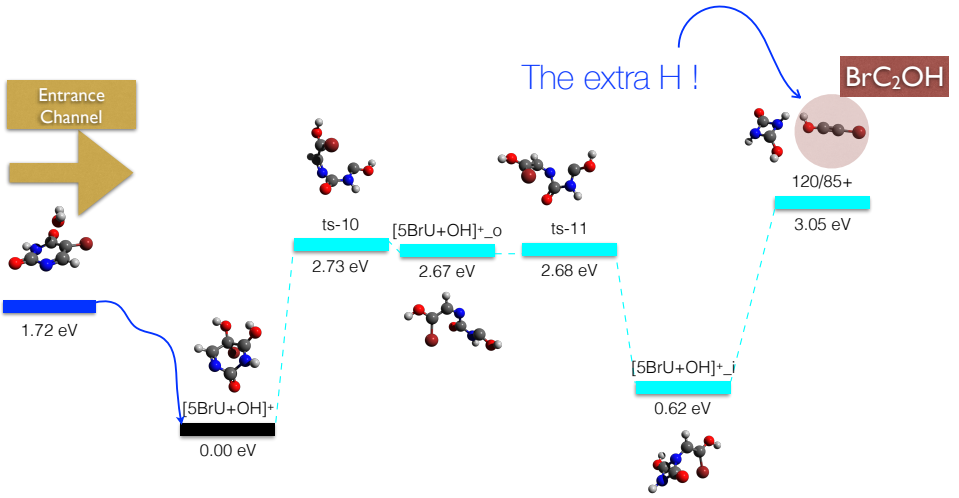


Figure 9.7: Formation pathway for the $BrC_2O + H$ fragment starting from the $[5BrU + OH]^+$ product.

The PES obtained are presented in Figures 9.8 (neutral 5BrU), 9.9 (ionized 5BrU), 9.10 (O7-protonated 5Bru), 9.12 (O8-protonated 5Bru) and 9.3 ($[5BrU - H]^+$). In order to provide a more guided analysis of the PES presented, Table 9.3 collects the m/z detected in the experiments and their corresponding formula as predicted by the different PES.

In addition, their minimized energies are collected in correspondence of the species from which they are originated, the barrier energies (if present) are in parenthesis. The energies are reported in eV. The red font for the energies indicates that the specific fragment comes from the enolic form, else

it means that it comes from the keto form. When * accompanies an m/z it means we are reporting the mass of a neutral fragment which has been detected in the experiments together with a protonated water cluster. Thus, the actual m/z detected will be the mass of the neutral fragment +1 a.u./e (proton) +18 a.u./e $*n$, where n is the number of waters in the charged cluster.

From our calculations, the lowest pathways for the formation of the peaks BrC_2O^+ and $[M-H_2NCO]^+$ pass through the $[5BrU-H]^+$ species (Figure 9.12) with barriers of 2.26 and 1.80 eV, respectively. Higher energy neutral BrC_2O and $[M-HNCO]$ (4.09 and 2.49 eV) may be obtained from the neutral compound. The same pathway of BrC_2O would also generate HCNH neutral, another hydrated peak detected. These three species belong to the series of neutral residues which need a protonated water cluster to be detected. The tautomers of the neutral 5BrU just gave isomerized products, but not interesting fragments.

On the other end, a low-energy channel for the formation of the $[M-HNCO]^+ + H$ is obtained from $5BrU+$. The fragment has 3.10 eV minimum energy (evidenced in blue, Figure 9.9) with 1.80 eV as the highest barrier to overcome on the pathway found. This mechanism proceeds through a keto-enol tautomerization, specifically the one that involves the H12 transfer from N3 to O8, and subsequent loss of neutral NCO. The tautomerization barrier is the limiting step of the pathway, although the second highest barrier is very close to this one (1.73 eV). This pathway seems to support the hypothesis that tautomerization plays a role in this intricate problem.

The same m/z fragment may originate from the tautomerization of both the protonated species to a same product (H+_tO7 in Figures 9.10 and 9.11). The pathway from the tautomer is the same and it involves a concerted release of neutral HNCO with a barrier of 4.66 eV to reach products at 1.06 eV. Nevertheless, we register a lower barrier for the H12-transfer from N3 to O7 than for the same transfer from N3 to O8 (1.39 vs 1.81 eV). This mechanism leads to a linear open structure, opposite to the case of the same m/z coming from 5BrU+, where the fragment is a four elements ring.

The $[M-HNCO]^+$ is also reported and it may be generated from $5BrU+$ from a barrier-less release of neutral HNCO. This reaction generated products at 2.65 eV. From this same molecules, when the enolic tO8-N1+ is

generated by the H11-transfer from N1 to O8, the ring opening at the level of the C5-C6 bond needs 5.91 eV to get to the isomerized 5 elements ring (iso) and then to the 147+ with the barrier-less HNCO loss. The final product has an energy of 1.88 eV. On the other hand, a fragment with the same m/z is originated from the 5BrUH+ compound from a concerted release of HNCO and H-transfer from the O7 to the N3 leaving with the HNCO residue. The barrier is of 4.22 eV here, with a 3.83 eV for the products.

The tautomerization of $5BrU+$, which involves the H12 transfer to O8 also gives the 163 a.u./e fragment (Figure 9.9) as the result of the concerted release of the H(11)N(1)C(6) neutral molecule.

To better understand the possible role played by the surrounding waters, the intramolecular H-transfer mechanism was analysed. Because of the great interest in this molecule, several studies on the keto-enol tautomerization reaction of uracil and derivates already exist. Among them, Hu et al. studied this mechanism in neutral clusters of U with a number of water molecules between 0 and 4 [70]. The interaction of water with the uracil is always favourable and the most stable clusters with $U + H_2O$ are obtained when water is in the S4 and S2. The $U(H_2O)^{S4}$ is the most stable combination; whereas for the enolic form the most stable possibility has the water in S2. Moreover, Hu's group could identify zone S1 and S2 as the ones that influence the most the tautomerization barrier. In particular, the tautomerization to the enolic form seems to be favoured when water is in S2; instead the presence of water in S1 protects uracil from the enolization. In fact, all the six different configurations considered for 1, 2 and 3 water molecules in the S1 zone have a water molecule acting as an H-donor towards the O7, such that when an H-transfer is attempted from N3, this results not very favoured as part of the electron density is already taken in the H-bond with a water molecule. On the other hand, placing 1 or more H_2O in the S2 zone they would act as a bridge for the H from N3 to O7, lowering the barrier for the transfer. If one H_2O is located in S2 and another in S1 interacting with O7 the effects of activating and protecting are balanced. A similar study was conducted on the neutral $5BrU(H_2O)$ by the same group [67]. Here, one water molecule is placed in the S1 and S2 sites. In the S2 zone water may act as an H-bond donor (to O7) and weak H-bond acceptor (from N3). Given the total charge density surfaces of water and the keto and enol of 5BrU, it

Table 9.3: Mass per charge ratios (m/z in a.u./e) of detected peaks in the experiments and corresponding formula as found from our computational results (Fragment). The relative energy (in eV) for each fragment found in the calculations is reported in correspondence of the species from the fragmentation of which a given fragment is originated. Numbers without parentheses indicate the energies needed to generate the fragments (*i.e.* the energy of the exit channels in figures 9.8-9.12). Numbers in red identify products whose formation pathways pass through a tautomerization. Numbers in black correspond to fragments directly obtained from the parent compound. Numbers in parentheses correspond to the highest energy barrier that needs to be overcome in the corresponding energy path. The * indicates that we are considering the mass of a neutral fragment which is detected in the experiments when in a cluster with protonated water.

m/z (a.u/e)	Fragment	$5BrU$	$5BrU+$	$5BrUH+$	$5BrUH+_O8$	$[5BrU-H]^+$	$[5BrU+OH]^+$
28*	$HCNH$	4.09					
119/121	BrC_2O^+					1.90 (2.26)	
119/121*	BrC_2O	4.09					
120/122	$BrC_2O^+ + H$		2.97 (3.11)				
120/122*	$BrC_2O + H$			5.32 (4.66)	5.32 (4.66)		3.05 (2.73)
129	$[M-Br]OH^+$		1.12				
128	$[M-HBr]OH^+$			3.46	3.04		
146/148	$[M-H_2NCO]^+$					1.10 (1.80)	
						2.19 (2.93)	
147/149	$[M-HNCO]^+$		2.65	3.83 (4.22)			
			1.88 (5.91)				
147/149*	$[M-HNCO]$	2.49 (2.77)					
148/150	$[M-HNCO]^+ + H$		3.10 (1.80)		1.06 (4.66)		
162/165	$[M-H_2CNH]H^+$			4.80			
163/165	$[M-HCN]^+$		1.64 (1.80)				

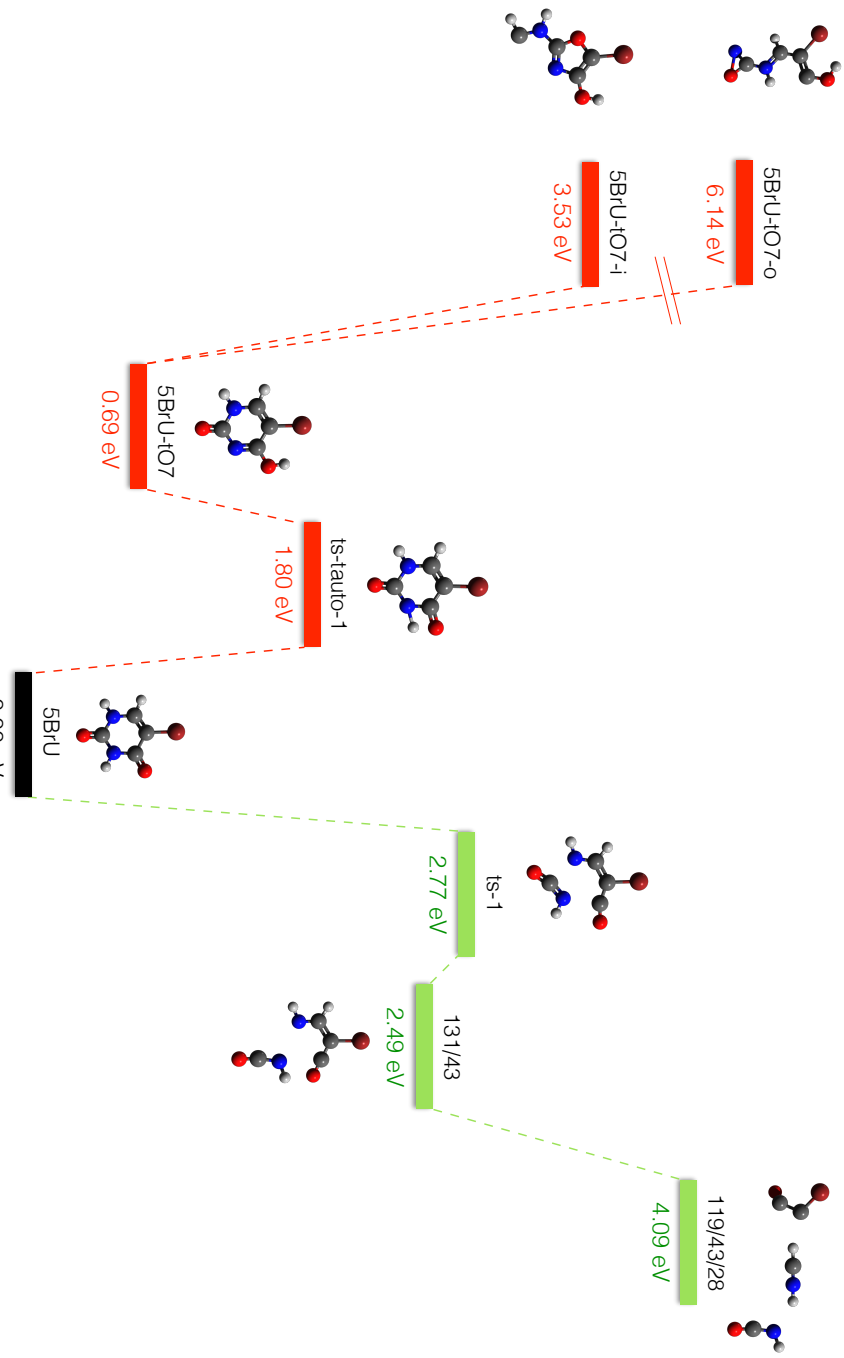


Figure 9.8: PES details for neutral 5BrU.

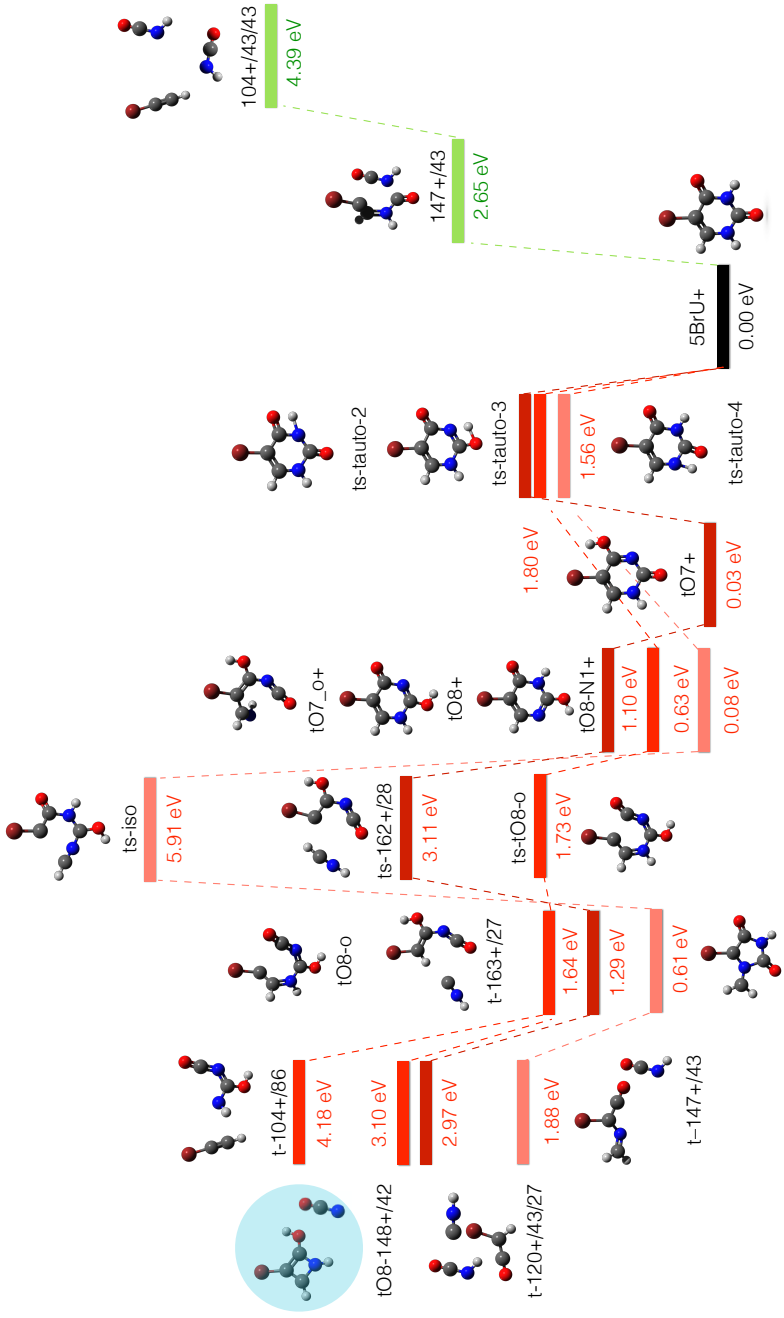


Figure 9.9: PES details for ionized 5BrU. The color code used differentiates the pathways in: the one that form fragments directly from the reactant (green), the ones that pass through a tautomerization (red). Different shades of red have been used for different tautomers.

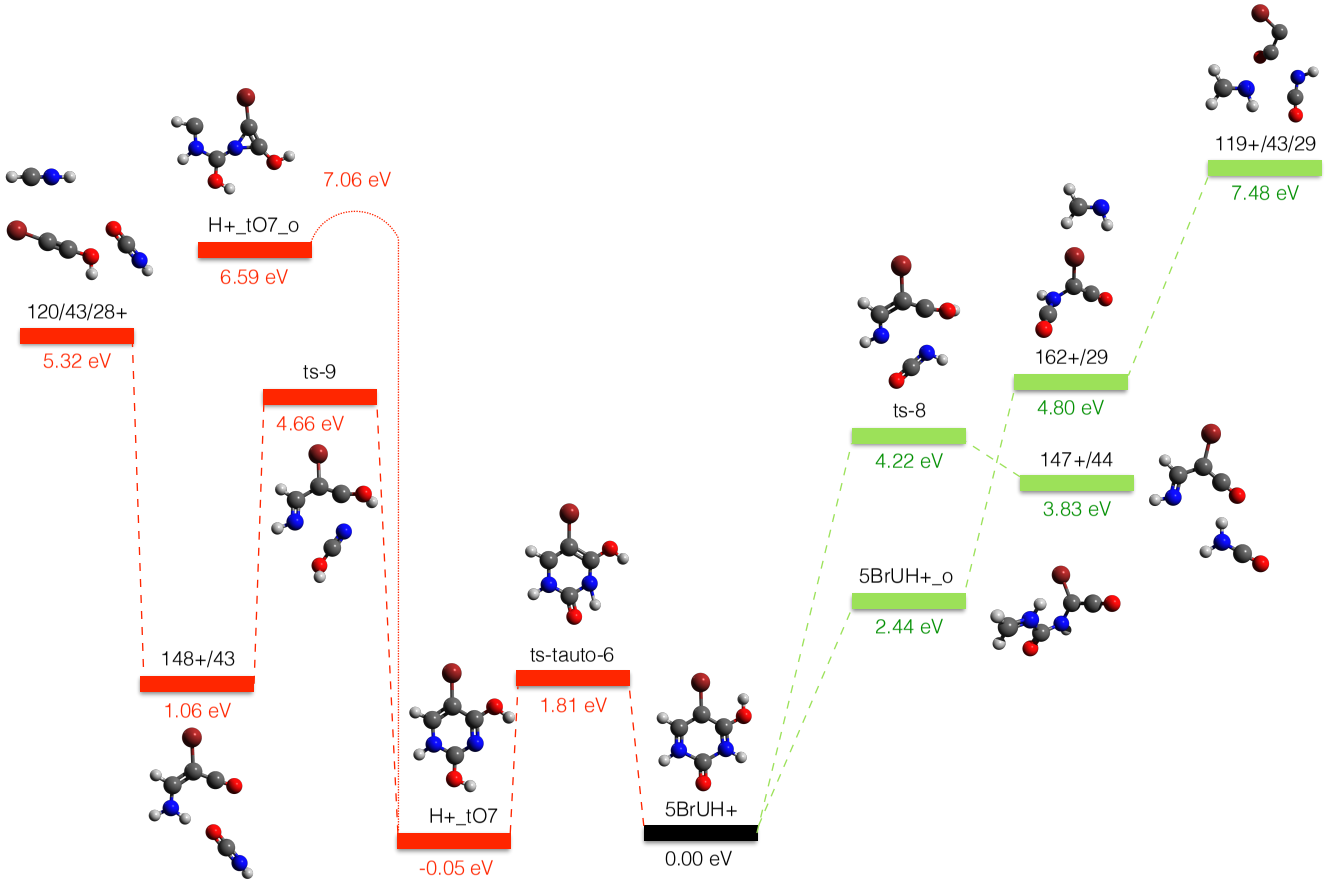


Figure 9.10: PES details for 5BrU protonated in O7. The curved line at 7.06 eV comes from the scan, as it seems to suggest the presence of a TS which was not found though.

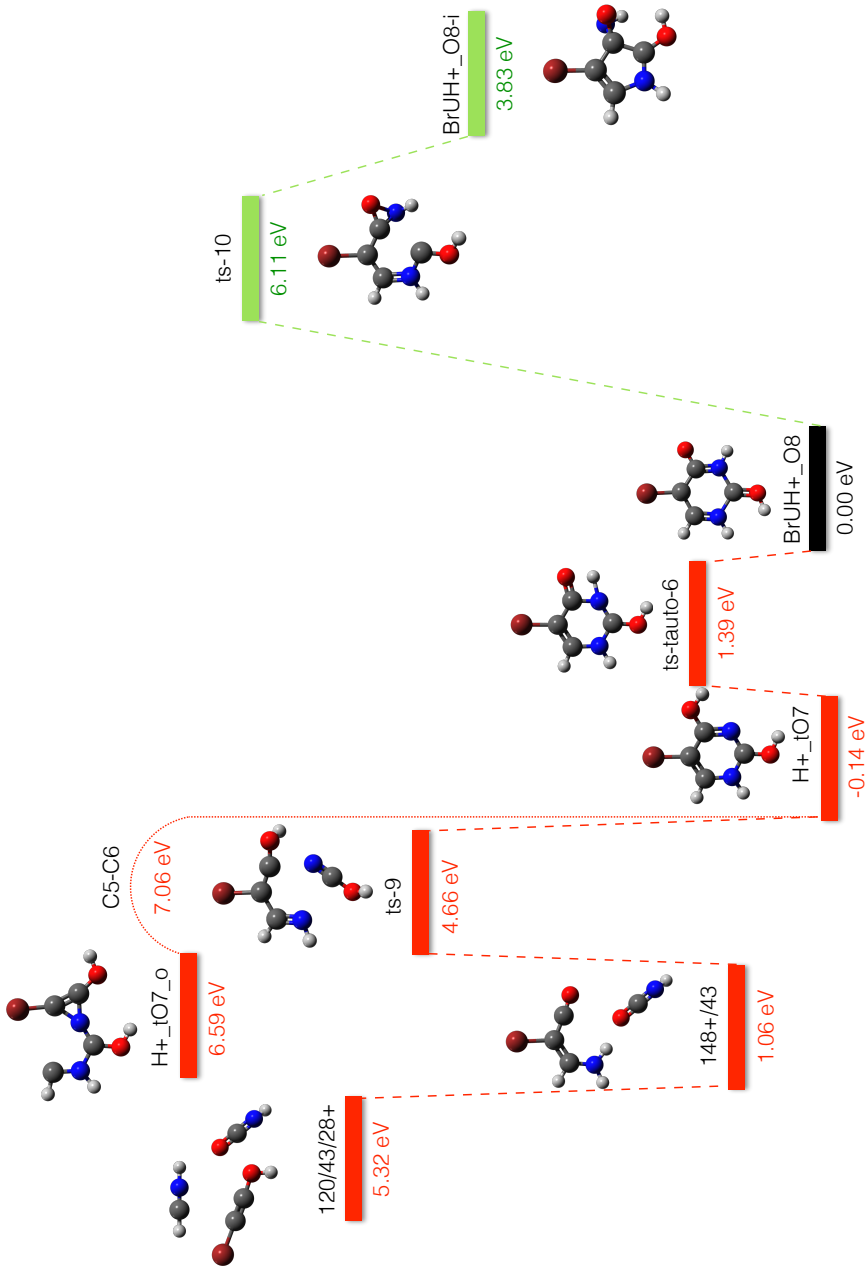


Figure 9.11: PES details for 5BrU protonated in O8.

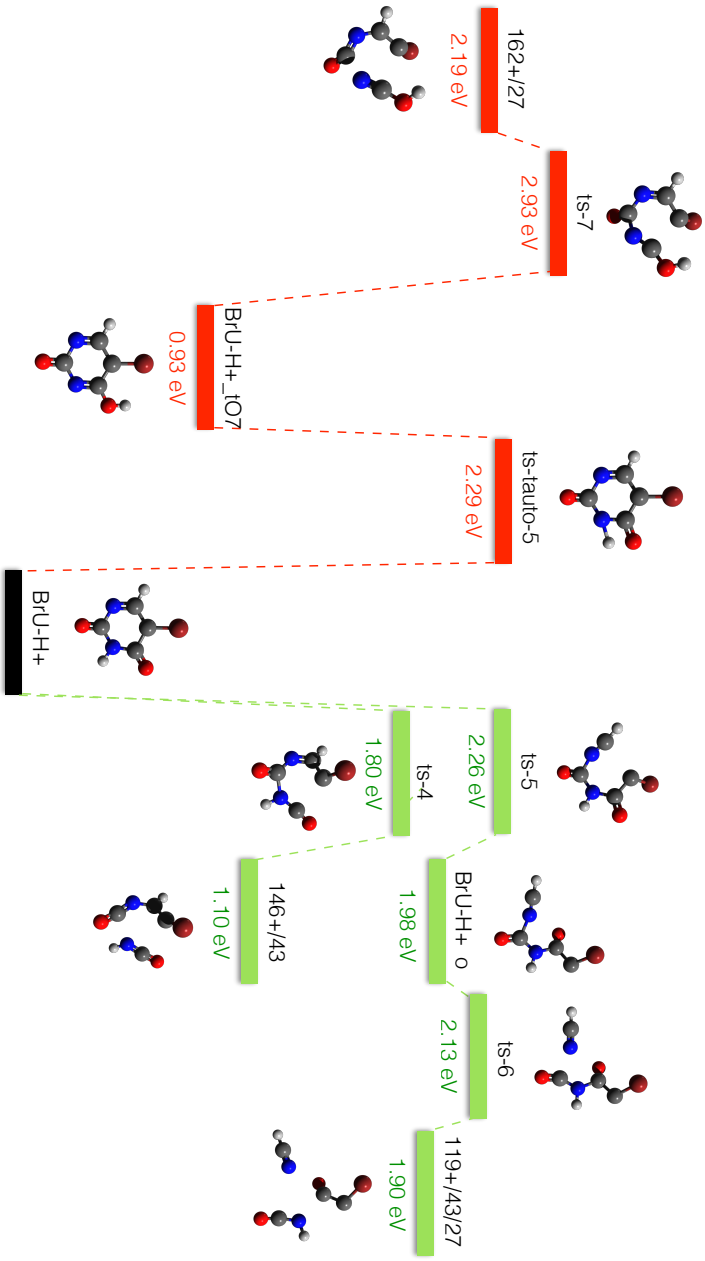
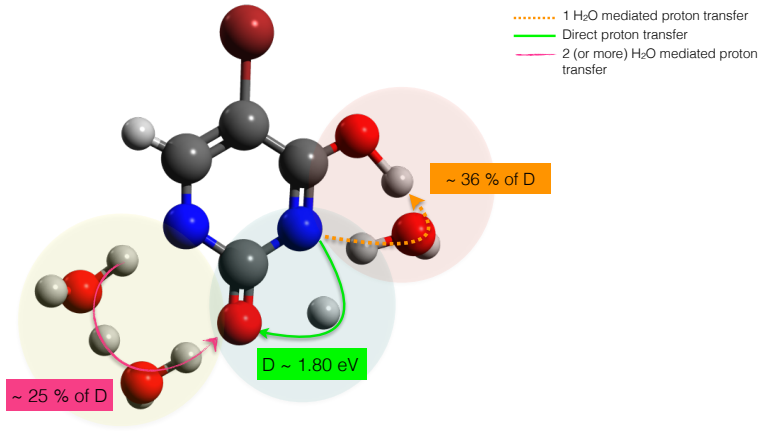


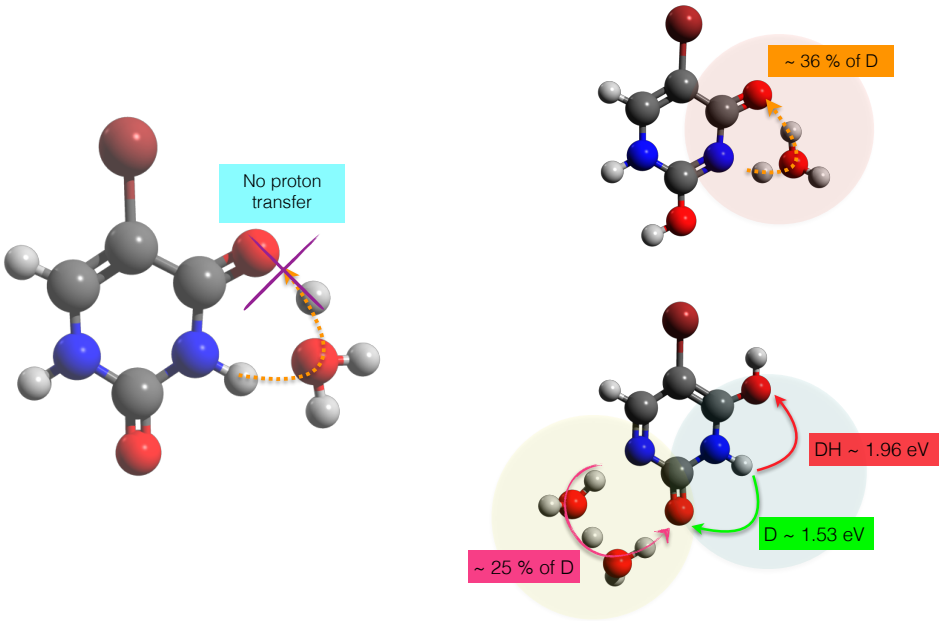
Figure 9.12: PES details for $[5BrU - H]^+$ (here BrU-H+).

may be understood that water has an higher stabilizing effect on the enol than on the keton, thanks to the $n \rightarrow \sigma^*$ interaction between the lone pair of the water oxygen and the antiperiplanar O7-H12. On the other end, the bulky, poorly electronegative and low in charge density Br is unable to form an H-bridge with water in S1. In this case the presence of bromine generates a larger negative potential zone in S1, than in the case of uracil. This has a repellant effect for the water, which hardly manages to sit in this area of the molecule. As a result, the protective effect towards the tautomerization provided from the hydration of zone S1 cancels out when H is substituted with Br in position 9.

In order to get more insights on this reaction, we have performed calculations on clusters of 5BrU with 1 to 5 H_2O coordinated. The H_2O molecules were distributed with various geometries in S1, S2, S3 and S4. The cluster was analysed in three states: neutral, charged and protonated. As already mentioned, the barriers found may lower if the surrounding water mediates the transfer. To this end different barriers for the direct and water mediated H-transfer were calculated. The largest H_2O bridge for the H-transfer had 3 water molecules. Figure 9.13 presents a scheme where the effect of the water assistance is summarised. Please, note that the scheme is a superposition of three cases: direct transfer, one water mediated transfer and 2 waters mediated transfer in just one figure, it does not refer to a $5BrU(H_2O)_3$ cluster where all the three transfers happen at the same time. The calculations showed that when the water molecules are located in other hydration zones, such that they are not directly involved in the H-transfer, they do not influence the barrier energy. The barrier for the H-transfer drops to the 36% of the initial value when one water mediates the H transfer. A pair of H_2O acting as a bridge for the H further decreases the barrier, but the effect is less dramatic than the one with the first water molecule (around 25% of the initial barrier). The TS energy reaches a plateau with the third water assisting the H transfer. Locating a water molecule in S1, such that O7 establishes an H-bridge with the water molecule, increases the H12-transfer to O7 barrier for the reasons already mentioned for the references [67, 70]. In agreement with this situation, when the proton has to be transferred to an O already involved in an H bridge or, worse, protonated, the barrier for the H-transfer to it rises. These statements are valid for the three cases.



(a) Direct and water mediated transfer scheme for charged and neutral 5BrU.



(b) Direct and water mediated transfer scheme for protonated 5BrU.

Figure 9.13: The water mediation decreases the barrier of H-transfer to the 36% of the direct transfer (D), a second water decreases it more, reaching a 25 % of D.

For the isolated molecule the keto-enol tautomerization is sensitive to the N and O involved in the process and to the state of BrU. We will now name $N3 \rightarrow O8$ the transfer of H12 from N3 to O8; $N1 \rightarrow O8$ the transfer of H11 from N1 to O8; and $N3 \rightarrow O7$ the transfer of H12 from N3 to O7. The direct H-transfer has a higher barrier for $N3 \rightarrow O8$ (1.72 eV) than for $N3 \rightarrow O7$ (1.58 eV) and $N1 \rightarrow O8$ (1.57 eV) in the case of the neutral. For the ionized 5BrU, these values become 1.92, 1.71 and 1.34 eV, respectively. The $N1 \rightarrow O8$ is the most favoured in the ionized 5BrU molecule (1.34 eV). When considering the protonated molecule, locating the proton on the O not involved in the tautomerization, the $N1 \rightarrow O8$ is the most unfavored one (1.83 eV), while $N3 \rightarrow O7$ and $N3 \rightarrow O8$ have same barrier energy to overcome (1.53 eV).

The scheme in Figure 9.13b shows that the decrease in the energy barriers when assisted with water is similar in the case of the protonated uracil. The barriers $N1 \rightarrow O8$ decreases to a 25% of the original value, while the $N3 \rightarrow O7$ to a 36%.

Chapter 10

The role of water insertion in uracil and 5-halo-substituted uracil

In chapter 9, the water attack on C5 produced a water-insertion reaction on the 5BrU or a substitution of Br by OH. Following these results, we have investigated the effects of this same nucleophilic attack in uracil and other 5-halo-subtistuted uracils (5-fluoro-uracil and 5-cloro-uracil). A scheme of the 5-X-substituted uracil molecule (5XU) is shown in Figure 10.1. As for chapter 9, the mechanisms of this same reaction with water were explored for the neutral ($5XU$), ionized ($5XU+$), protonated ($5XUH+$ and $5XUH+_O8$) and the $(5XU-H)^+$ species. The results obtained were then compared with the 5-bromouracil results.

We observed that the presence of additional waters around the 5BrU did not influence the barrier energy or the reaction mechanism, but it provided an extra-stabilization of the products with respect to the reactants (Table 9.2 and Figure 9.5). Thus, we decided to investigate the role that water molecules coordinated with the nucleophile may play in the reaction. Therefore, three cases have been considered: 1 water molecule ($1w_0hb$), 3 water molecules forming just one hydrogen bond between them ($3w_1hb$) and 3 water molecules forming two hydrogen bond between them ($3w_2hb$). The position of the water molecules in the mentioned clusters is illustrated

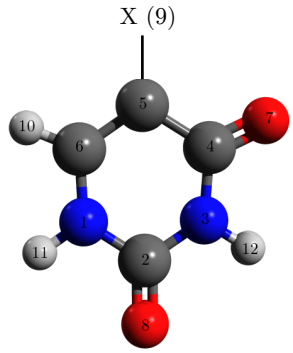


Figure 10.1: Schematic structure for 5-X-uracil (5XU), substituting X with H, F, Cl or Br one obtains the molecules explored in this chapter.

in Figure 10.2. These structures were the initial geometries used for the scans performed in all the cases explored.

In order to study these reactions, we have performed relaxed scans of the distance C5-O_w, where O is the oxygen of the water molecule acting as a nucleophile. The mentioned scans have been used to locate the transition states. Once obtained the TSs, the corresponding IRC were carried out to verify the validity of the proposed pathways. The calculations were performed at M06-2X/6-311++G(d,p) level of theory.

10.1 Is the nucleophilic addition sensitive to the halo-substitution ?

Table 10.1 presents the barrier energies (ΔE_{el}^{TS} and ΔG^{TS}) and the reaction energies (ΔE_{el}^{Pr} and ΔG^{Pr}) in total electronic energies (E_{el}) and Gibbs energies (G) for the 5XU species explored. As already pointed out in the case of the Br, the presence of the inductive electron-withdrawing halogen in position 5 seems to enhance the C5 nucleophilicity.

Apart from the effect given by the charge state, the protonation and the loss of H11 already evidenced in chapter 9, the different substitution in position 5 also influences the reaction mechanism. In Figure 10.3 we show the geometries of the transition states (grey rows) and the products of the water attack on $5XU$, $5XU+$, $5XUH+$, $5XUH+_O8$ and $[5XU-H]^+$ (with X = H, F, Cl, Br), as obtained from our electronic structure calculations.

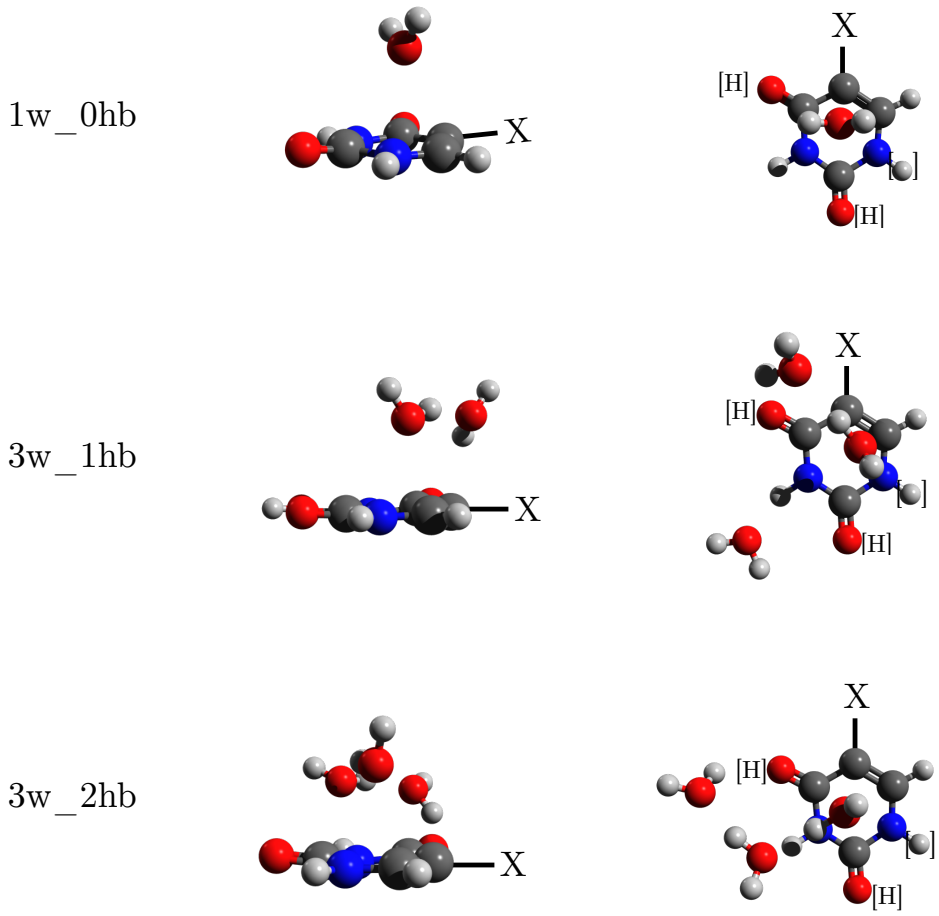


Figure 10.2: Scheme of the input geometries of the $1w_0hb$ (top row), $3w_1hb$ (middle row) and $3w_2hb$ (bottom row) 5-X-uracil (5XU) water clusters used in the relaxed scans. For each scheme two different perspectives are shown: in the molecular plane (on the left) and on top of the molecular plane (on the right). The 5XU was analysed when present in the cluster as 5XU neutral, $5XU+$, $5XUH+$, $5XUH+_O8$, $[5XU-H]^+$. The [H] indicate the position of H in the case of the two protonated species, respectively on O7 and O8 for $5XUH+$ and $5XUH+_O8$. The brackets around the H11 indicate the H removed when considering the $[5XU-H]^+$ species.

Table 10.1: Barrier energies for the nucleophilic attack of one water molecule on the $5XU$, $5XU+$, $5XUH+$, $5XUH+_O8$ (with X = H, F, Cl, Br) given as the differences between the total electronic energies and the total Gibbs energy of the transition state and the reactant, ΔE_{el}^{TS} and ΔG^{TS} respectively. ΔE_{el}^{Pr} and ΔG^{Pr}, instead, are the differences between the total electronic energies and the total Gibbs energies of the products and the reactants. Please, note that ΔG^{Pr} is equal to ΔG_r^0, which is a standard notation for this quantity. The electronic and Gibbs energies reported have been corrected for the zero-point energy. The ΔG^Y are always reported in parenthesis.

	ΔE_{el}^Y (ΔG^Y) [eV]							
Species	**5XU**		**5XU+**		**5XUH+**		**5XUH+_O8**	
	Y = TS	Pr	TS	Pr	TS	Pr	TS	Pr
X = H	3.45 (3.48)	0.503 (0.539)	0.860 (0.951)	0.240 (0.312)	3.85 (3.88)	2.56 (2.56)	3.37 (3.41)	0.719 (0.766)
F	2.79 (2.89)	1.31 (1.40)	0.795 (0.896)	-0.074 (0.005)	3.50 (3.57)	-0.008 (-0.010)	2.67 (2.75)	1.72 (1.79)
Cl	2.68 (2.76)	-0.177 (-0.174)	1.12 (1.22)	0.205 (0.288)	3.05 (3.11)	0.178 (0.131)	3.10 (3.16)	0.700 (0.332)
Br	2.53 (2.61)	-0.077 (-0.077)	1.12 (1.21)	-0.169 (-0.157)	3.46 (3.51)	0.958 (0.571)	3.04 (3.09)	0.581 (0.534)

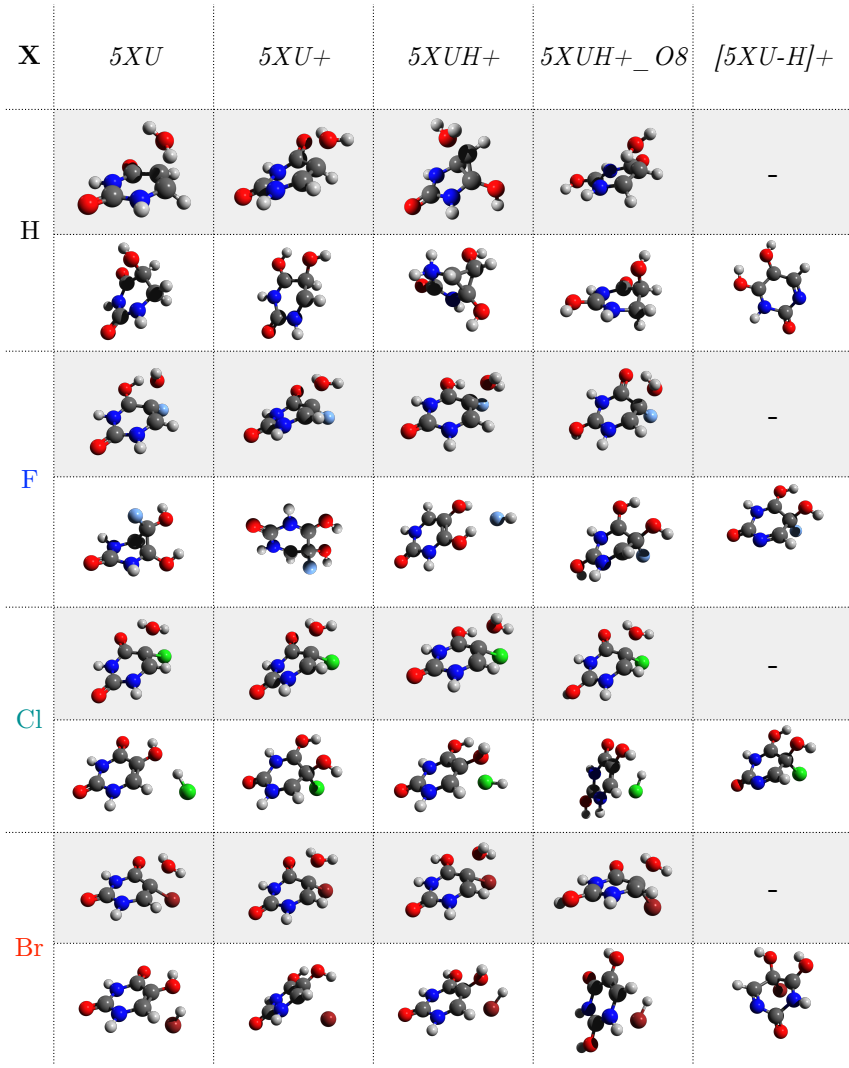


Figure 10.3: Transition state (grey cells) and product (white cells) geometries resulting from the nucleophilic attack of water on (from top to bottom): uracil; 5-fluoro-uracil; 5-cloro-uracil and 5-bromo-uracil. For each of these compounds the $5XU$, $5XU+$, $5XUH+$, $5XUH+_O8$ and $[5XU-H]^+$ states were considered.

In the case of $5XU$ neutral, the nucleophile adds up to the ring generating an insertion reaction for U and 5FU; whereas the mechanism becomes a nucleophilic substitution with release of neutral HCl and HBr in the cases of 5ClU and 5BrU, respectively, and the obtention of an alcohol. This is in agreement with the fact that Cl and Br are better leaving groups than H or F, as less electronegative and with a bigger atomic radius. In addition, the resonant electron-donating effect is predicted to be stronger in F than in Cl and Br, as the 2p of F better overlaps the 2p of C5 than the 3p of Cl or the 4p of Br (even worse). This makes the C5-F bond stronger than the equivalent bond with Cl and Br. In the case of the insertion reaction (U and 5FU) mechanisms, the final products are quite different. Surprisingly, for U the H_w^+ (the subscripted w indicates a proton proceeding from the water) is transferred to the C6, which differers from the mechanism previously observed for $[5BrU-H]^+$ in chapter 9. This H_w^+ transfer induces a distortion of the planar ring, as C5 moves from the molecular plane. The insertion proceeds through a different mechanism in the case of 5FU, as, due to the presence of F, the most electrophilic site becomes the O7, which receives the H_w^+. Here, the distortion induced from the addition of water goes even further; in fact, not only the planarity is loss, but a covalent bond is formed between C4 and C6, leading to a double ring structure. In addition, the U and 5FU molecules also produce the most unstable products in the neutral series. Therefore, when analysing the numbers given in Table 10.1, one has to keep in mind that three different products may be obtained: substitution of X by OH, as in the case of X = Cl and Br, addition of OH to C5, as in U, addition of OH to C5 and formation of a double ring structure, as in X = F. As a final remark, we shall point out that even if the barrier for the water insertion in the case of uracil is pretty high (3.45 eV) compared to the other reactions, making it less likely to happen, the water insertion barrier for 5FU is very similar to the one of X-substitution and 19% lower than the uracil water-insertion.

When adding a positive charge to the system, overall, the barriers decrease (a percentage between 55 an 75 %) with respect to the highest barrier in the neutral case. The greatest decreases are registered for U (71 %) and 5FU (75 %). As a matter of fact, the presence of the positive charge changes the mechanisms acting on the 5XU compounds. The reaction

becomes here an insertion for U, 5FU and 5ClU with the H_w^+ transferred to the O7, with a consequent distortion of the ring, similar to what observed in the $[5BrU - H]^+$ case. Thus, the positive charge seems to be delocalized between C5 and C6 and N3, as the C6 is not anymore the most electrophilic site for uracil, and O7 becomes the atom which always receives the proton, even when there is a nucleophilic substitution. In the case of 5BrU, there is a substitution reaction with the sole Br as leaving group and H_w^+ transferred to the O7. The addition barrier is low in the case of uracil and lower for F, while it rises a bit for the Cl case. The addition reaction presented in the 5ClU appears like an half-way between the 5FU and the 5BrU reaction mechanisms; in fact, the C5-Cl bond is sensibly stretched, but not broken as in the case of 5BrU, where a diol is formed, recovering the planarity of the ring in the product structure.

For the 5XU species protonated in O7, the additional H is expected to stabilize the charge as a keto group becomes an enol group, passing from electron-withdrawing to electron-donor. The reaction mechanism is a nucleophilic substitution of X per OH for all the halogens. In fact, the presence of H on O7 and the charge delocalized between C6 and N3 do not allow the H_w^+ transfer on O7 or C6. Thus, the most electrophilic site available where to transfer the H_w^+ becomes the halogen, which is then released as neutral HX. The situation is different for U, here the absence of an halogen in position 5 drives the reaction towards an H_w^+ transfer to N1, which leads to a very unstable double ring, similar to the case of neutral $5FU$. The barrier energies are higher here than in any other case. This is probably due to the limits imposed by the H on the O7, which rises the barrier of the transfer on this site and forces the water to transfer H_w^+ on the halogen, which implies an higher barrier energy.

One important consequence of the change of the protonation site from O7 to O8 is the variation of the reaction mechanism and the of the final products observed. As a matter of fact, when O8 is protonated, the positive charge is delocalized among N1-C2-N3-O8. This allows the H_2O to transfer H_w^+ on the C6 of the uracil, restoring the mechanism already observed in the neutral case for the uracil (in fact, a similar barrier is observed in both cases). For the 5-fluoro-uracil too now the H_w^+ may be transferred to the O7, which comes back as the lowest energy channel for this nucleophilic attack.

In this case, the reaction pathway does not resemble the one observed for neutral 5FU, as the double ring product is not formed. The reaction induced is a nucleophilic substitution for $5ClUH + _O8$ and $5BrUH + _O8$ with loss of neutral HCl and HBr, respectively, and diol formation.

The effects on species $[5BrU - H]^+$ were also investigated. In all cases, the TS search did not give results, and according to the relaxed scans performed, the reaction proceeds through the barrierless water insertion with OH^- addition on C5 and H_w^+ transfer on O7, and consequent formation of a product, which is more stable than the reactant. Table 10.2 reports the ΔE_{el}^{Pr} and ΔG^{Pr} for the insertion products obtained. The values evidence how the $[5BrU + OH]^+$ is the less stabilized product, -1.72 eV vs -3.35 eV (X=H), -3.69 eV (X=F), -3.31 eV (X=Cl).

Table 10.2: Total electronic energies and Gibbs energies differences between products and reactants, ΔE_{el}^{Pr} and ΔG^{Pr}, for the $[5BrU - H]^+$ species, with X = H, F, Cl, Br.

Species	ΔE_{el}^{Pr} (ΔG^{Pr}) [eV] $[5BrU - H]^+$
X = H	-3.35 (-2.89)
F	-3.69 (3.08)
Cl	-3.31 (-2.71)
Br	-1.72 (-1.04)

10.2 Water network effects

As already mentioned, in order to elucidate the role of the water environment in this reaction, we have studied systems where 1 or 2 waters are connected to the one acting as a nucleophile in the reaction.

In Figure 10.4 we show the barrier energies for the reaction with water for the $1w_0hb$ (brown), $3w_1hb$ (green) and $3w_2hb$ (yellow) for the all the $5XU$ (top left panel), $5XU+$ (top right panel), $5XUH+$ (bottom left panel) and $5XUH + _O8$ (bottom right panel) states with X = H, F, Cl, Br. Considering that the reactions described in this chapter involve an H-transfer to another atom of the 5XU system, one would expect that, as for the case of the tautomerization in chapter 9, the presence of a bridge of n water molecules would lower the H-transfer barrier. From the observation

of the barriers, we may notice that adding up extra-water molecules lowers the barrier energy with respect to the $1w_0hb$. Nonetheless, often, the barrier energy variation occurring between the $3w_2hb$ and the $3w_1hb$ clusters is not obvious. This anomaly may be explained considering also the TS structures presented in Figure 10.5. Here, the cyan rows correspond to $3w_1hb$ TS, while the yellow rows correspond to $3w_2hb$ TSs. Comparing these geometries with the one in Figure 10.3, one can realise that some of the TSs are different. For instance, the reaction products change in the neutral $5XU$ when passing to the $3w_1hb$ and $3w_2hb$ clusters. Thus, the presence of extra-water connected to the nucleophile may also influence the reaction mechanism.

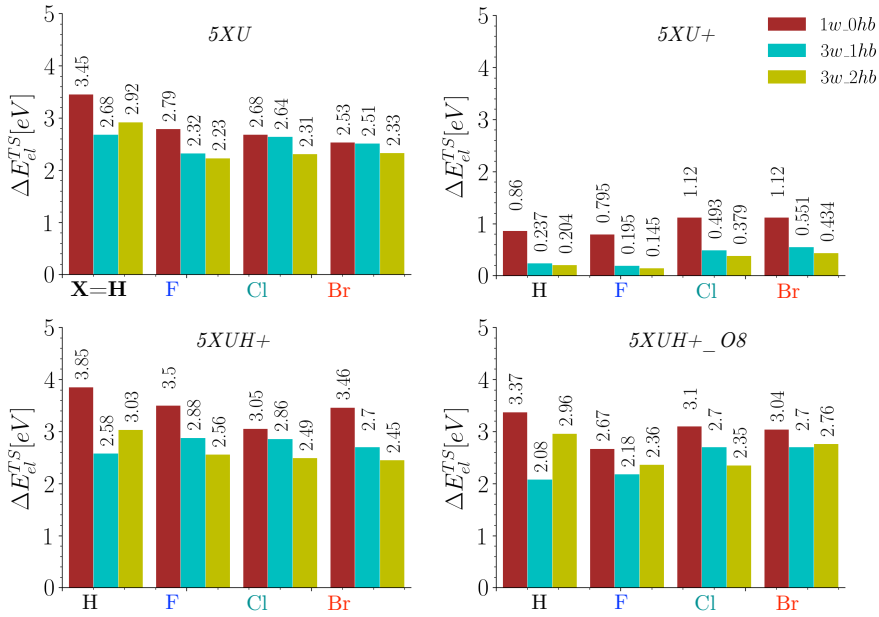


Figure 10.4: ΔE_{el}^{TS} for: $5XU$ (top left), $5XU+$ (top right), $5XUH+$ (bottom left) and $5XUH+_O8$ (bottom right) states. Each of the H, F, Cl, Br clusters contains 3 barriers: one which refers to the reaction with 1 water ($1w_0hb$, brown bar); one for the reaction with one 1 water H-bonded to the nucleophile ($3w_1hb$, cyan bar); and one which accounts for the case where 2 water molecules are bound to the nucleophile through H-bonds ($3w_2hb$, yellow bar). The labels on top of the bars show the values of ΔE_{el}^{TS} in eV.

The mentioned decrease of the H-transfer barrier would be expected to favour the H_w^+ transfer to the O7, thus, the insertion reaction with respect to the substitution one. Our results not always meet this expectation. Figure

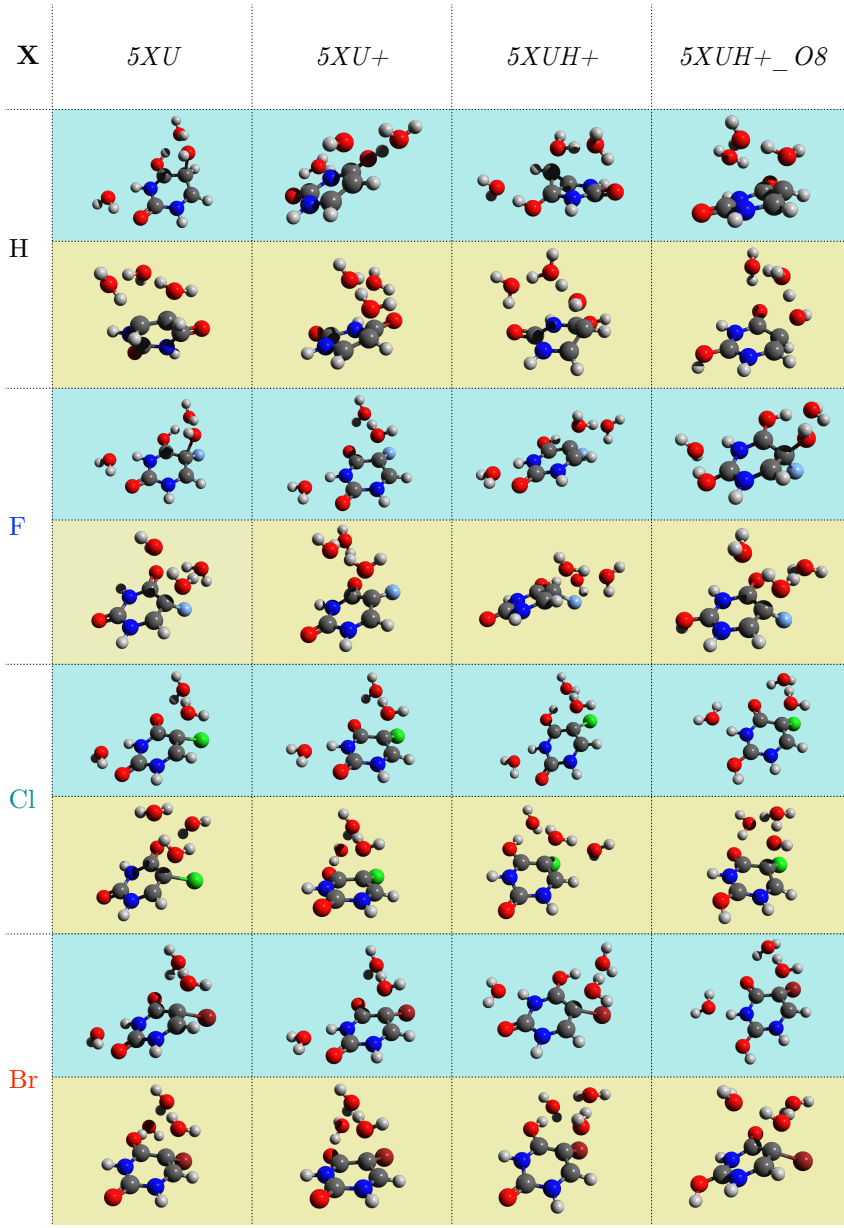


Figure 10.5: Transition states geometries for $3w_1hb$ and $3w_2hb$, as found from electronic structure calculations at M06-2X/6-311++G(d,p) level of theory. Coherently with Figure 10.4: the cyan rows refer to $3w_1hb$ TSs; and the yellow rows show $3w_2hb$ transition states.

10.7 presents the reaction products structures and it helps the understanding of the reaction mechanism. Here, the same colour code of Figure 10.5 has been used.

In the neutral uracil, for example, the presence of a water bounded to the one acting as a nucleophile does favour the H_w^+ transfer to O7, forming a double ring structure as a product. However, the presence of a third water molecule in the H-bounded network of waters, drives the reaction towards the formation of a water bridge with the C6 rather than O7, forming the same reaction product as in $1w_0hb$. The ΔE_{el}^{TS} is decreased with respect to the $1w_0hb$ case. The effect of adding 1 or two water molecules to the 5FU-water neutral cluster is to favour the substitution mechanism, in which HF is lost. In both the 5ClU and 5BrU neutral the HCl and HBr are lost.

The most dramatic changes in the reaction barriers are registered when 5XU is in a simply charged, doublet state. In particular, for the cases of the uracil and 5-fluoro-uracil, the $3w_1hb$ and $3w_2hb$ barriers decrease as much as 1/3 of the $1w_0hb$ barrier. On the other hand, the $3w_1hb$ and $3w_2hb$ barriers are about half of the $1w_0hb$ one for X = Cl and Br. As the reaction implied in all cases the transfer of H_w^+ to O7, the barrier is progressively lowered, so here the trend predicted is observed. Coherently with what already observed for the tautomerization reactions in chapter 9, the barrier lowers significantly when passing from a direct to a water mediated transfer, whereas the effect of the second water mediating is not very significant.

In $UH+$, the presence of the water network allows the transfer of H_w^+ to the O8 rather than the C6. The mentioned transfer is easier in the case of $3w_1hb$. For $5FUH+$, the presence of extra-water does not change the mechanism of the reaction, but it seems to stabilize the TS progressively lowering the barrier energy. The same happens for $5ClUH+$. Whereas, the sole Br is lost from $5BrUH+$, when in $3w_2hb$.

For $5XUH + _O8$, again when X=Cl or Br, the mechanism does not change and a decrease in the barrier is observed. For uracil the water-bridge allows the transfer to O7 rather than C6 in the $3w_1hb$ cluster, as obtained from the scan. Nevertheless, the H_w^+ transfer to C6 is restored in the $3w_2hb$ cluster. The F case is surprising, as the product appears stabilized by the H-bridges with the waters nearby, and it presents a weak interaction of one

water with C5 and a stretched C5-F bond. This structure is very similar to the TS for the reaction.

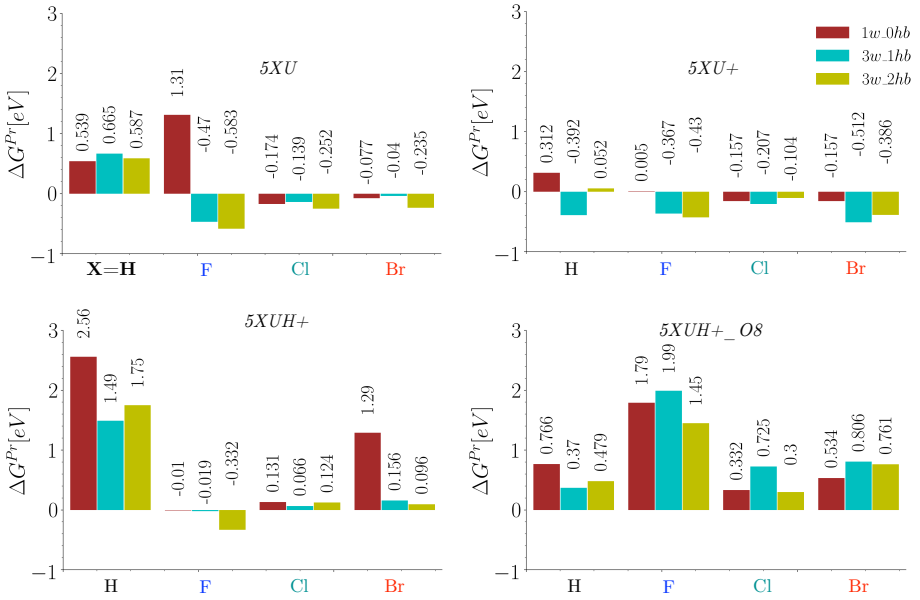


Figure 10.6: ΔG^{Pr} for: $5XU$ (top left), $5XU+$ (top right), $5XUH+$ (bottom left) and $5XUH+_O8$ (bottom right) states were considered. Each of the H, F, Cl, Br cluster contains the ΔG_r^0 for the reaction with 1 water ($1w_0hb$, brown bar) and the one with the water network ($3w_1hb$ and $3w_2hb$, cyan and yellow bars respectively).

Figure 10.6 shows the ΔG^{Pr} for the reactions in the $1w_0hb$, $3w_1hb$ and $3w_2hb$ clusters. Comparing these values with the optimized geometries of the products reported in Figure 10.7, one may notice how the most distorted structures are also the most unstable ones, whereas the inserted and substituted products appear to be stabilized by the presence of the water clusters and the formation of H-bonds with them. Finally, it is worth noticing how for the $3w_2hb$ $5XU+$ cluster the H_w^+ is not transferred to O7, but it remains in the water cluster. In this way, an alcohol rather than a diol is formed.

The presence of extra-water molecules does not affect the $[5XU-H]^+$ mechanism according to our scans. In fact, the reaction remains a barrierless water insertion, where the $[5XU+OH]^+$ product is much lower in energy than the reactant. Figure 10.8 shows the ΔG^{Pr} for the mentioned reactions. The structure of the $[5XU+OH]^+$ is the same in all cases, with OH^-

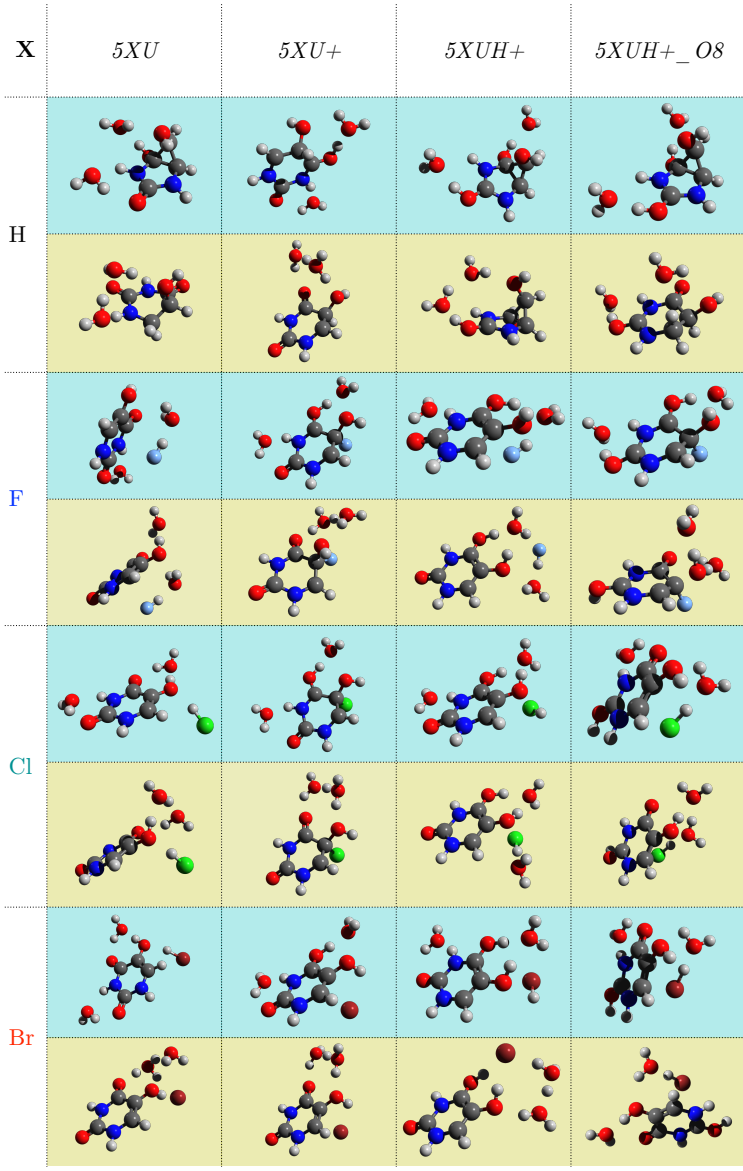


Figure 10.7: Products geometries for $3w_1hb$ and $3w_2hb$, as found from electronic structure calculations at M06-2X/6-311++G(d,p) level of theory. Coherently with Figure 10.5: the cyan rows refer to $3w_1hb$ products; and the yellow rows show $3w_2hb$ products.

added on C5 and H_w^+ transferred to O7. From the bar plot it is possible to appreciate that increasing the number of water molecule in the cluster increases the stability of the $[5XU + OH]^+$ with respect to the $[5XU - H]^+$ reactant. A trend already observed in the $5BrU - H^+$ water clusters (see Chapter 9. The highest stability of the final products is registered in the cases of the U and 5FU molecules.

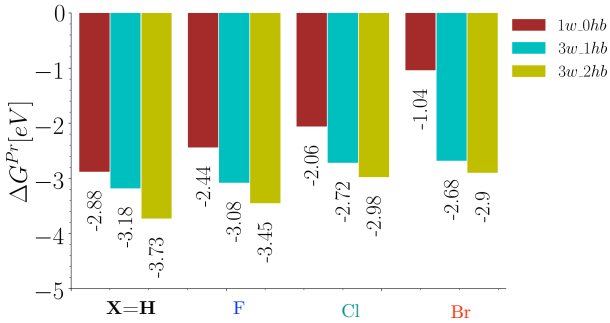


Figure 10.8: ΔG^{Pr} for the $[5XU - H]^+$ states. Each of the H, F, Cl, Br cluster contains the ΔG_r^0 for the reaction with 1 water ($1w_0hb$, brown bar) and the one with the water network ($3w_1hb$ and $3w_2hb$, cyan and yellow bars respectively).

Part IV

Final Remarks

Chapter 11

Conclusions and Perspectives

Some general considerations on the works presented may be summarized as follows:

- The knockout driven molecular growth processes shown are believed to be quite general and will most likely occur for different types of projectiles depending on the amount of energy deposited through nuclear scattering;
- Prompt atom knockout is likely to play a role also in the radiation damage of more complex biomolecules;
- The BOMD and ADMP approaches are effective to study statistical decay of molecules and clusters in the present experimental conditions;
- The environment may play a key role in the collision induced decay of biomolecules: determining the predominance of a pathway with respect to another through a catalytic action or actively reacting with the biomolecules.

More specific conclusions may also be drawn for each study:

- As for butadiene clusters, keV ions colliding with loosely bound clusters of 1,3-butadiene can induce bond-forming reactions. In fact, the fragments originated from knockout are often highly reactive and form covalent bonds with neighbouring molecules on sub-picosecond timescales. Our DFT calculations show that the most stable isomers

of the most abundant reactions products have ring structures. Further, the measured final size distribution of the reaction products can be understood in terms of their calculated relative stabilities. These findings suggest that we are indeed forming cyclic rings from initially linear molecules. The similarities between our experimental and simulated mass spectra suggest that the very fast atom knockout processes that are included in these simulations play important roles. However, it cannot be excluded that also other types of processes, such as electronic excitation processes or ionization processes also may be able to initiate molecular growth - in particular in clusters of smaller molecules. As a future perspective, it would be interesting to systematically investigate the relative importance of ionization and of electronic and nuclear stopping processes for molecular growth inside clusters of small hydrocarbon molecules.

- Prompt atom knockout was also shown to play a role in collisions between tetraphenylporphyrin ions and noble gas atoms at center-of-mass energies in the range 50 - 110 eV. Through comparisons with the results from classical molecular dynamics simulations, we have found out that about 40% and 70% of the total fragmentation cross sections are due to such processes in collisions with He and Ne, respectively. Interesting insights might turn up by the analysis of the embedding effect of such molecules in a surrounding (cluster or solvent) environment. The latter may lead to efficient molecular growth processes, which have been seen in butadiene clusters too.

- The combination of molecular dynamics and sampling of the potential energy surfaces has proven to be effective for the cases of small enough system in both gas phase (such as adenine) and in water cluster (nanohydrated 5-bromo-uracil). In particular, in the case of adenine, where just about the 10% of the events are initiated by prompt atom knockouts, the dynamics could help the identification of the mass/charge obtained by the experiment and proposing mechanism for the fragmentation. A natural follow up for this project would involve the consideration of the solvation effects.

- For the case of 5BrU, this combined dynamic and electronic structure approach could confirm that when exposed to energetic multiple ionized atoms (such as C^{4+}), this mutagen tends to tautomerize. This strongly supports the idea of its mutagenic mechanism to be initiated by a water-mediated tautomerization, which significantly lowers the intramolecular H-transfer barriers. The dynamics performed revealed the presence of a second reaction mechanism, where the water acts like a nucleophile and performs an attack to the Br-substituted carbon. Depending on the charge and number of hydrogen into the system, this process may result in a barrier-less water insertion mechanism, or in a nucleophilic substitution, where either neutral Br or HBr are expelled from the initial compound.

- The same type of reactions were shown to take place in uracil and the its F- and Cl-substituted. The different substitution in position C5 produced an higher variety of products, for example double ring compounds. Considering the stability of some of the reaction products which lose the halogens and the relatively low barriers, this may constitute a deteriorating process for the radiosensitizer / mutagen. On the other side, the inserted and double ring structures are higher in energy and more tense, so they may activate other reactions. Investigating the evolution of these new compounds would elucidate their roles in biological systems.

Chapter 12

Conclusiones y Perspectivas

Algunas consideraciones generales sobre los estudios presentados pueden resumirse como sigue:

- Los procesos de crecimiento molecular conducidos por knockout son bastante generales y probablemente ocurran para diferentes tipos de proyectiles, dependiendo de la cantidad de energía depositada a través de la dispersión nuclear;
- Es probable que los processos de knockout también jueguen un papel en el daño por radiación de biomoléculas más complejas;
- Los métodos BOMD y ADMP son efectivos para estudiar la descomposición estadística de moléculas y cluster en las condiciones de los experimentos analizados;
- El entorno puede desempeñar un papel clave en la descomposición inducida por colisión de las biomoléculas: determinar el predominio de una vía con respecto a otra a través de una acción catalítica o reaccionar activamente con las biomoléculas.

También se pueden sacar conclusiones más específicas:

- En cuanto a los clusters de butadieno, los iones a energías del orden de keV que colisionan con clusters de 1,3-butadieno pueden inducir reacciones de formación de enlaces. Los fragmentos que se forman

inicialmente de esta manera, a menudo, son altamente reactivos y forman enlaces covalentes con moléculas vecinas en escalas de tiempo de subpicosegundos. Nuestros cálculos de DFT muestran que los isómeros más estables de los productos de reacciones más abundantes tienen estructuras de anillo. Además, la distribución experimental del tamaño de los productos de reacción se puede entender en términos de las estabilidades relativas calculadas para dichos productos. Estos hallazgos sugieren que estamos formando anillos cíclicos a partir de moléculas inicialmente lineales. Las similitudes entre nuestros espectros de masas experimental y simulado sugieren que los procesos de eliminación de átomos muy rápidos que se incluyen en estas simulaciones desempeñan un papel importante. Sin embargo, no se puede excluir que también otros tipos de procesos, como los procesos de excitación electrónica o los procesos de ionización, también puedan iniciar el crecimiento molecular, en particular en cluster de moléculas pequeñas. Como perspectiva futura, sería interesante investigar sistemáticamente la importancia relativa de la ionización y de los procesos de stopping electrónicos y nucleares para el crecimiento molecular dentro de cluster de hidrocarburos.

- También se demostró que el proceso de knockout desempeña un papel en las colisiones entre los iones tetrafenilporfirina y los átomos de gas noble de energías (referidas al centro de masa) de 50 a 110 eV (ver capítulo 7). A través de comparaciones con los resultados de simulaciones de dinámica molecular clásica, hemos descubierto que aproximadamente el 40% y el 70% de las secciones transversales de fragmentación total se deben a dichos procesos en colisiones con He y Ne, respectivamente. El análisis del efecto de embedding de dichas moléculas en su entorno (cluster o solvente) puede ser relevante para inducir procesos de crecimiento molecular eficientes, similares a los observados en clusters de butadieno.

- La combinación de dinámica molecular y muestreo de las superficies de energía potencial han demostrado ser efectivas para los casos de sistemas lo suficientemente pequeños en ambas fases gaseosas (como adenina) y cluster con agua (5-bromo-uracilo nano-hidratado). En

particular, en el caso de la adenina, donde aproximadamente el 10% de los eventos se inician mediante knockouts rápidos de átomos, la dinámica podría ayudar a identificar la masa/carga obtenida por el experimento y proponer un mecanismo para la fragmentación. Un seguimiento natural de este proyecto implicaría la consideración de los efectos de la solvatación.

- Para el caso de 5BrU, este metodo combinado dinámica-estructura electronica confirma que cuando se expone a múltiples átomos ionizados energéticos (como C^{4+}), este mutágeno tiende a tautomerizarse. Esto apoya la idea de que su mecanismo mutagénico es iniciado por una tautomerización mediada por agua, que reduce significativamente las barreras de transferencia intramolecular de H. Las dinámicas realizadas revelaron la presencia de un segundo mecanismo de reacción, donde el agua actúa como un nucleófilo y realiza un ataque al carbono sustituido con Br. Dependiendo de la carga y la cantidad de hidrógeno en el sistema, este proceso puede resultar en un mecanismo de inserción de agua sin barreras, o en una sustitución nucleofílica, donde se expulsan Br o HBr neutros del compuesto inicial.

- Se ha demostrado en este trabajo que el mismo tipo de reacciones tienen lugar en uracilo y sus halo-substituidos en F y Cl. La diferente sustitución en la posición C5 produjo una variedad más alta de productos, por ejemplo, compuestos de doble anillo. Teniendo en cuenta la estabilidad de algunos de los productos de reacción que pierden los halógenos y las barreras relativamente bajas, esto puede constituir un proceso de deterioro para el mutágeno o de desactivación del radiosensitizer. Por otra parte, las estructuras de anillo insertado y doble tienen mayor energía y son más tensas, por lo que pueden activar otras reacciones. Investigar la evolución de estos nuevos compuestos ayudaría a establecer su impacto en sistemas biológicos.

Bibliography

[1] M. Gatchell and H. Zettergren. In: *J. Phys. B At. Mol. Opt. Phys.*, 49.162001 (2016).

[2] J. F. Ziegler and J. P. Biersack. Springer, Boston, MA, 1985.

[3] J. F. Ziegler. URL: http://srim.org.

[4] P. Swings and L. Rosenfeld. In: *Astrophys. J.*, 86 (1932), pp. 483–486.

[5] J. Cabannes et al. In: *Nature*, 142 (1938), p. 718.

[6] A. McKellar. In: *Publ. Astron. Soc. Pacific*, 52.307 (1940), p. 187.

[7] A. G. G. M. Tielens. In: *Rev. Mod. Phys.*, 85.3 (2013), pp. 1021–1081.

[8] J. M. Shull and S. Beckwith. In: *Annu. Rev. Astron. Astrophys.*, 20.1 (1982), pp. 163–190.

[9] T. Chen et al. In: *J. Chem. Phys.*, 142.14 (2015), p. 144305.

[10] M. Gatchell et al. In: *J. Phys. Conf. Ser.*, 583.1 (2015), p. 12011.

[11] J. F. Ward. In: *Prog. Nucleic Acid Res. Mol. Biol.* Ed. by W. E. Cohn et al. Vol. 35. Academic Press, 1988, pp. 95–125.

[12] K. Apel and H. Hirt. In: *Annu. Rev. Plant Biol.*, 55.1 (2004), pp. 373–399.

[13] S. Gregoli et al. In: *Radiat. Res.*, 89.2 (1982), pp. 238–254.

[14] D. Becker and M. D. Sevilla. In: *Adv. Radiat. Biol.*, 17 (1993), pp. 121–180.

[15] M. G. Debije et al. In: *Radiat. Res.*, 154.2 (2000), pp. 163–170.

[16] T. Elsässer et al. In: *Int. J. Radiat. Oncol.*, 78.4 (2010), pp. 1177–1183.

[17] W. P. Levin et al. In: *Br. J. Cancer*, 93 (2005), p. 849.

[18] M. Lodge et al. In: *Radiother. Oncol.*, 83.2 (2007), pp. 110–122.

[19] F. M. Johnson. Springer, Dordrecht, 1995.

[20] Z. Suo et al. In: *Astrobiology*, 7.4 (2007), pp. 605–615. URL: https://doi.org/10.1089/ast.2006.0120.

[21] B. H. Huynh et al. In: *J. Chem. Phys.*, 61.9 (1974), pp. 3750–3758.

[22] M. F. Perutz. In: *Proc. R. Soc. London B Biol. Sci.*, 208.1171 (1980), pp. 135–162.

[23] K. M. Kadish et al., eds. Amsterdam: Academic Press, 2003, pp. 1–20.

[24] S. Feil et al. In: *Int. J. Mass Spectrom.*, 255-256 (2006), pp. 232–238.

[25] A. D. Adler et al. In: *Org. Mass Spectrom.*, 3.7 (1970), pp. 955–962.

[26] B. D. Beato et al. In: *Chem. Geol.*, 91.2 (1991), pp. 185–192.

[27] J. D. Laycock et al. In: *J. Mass Spectrom.*, 32.9 (1998), pp. 978–983. URL: http://2-0.

[28] M. Meot-Ner et al. In: *Ann. N. Y. Acad. Sci.*, 206.1 (1973), pp. 641–648.

[29] T. Gozet et al. In: *Int. J. Mass Spectrom.*, 279 (2009), pp. 113–118.

[30] L. A. Hayes et al. In: *Int. J. Mass Spectrom.*, 227 (2003), pp. 111–120.

[31] V. Bernigaud et al. In: *Eur. Phys. J. D*, 51 (2009), pp. 125–130.

[32] L. M. Nuwaysir and C. L. Wilkins. In: *Anal. Chem.*, 61.7 (1989), pp. 689–694.

[33] J. A. Castoro et al. In: *Anal. Chem.*, 64.19 (1992), pp. 2238–2243.

[34] Z. Deng et al. In: *J. Chem. Phys.*, 127.14 (2007), p. 144715.

[35] Z. Deng et al. In: *Phys. Rev. Lett.*, 95.15 (2005), p. 153201.

[36] H. M. Dang et al. In: *Eur. Phys. J. D*, 63 (2011), pp. 359–367.

[37] L. Giacomozzi et al. In: *Phys. Chem. Chem. Phys.*, 19.30 (2017), pp. 19750–19755.

[38] S. Pilling et al. In: *Proc. Int. Astron. Union*, 4.S251 (2008), pp. 371–376.

[39] C. A. Cole et al. In: *J. Phys. Chem. A*, 119.2 (2015), pp. 334–343.

[40] Z. Martins et al. In: *Earth Planet. Sci. Lett.*, 270.1-2 (2008), pp. 130–136.

[41] M. P. Callahan et al. In: *Proc. Natl. Acad. Sci.*, 108.34 (2011), 13995 LP –13998.

[42] J. Kissel and F. R. Krueger. In: *Nature*, 326 (1987), p. 755.

[43] S. Chakraborty et al. In: *Int. J. Quantum Chem.*, 118.2 (2017), e25459.

[44] S. Pilling et al. In: *J. Phys. Chem. A*, 113.42 (2009), pp. 11161–11166.

[45] R. Glaser et al. In: *Astrobiology*, 7.3 (2007), pp. 455–470.

[46] D. Roy et al. In: *Proc. Natl. Acad. Sci.*, 104.44 (2007), 17272 LP –17277.

[47] K. M. Merz et al. In: *J. Phys. Chem. A*, 118.20 (2014), pp. 3637–3644.

[48] A. M. Kamel and B. Munson. In: *Eur. J. Mass Spectrom.*, 10.2 (2004), pp. 239–257.

[49] J. M. Rice and G. O. Dudek. In: *J. Am. Chem. Soc.*, 89.11 (1967), pp. 2719–2725.

[50] C. C. Nelson and J. A. McCloskey. In: *J. Am. Chem. Soc.*, 114.10 (1992), pp. 3661–3668.

[51] H.-W. Jochims et al. In: *Chem. Phys.*, 314.1 (2005), pp. 263–282.

[52] J. Tabet et al. In: *Int. J. Mass Spectrom.*, 292.1 (2010), pp. 53–63.

[53] B. F. Minaev et al. In: *J. Chem. Phys.*, 140.17 (2014), p. 175101.

[54] D. T. Major et al. In: *J. Org. Chem.*, 67.3 (2002), pp. 790–802.

[55] B. Boudïaffa et al. In: *Science (80-.).*, 287.5458 (2000), 1658 LP –1660.

[56] F. Martin et al. In: *Phys. Rev. Lett.*, 93.6 (2004), p. 68101.

[57] T. Schlathölter et al. In: *ChemPhysChem*, 7.11 (2006), pp. 2339–2345.

[58] J.-C. Poully et al. In: *Phys. Chem. Chem. Phys.*, 17.11 (2015), pp. 7172–7180.

[59] M. A. Huels et al. In: *J. Am. Chem. Soc.*, 125.15 (2003), pp. 4467–4477.

[60] K. K. Khanna and S. P. Jackson. In: *Nat. Genet.*, 27 (2001), p. 247.

[61] J. F. Fowler et al. In: *Cancer Treat. Rev.*, 3.4 (1976), pp. 227–256.

[62] E. B. Douple and R. C. Richmond. In: *Br. J. Cancer. Suppl.*, 3 (1978), pp. 98–102.

[63] J. D. Watson and F. C. H. Crick. In: *Nature*, 171.8 (1953), pp. 737–738.

[64] A. J. F. Griffiths et al. New York: W.H. Freeman and Company, 2012.

[65] N. V. Bhagavan and C.-E. Ha. Academic Press, 2011.

[66] S. Brenner and J. H. Miller. Academic Press, 2001.

[67] X. Hu et al. In: *Biochemistry*, 43 (2004), pp. 6361–6369.

[68] V. I. Danilov et al. In: *J. Phys. Chem. A*, 113.11 (2009), pp. 2233–2235.

[69] T. van Mourik et al. In: *Theor. Chem. Acc.*, 125.3 (2010), pp. 233–244.

[70] X. Hu et al. In: *J. Phys. Chem. B*, 108 (2004), pp. 12999–13007.

[71] M. Orozco et al. In: *J. Phys. Chem. B*, 102 (1998), pp. 5228–5233.

[72] M. Palafox et al. In: *Spectrosc. Lett.*, 44 (2011), pp. 300–306.

[73] G. Fogarasi. In: *Chem. Phys.*, 349.1-3 (2008), pp. 204–209.

[74] D.-S. Ahn et al. In: *Chem. Phys. Lett.*, 390.4 (2004), pp. 384–388.

[75] M. D. Topal and J. R. Fresco. In: *Nature*, 263.5575 (1976), pp. 285–289.

[76] E. Kryachko et al. In: *J. Phys. Chem. A*, 105 (2001), pp. 1288–1295.

[77] E. Kryachko et al. In: *J. Phys. Chem. A*, 105 (2001), pp. 1288–1295.

[78] J. F. Ward. In: *Adv. Radiat. Biol.*, 5 (1975), pp. 181–239.

[79] J. D. Zimbrick et al. In: *Int. J. Radiat. Biol. Relat. Stud. Physics, Chem. Med.*, 16.6 (1969), pp. 505–523.

[80] T. S. Lawrence et al. In: *Semin. Radiat. Oncol.*, 13.1 (2003), pp. 13–21.

[81] T. Bergen et al. In: *Rev. Sci. Instrum.*, 70.8 (1999), pp. 3244–3253.

[82] N. De Ruette et al. In: *Rev. Sci. Instrum.*, 89.7 (2018), pp. 1–12.

[83] R. D. Thomas et al. In: *Rev. Sci. Instrum.*, 82.6 (2011), p. 65112.

[84] H. T. Schmidt et al. In: *Rev. Sci. Instrum.*, 84.5 (2013), p. 55115.

[85] W. C. Wiley and I. H. McLaren. In: *Rev. Sci. Instrum.*, 26.12 (1955), pp. 1150–1157.

[86] N. A. Mautjana et al. In: *Electrochim. Acta*, 55.1 (2009), pp. 52–58.

[87] D. J. Douglas and J. B. French. In: *J. Am. Soc. Mass Spectrom.*, 3.4 (1992), pp. 398–408.

[88] D. Gerlich. 2007.

[89] C. Peng and H. Bernhard Schlegel. In: *Isr. J. Chem.*, 33.4 (1993), pp. 449–454.

[90] K. Fukui. In: *Acc. Chem. Res.*, 14.12 (1981), pp. 363–368.

[91] M. J. Frisch et al.

[92] M. Born and J. R. Oppenheimer. In: *Ann. Phys.*, 84.458 (1927).

[93] P. Hohenberg and W. Kohn. In: *Phys. Rev.*, 136.3B (1964), B864–B871.

[94] W. Kohn and L. J. Sham. In: *Phys. Rev.*, 140.4A (1965), A1133–A1138.

[95] C. J. Cramer. 2nd. John Wiley & Sons Ltd, 2004.

[96] J. Slater. In: *Quantum Theory Mol. solids*, 4 (1974).

[97] J. P. Perdew and W. Yue. In: *Phys. Rev. B*, 33.12 (1986), pp. 8800–8802.

[98] E. Engel and R. M. Dreizler. New York: Springer Heidelberg Dordrecht, 2011, pp. 255–368.

[99] S. H. Vosko et al. In: *Can. J. Phys.*, 58.8 (1980), pp. 1200–1211.

[100] A. D. Becke. In: *Phys. Rev. A*, 38.6 (1988), pp. 3098–3100.

[101] K. P. Huber and G. L. Herzberg. New York: Van Nostrand Reinhold, 1979.

[102] J. Heyd et al. In: *J. Chem. Phys.*, 123.17 (2005), p. 174101.

[103] M. Städele et al. In: *Phys. Rev. B*, 59.15 (1999), pp. 10031–10043.

[104] C. J. Umrigar and X. Gonze. In: *Phys. Rev. A*, 50.5 (1994), pp. 3827–3837.

[105] F. Herman et al. In: *Phys. Rev. Lett.*, 22.16 (1969), pp. 807–811.

[106] F. Herman et al. In: *INt. J. Quantum Chem.*, IIIS (1970), p. 827.

[107] J. P. Perdew et al. In: *Phys. Rev. Lett.*, 82.12 (1999), pp. 2544–2547.

[108] J. Tao et al. In: *Phys. Rev. Lett.*, 91.14 (2003), p. 146401.

[109] Y. Zhao et al. In: *J. Chem. Theory Comput.*, 2.2 (2006), pp. 364–382.

[110] A. D. Becke. In: *J. Chem. Phys.*, 98.2 (1993), pp. 1372–1377.

[111] C. Lee et al. In: *Phys. Rev. B*, 37.2 (1988), pp. 785–789.

[112] J. P. Perdew. In: *Electron. Struct. Solids '91*. Ed. by P. Ziesche and H. Eschrig. Vol. 17. Physical Research. Berlin: Akademie Verlag, 1991, pp. 11–20.

[113] Y. Wang and J. P. Perdew. In: *Phys. Rev. B*, 43.11 (1991), pp. 8911–8916.

[114] F. Jensen. West Sassex, England: John Wiley & Sons Ltd, 2007, p. 624.

[115] A. D. McLean and G. S. Chandler. In: *J. Chem. Phys.*, 72.10 (1980), pp. 5639–5648.

[116] R. Krishnan et al. In: *J. Chem. Phys.*, 72.1 (1980), pp. 650–654.

[117] J. Tersoff. In: *Phys. Rev. Lett.*, 56.6 (1988).

[118] S. J. Stuart et al. In: *J. Chem. Phys.*, 112.14 (2000), pp. 6472–6486.

[119] D. W. Brenner et al. In: *J. Chem. Phys.*, 783 (2002).

[120] M. Wolf et al. In: *Eur. Phys. J. D*, 70.4 (2016), p. 85.

[121] S. Plimpton et al. URL: https://lammps.sandia.gov/.

[122] M. Elstner and G. Seifert. In: *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 372.2011 (2014), pp. 20120483–20120483.

[123] J. C. Slater and G. F. Koster. In: *Phys. Rev.*, 94.6 (1954), pp. 1498–1524.

[124] M. Elstner et al. In: *Phys. Rev. B*, 58.11 (1998), pp. 7260–7268.

[125] S. Froyen and W. A. Harrison. In: *Phys. Rev. B*, 20.6 (1979), pp. 2420–2422.

[126] W. A. Harrison. In: *Phys. Rev. B*, 34.4 (1986), pp. 2787–2793.

[127] D. J. Chadi. In: *Phys. Rev. Lett.*, 43.1 (1979), pp. 43–47.

[128] J. Harris. In: *Phys. Rev. B*, 31.4 (1985), pp. 1770–1779.

[129] T. Helgaker et al. In: *Chem. Phys. Lett.*, 173.2 (1990), pp. 145–150.

[130] E. Uggerud and T. Helgaker. In: *J. Am. Chem. Soc.*, 114.11 (1992), pp. 4265–4268.

[131] S. S. Iyengar et al. In: *J. Chem. Phys.*, 115.22 (2001), pp. 10291–10302.

[132] H. B. Schlegel et al. In: *J. Chem. Phys.*, 114.22 (2001), pp. 9758–9763.

[133] H. B. Schlegel et al. In: *J. Chem. Phys.*, 117.19 (2002), pp. 8694–8704.

[134] D. Marx and J. Hutter. In: *Mod. Methods Algorithms Quantum Chem.* Ed. by J. Grotendorst. John von Neumann Institute for Computing, 2000, pp. 301–449.

[135] W. Miller. Springer US, 1976.

[136] J. C. Tully. In: *Mod. Methods Multidimens. Dyn. Comput. Chem.* 1998, pp. 34–72.

[137] J. C. Tully. Singapore: World Scientific, 1998, pp. 385–404.

[138] H. Goldestein. Wiesbaden: Aula-Verlag, 1987.

[139] F. Furche et al. URL: http://www.turbomole.com/.

[140] E. Peeters. In: *Proc. Int. Astron. Union*, 7.S280 (2011), pp. 149–161.

[141] O. Berné and A. G. G. M. Tielens. In: *Proc. Natl. Acad. Sci.*, 109.2 (2012), 401 LP –406.

[142] M. Gatchell et al. In: *Phys. Chem. Chem. Phys.*, 19.30 (2017), pp. 19665–19672.

[143] C. J. Henning et al. In: *Astrophys. J.*, 696.1 (2009), p. 706.

[144] G. Y. Gerasimov. In: *J. Eng. Phys. Thermophys.*, 82.3 (2009), pp. 432–441.

[145] R. Peverati et al. In: *J. Phys. Chem. A*, 118.43 (2014), pp. 10109–10116.

[146] Y. Carpentier et al. In: *J. Phys. Chem. A*, 117.39 (2013), pp. 10092–10104.

[147] R. P. Head-Gordon et al. In: *Astrophys. J.*, 830.2 (2016), p. 128.

[148] C. S. C. Salama and Farid. In: *Astrophys. J. Suppl. Ser.*, 208.1 (2013), p. 6.

[149] H. Richter and J. B. Howard. In: *Prog. Energy Combust. Sci.*, 26.4 (2000), pp. 565–608.

[150] B. Shukla and M. Koshi. In: *Combust. Flame*, 158.2 (2011), pp. 369–375.

[151] P. P. Bera et al. In: *Phys. Chem. Chem. Phys.*, 17.3 (2015), pp. 1859–1869.

[152] A. M. Mebel et al. In: *J. Am. Chem. Soc.*, 130.41 (2008), pp. 13618–13629.

[153] S. Irle et al. In: *J. Phys. Chem. B*, 110.30 (2006), pp. 14531–14545.

[154] J. Z. Tielens et al. In: *Astrophys. J. Lett.*, 797.2 (2014), p. L30.

[155] O. Berne et al. In: *Astron. Astrophys.*, 133 (2015), pp. 1–9.

[156] E. R. Micelotta et al. In: *Astron. Astrophys.*, 510 (2010), A36.

[157] F. Seitz et al. In: *J. Chem. Phys.*, 139.3 (2013), p. 34309.

[158] R. Delaunay et al. In: *J. Phys. Chem. Lett.*, 6.9 (2015), pp. 1536–1542.

[159] Y. Wang et al. In: *Phys. Rev. A*, 89.6 (2014), p. 62708.

[160] H. Zettergren et al. In: *Phys. Rev. Lett.*, 110.18 (2013), p. 185501.

[161] J. C. Robinson et al. In: *J. Am. Chem. Soc.*, 124.34 (2002), pp. 10211–10224.

[162] P. Linstrom and W. Mallard. 2005.

[163] J. R. Premkumar et al. In: *Indian J. Chem.*, 53 (2014), pp. 985–991.

[164] P. Franceschi et al. In: *Int. J. Mass Spectrom.*, 220.2 (2002), pp. 281–288.

[165] H. Zettergren et al. In: *J. Chem. Phys.*, 133.10 (2010), p. 104301.

[166] J. P. Sánchez et al. In: *J. Phys. Chem. A*, 120.4 (2016), pp. 588–605.

[167] N. F. Aguirre et al. In: *J. Chem. Theory Comput.*, 13.3 (2017), pp. 992–1009.

[168] F. J. Gonzalez. In: *Pharmacol. Rev.*, 40.4 (1989).

[169] J. H. Dawson and M. Sono. In: *Chem. Rev.*, 87.5 (1987), pp. 1255–1276.

[170] C. B. Nielsen et al. In: *J. Phys. Chem. A*, 109.17 (2005), pp. 3875–3879.

[171] V. Bernigaud et al. In: *J. Am. Soc. Mass Spectrom.*, 19.6 (2008), pp. 809–813.

[172] M. K. Lykkegaard et al. In: *J. Phys. Chem. A*, 113.8 (2009), pp. 1440–1444.

[173] M. H. Stockett et al. In: *Phys. Chem. Chem. Phys.*, 16.40 (2014), pp. 21980–21987.

[174] S. Martin et al. In: *Phys. Rev. Lett.*, 110.6 (2013), p. 63003.

[175] S. Martin et al. In: *Phys. Rev. A*, 85.5 (2012), p. 52715.

[176] A. I. S. Holm et al. In: *J. Chem. Phys.*, 134.4 (2011), p. 44301.

[177] M. H. Stockett et al. In: *Phys. Rev. A*, 89.3 (2014), p. 32701.

[178] N. Russo et al. In: *J. Comput. Chem.*, 19.9 (1998), pp. 989–1000.

[179] F. Turecček and X. Chen. In: *J. Am. Soc. Mass Spectrom.*, 16.10 (2005), pp. 1713–1726.

[180] C. Marian et al. In: *Phys. Chem. Chem. Phys.*, 7.18 (2005), pp. 3306–3316.

[181] M. C. Castrovilli et al. In: *Phys. Chem. Chem. Phys.*, 19.30 (2017), pp. 19807–19814.

[182] A. D. Huber et al. In: *J. Phys. Conf. Ser.*, 373.1 (2012), p. 12005.

[183] S. K. Kim et al. In: *J. Phys. Chem.*, 100.19 (1996), pp. 7933–7937.

[184] N. J. Kim et al. In: *Int. J. Mass Spectrom.*, 219.1 (2002), pp. 11–21.

[185] P. Markush et al. In: *Phys. Chem. Chem. Phys.*, 18.25 (2016), pp. 16721–16729.

[186] S. Maclot et al. In: *ChemPhysChem*, 12.5 (2011), pp. 930–936.

Acknowledgements

I would like to thank all the people that have worked with me on the work presented in this thesis, starting from the bosses, my supervisors: prof. Pedro Alexandrino Fernandes and prof. Manuel Alcamí. I would like to especially thank Manuel Alcamí for his infinite patience and the attitude he has always had with me, trying to alleviate all the additional bureaucratic tasks who sometimes have pressured me quite a lot. A special thanks goes also to prof. Henning Zettergren, who, even if not an "official" supervisor, has been of great support during these three years. During my PhD I was also assigned some advisors, thanks then to Michael Gatchell and Sergio Sousa for their support and the interesting discussions I could have with them.

Growing up I got more skeptical and cynical, this is probably something many people can relate with. However, (I hope) everyone of us knows somebody who is exceptional, somebody who is genuinely keen on her/his profession, who is passionate and inspires the same passion in others. "My person" during this last year and a half has been Sergio Díaz-Tendero. My most sincere thanks for all the support and advising you have provided to this work so far.

I, finally, would like to thank all the people from the atom fysik group at Stockholm University, from the BioSim and the rest of Maria João Ramos group in Porto University and the group of Madrid for the meals shared, the light (or not) chats and all their niceness.

I remember reading in my brother's thesis how meticulous we are in mentioning strangers who have worked in our same field, while the same attention is not applied when citing our loved ones who also helped us achieving our goals. I am not going to copy the idea of the "people important in my life" bibliography, but I am going to try to give all of you the credit you deserve.

I will start with my family. You will pardon me if I will briefly switch to Italian now. Ho le mie incomprensioni con mia madre, nonostante questo sono sempre accolta e, recentemente, persino supportata nelle mie scelte di vita. Grazie per questo tuo amore ansioso (a tratti soffocante). Mio fratello è sempre impegnato, una tabella di marcia pianificata nei minimi dettagli, non foss'altro che è sempre in ritardo (anche se io di più). Grazie per rispondermi sempre quando vado in crisi e per sopportarmi quando ti rimprovero per non fare quello che farei io al posto tuo (una cosa veramente stupida da fare). Vorrei inoltre ringraziare mio zio Tonino e mia zia Renata che mi sono sempre vicini a modo loro, anche se mia zia Renata un poco poco instabile è, ma ci piaci così. Se sono una femminista, lo devo al fatto di essere cresciuta tra le donne forti della mia famiglia, quasi tutte. In particolare, mia nonna Mariella, che è tra le persone più care che ho sin da quando ero bambina. Ringrazio mio zio Carmine e mia zia Pina che mi accolgono sempre calorosamente e che in me ci credono da sempre.

Inizio con la famiglia che mi è toccata finisco con la famiglia che mi sono scelta. Prima di dire qualsiasi cosa su ognuno di voi, vorrei ringraziarvi per l'amicizia, che è quella cosa per cui ti vedi in persona due volte all'anno, ma sembra che ti sia lasciato ieri. In effetti, sono io quella che non c'è mai, la cosa bella è sapere di essere già perdonata.

Ringrazio tutti i miei "amici del paese (Grumo Nevano, la piccola Parigi)". Oreste, per il suo essere sempre dolce nei miei riguardi. Pietro, che si sopporta anche il peggio di me, però continua ad esserci, non so bene perché. Stefania che è Stefania, un pronto intervento di allegria, riesce a farti ridere anche dei suoi guai, che per fortuna non sono veramente così tanti come lei pensa. Teresa e Antonio, the royal graduation. Antonio, meno numeroso di prima, ma comunque sempre "avviluppante". Teresa, che sa i fatti nostri prima che noi stessi abbiamo tempo di rendercene conto, quindi empatica e comprensiva (ma anche capera pro). Gigi, a cui devo la mia iniziazione alla fotografia seria. Gina, che nonostante la sua logorrea, trova anche tempo di ascoltarmi (a volte). Riccardo, con cui sono cresciuta e che ha soffiato per anni sul fuocherello della mia fame di conoscere il mondo, facendo esplodere un incendio (sorry, mom). Ringrazio inoltre: Chiara la tranquillona, Mà la modesta, Danila l'amica di Stefania e Marco stilista di successo per la loro amicizia.

Dovrei aggiungere a questa famiglia scelta anche la mia famiglia scientifica. Vorrei ringraziare Ana e Michele, "galeotti" per avermi fatto innamorare della ricerca, che non solo sono miei mentori, ma anche miei amici. Eduagna, un bachelor e un master condivisi (coppia Schiavo - De Angelis, "ma siete fratelli?"), che continua ad essere mio amico, per sfortuna per lui, pure a distanza; ed Arianna, che quest'anno si è sopportata le Eduard's tasks pazientemente per aiutarmi con le mie "carte e cartuscelle".

A tutte queste persone devo aggiungere anche i miei fratelli acquisiti, Aldo e Michele. Grazie per essermi stati vicini nei momenti peggiori e in quelli migliori, di molti dei quali siete protagonisti. Principalmente Aldo, perché Michele è un traditore che si perde la mia difesa !!!

I would like to thank the people I met during my travelling and who became very good friends to me. Gerard, who has the merit of ending my period of loneliness in Stockholm with a lot of irish dances, chillith, dinners and a friendship that lasts until today. Irina, who always asks, listens and never judges. She has this tremendous power of always making me feel understood and supported. Joana, who was such a good friend to give me a roof when I was trying to find a place in Porto, and made me feel very welcome in a country new for me at that point. Fabian, who is always extremely gentle with me and always has interesting stories or data to share (sometimes made up). Vlad, who is kind of a new entry, but who already has my heart. I also would like to thank Colin, Alley, Alekz, Shreashy, Walid, Ferran and all the "Desert Island" once "Strong Princesses" members for all the good moments we shared in Sweden. Voglio inoltre ringraziare i miei coinquilini *ad interim* di Stoccolma per la pazienza e l'ospitalità: Karin (ormai veterana) e Andrea. I shall mention now Eduardo Oliveira with whom I shared very good time in Porto, together with the best rommies one could ever have, Deni and Patti. I would like to thank all my peer-PhD students in Madrid, with whom I shared "Comidas Cenas Cientificas" (CCC). Thanks to Eva, Darío (por enseñarnos a "The room"), Paula, Fernando, Enrique, Ransel, Pablo (que me odia, aunque yo le quiera), Marcos (al cual van gracias adicionales por suportar mis mismas causas y hacer las justas preguntas), Juni, Guille (por su alegria y compañerismo constante), Arturo (por iniciarme al mercado de Vallehermoso y a más cosas de Madrid, entre las cuales su piso), Pedro, Vicente and Evaristo. Gracias a Luis y a Mikel,

por las muchas conversas, a veces profundas y dolidas. To end this part, I would like to thank my co-ESRs for the adventures around Europe, especially Dima, Neus and Ewa.

I think it is clear, in my case, that if sleep is the balm of hurt minds, friendship is the balm of wandering minds. In fact, I am a rolling stone, but a lucky one, because I feel home whenever I am with you, Massi. Grazie per esserci ormai da sempre.

I finally want to thank my father, who I guess is the main responsible for what I am now. I was born privileged, as I was given the opportunity to shape my future as I wanted and I have always felt accepted simply for what I am, not exactly a standard person from a small italian village. Sinceramente, grazie.

In the acknowledgment of my MSc thesis, I have offered one of my kidneys to the people I was grateful to. Quick update: nobody has claimed it so far, and my kidneys have sensibly improved their quality. So, if you are ever going to need one...