Modelling Interstellar Complex Molecules : Developments and Simulations

Maysá Yusef Buey

To cite this version:

Maysá Yusef Buey. Modelling Interstellar Complex Molecules : Developments and Simulations. Chemical Sciences. Université Paul Sabatier - Toulouse III, 2023. English. NNT : 2023TOU30245 . tel-04550742

HAL Id: tel-04550742
https://theses.hal.science/tel-04550742

Submitted on 18 Apr 2024

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Maysá YUSEF BUEY

Le 10 novembre 2023

Modélisation de molécules interstellaires complexes : développements et simulations

Ecole doctorale : **SDM - SCIENCES DE LA MATIERE - Toulouse**

Spécialité : **Physico-Chimie Théorique**

Unité de recherche :
LCPQ - Laboratoire de Chimie et Physique Quantiques

Thèse dirigée par
Mathias RAPACIOLI et Tzonka MINEVA

Jury
M. Sergio DíAZ TENDERO, Rapporteur
M. Giacomo MULAS, Rapporteur
Mme Carine CLAVAGUERA, Examinatrice
Mme Phuong Mai DINH, Examinatrice
M. Mathias RAPACIOLI, Directeur de thèse
Mme Tzonka MINEVA, Co-directrice de thèse

PhD Thesis

Université Toulouse 3 – Paul Sabatier
Toulouse, France

Modelling Interstellar Complex Molecules: Developments and Simulations

10th November 2023

Maysá YUSEF BUEY

Thesis supervised by
Mathias RAPACIOLI and Tzonka MINEVA

Jury members:

M. Sergio DÍAZ TENDERO
M. Giacomo MULAS
Mme. Carine CLAVAGUERA
Mme. Phuong Mai DINH
M. Mathias RAPACIOLI
Mme. Tzonka MINEVA
Mme. Dahbia TALBI

من سماء إلى أختها يعبر الحالمون
حاملين مرايا من الماء حاشية للفراشة
في وسعنا أن نكون كما ينبغي أن نكون
من سماءٍ
إلى أختها
يعبر الحالمون ...

From one sky to another the dreamers pass —
the butterfly's attendants carry mirrors of water.
We could be what we should be.
From one sky to another the dreamers pass ...

Mahmoud Darwish

¡Bienvenida sea la risa
que deja alegría por donde pisa!
¡Que venga la risa
y su prima la sonrisa!

Gloria Fuertes

Acknowledgements

As Antonio Machado said: *"Caminante, no hay camino, se hace camino al andar"–"Traveler, there is no road; you make your own path as you walk."* Indeed, during the three years of my PhD, I walked, but I did not walk alone. There is a great number of people that I want to acknowledge their support and encouragement throughout my thesis time. I dedicate this work to them.

I would like to express my honest gratitude to my supervisors, Mathias Rapacioli, and Tzonka Mineva. Thank you for your trust, patience, and great scientific discussions. Thank you for the knowledge and good scientific practices that you have taught me during these three years. Most importantly, I would like to thank you for your kindness and humanity throughout this journey.

Thanks to the jury members of my thesis for accepting to study my work, in particular Sergio Díaz-Tendero Victoria and Giacomo Mulas. I appreciate all their valuable questions and suggestions.

These acknowledgements extend to Aude Simon, Dahbia Talbi, Fernand Spiegelman, and Auréliene de la Lande for their great advices, suggestions, and kindness. I enjoyed and learnt a lot from our exchanges and collaborations. In addition, I am grateful to all the members of the ANR-RUBI project and deMon community. Our discussions during the conferences were very inspiring.

I cannot forget to thank Amandine Laurent and Patricia Bordais who have been of great help throughout the administrative part of this research, David Sanchez for the informatics help and Eric Colledani for the technical support. Also great thanks to Arya for editing my English. I would also like to express my gratitude to my colleagues in MAD group, LCPQ, and FERMI Federation, specially to the PhD students and Post-Docs that I encountered during these years. I am glad it helped me to find new friends.
I would also like to thank my previous mentors and supervisors, specially Marina P. Arrieta, Irene Natalia, Cyril Martins and Nicolas Suaud for sharing their research experience during my undergraduate and postgraduates studies.

Many thanks to my family. Thank you very much for your immense support and for providing me with the best of two cultures.
To my family in Spain: your endless interest at every gathering and every phone call encouraged me all along my work. Thanks to Gloria, Javier, and Daniel for being present at the day of my PhD defence. I cannot imagine a better family with whom I can share this achievement. *¡Muchas gracias!*
To my family in Palestine: thank you for all the good moments, surrounded by tea, laughs, and olive trees. *Shukran iktir.*
To my grandparents, who still make part of everything.

Thanks to my new family in Toulouse. Marie, Olivier, Pablo, Simone. Thank you for your warm welcoming and all your help during these years. An immense thanks to Jordi, Sandra, Zak, Raquel, Cloé, Marine, Toño and *la bande de copains et copines de Toulouse.* Thanks for all the great moments: discussions around beers, hikes and ski at the Pyrenees, cinema, music concerts, and breaking free dances. This thesis would not have been done without all of you. *Merci enormement !*

Thanks to Blandine. I never imagined that the time I spent in Italy would reward me with a great friend like you. *Grazie mille cara amica!*

Thanks to Emma. Thanks for sharing me the literature part of life. Our conversations about Chimamanda, Audre Lorde, Virginia Woolf, Murakami, Kawabata, Olga Tokarczuk and many others, kept my feet on the ground. *Dankjewel!*

I am incredibly grateful to my friends in Madrid. Raquel y Cris (*mis queridísimas apumis*), Sergio, Xexu, Cristitina (*casi finesa*), Vicent, Sarai. I still remember the time we spent in high school, *en el barrio*, at the library, holidays together, climbing, music festivals, or our cozy reunions every time I came back to Madrid. Thanks for all those moments filled with encouragement, listening, empathy, laughs, love, and caring. *¡Muchas gracias amigas y amigos!*

Many thanks to all the nice people I encountered in my time in Italy, Belgium and France.

I would like to acknowledge the Spanish, Italian, Belgian and French public education system. Thanks to the teachers and professionals working in these institutions. This work would not have been possible without a solid, accessible and public education system.
I would also like to dedicate some lines thanking the women in science and in every other disciplines and professions. Thanks for their persistence and their resistance, opening the way for those coming after.

Thanks to every journalist in the radio, musician, film maker, and writer who enlivened and accompanied during my research work and my thesis writing.

Thanks (and apologies) to all the people that I forgot to list here. Thank you very much !

To finish, I want to thank and dedicate this thesis to my four cornerstones: my parents, my sister, and Fernand. Thank you for your unconditional love and encouragement. Once I read that the language of the heart is the one taught by your parents. I would like to finish with some words in Spanish.

A mi madre. Gracias por enseñarme la importancia de estudiar pero también de jugar. Gracias por sostenerme, siempre, y por enseñarme a querer y cuidar a la gente que te quiere.

A mi padre. Gracias por enseñarme la importancia del esfuerzo y de la honestidad en el trabajo. Gracias por transmitirme tu seguridad y generosidad.

A mi hermana, Mayi. Gracias por ser mi compañera de juegos, de estudios, por compartir conmigo cada momento. Gracias por todo tu apoyo, en forma de galletas, de llamada de teléfono, en tarde de hermanas o en forma de abrazo.

A Fernand, mi compañero de viaje. Gracias simplemente por estar conmigo y por transmitirme tu inmensa tranquilidad. Je te garde comme la derniere allumente.

En definitiva, gracias por vuestro amor incondicional.

Muchas gracias de corazón a todas y todos.

Abstract

Complex molecules present in the Interstellar Medium such as Complex Organic Molecules (COMs) and polycyclic aromatic hydrocarbons (PAHs) have been the long-sought astronomical species. The understanding of their formation, their behavior in/on interstellar ices and their evolution under astronomical sources of irradiations, have gain great importance as they are believed to play one of the key role in the appearance of life on Earth.

This thesis, is mostly devoted to the study of glycine prebiotic molecule immersed in a low density amorphous (LDA) interstellar ice and isolated PAHs, using theoretical chemistry methods. This is a challenging work, considering the variety of conformations of COMs and their possible evolutions under different interstellar conditions. This complexity necessitated further developments of theoretical tools that were implemented in deMonNano software package. Firstly, in order to deal with the large system sizes, the combination of Quantum Mechanics (QM) and Molecular Mechanics (MM) approaches have been implemented where DFTB level of theory has been used for the energy and forces computation in the QM part, and class 1 force fields potential has been included in the MM description. Furthermore, for the study of ultrafast evolution phenomena, the Real Time-Time Dependent-DFTB (RT-TD-DFTB) method has been implemented. Both DFTB(QM)/MM and RT-TD-DFTB implementations were combined.

The dynamic evolution of glycine molecule, was first studied by embedding glycine in LDA ice and using the QM/MM approach. The questions of glycine conformational space and the related transformations were addressed under relevant astrophysical conditions. The results confirmed that the large water network around glycine as well as the various glycine isomers must be explicitly considered in order to capture the tautomerization mechanisms. From the Astrochemistry point of view, this research suggests that glycine in the cold interstellar ices should remain in its neutral form up to at least 250K.

Finally, the preliminary studies of the ultrafast electronic dynamics induced by ion collision were carried out using the RT-TD-DFTB method, where the N-Glycine molecule in LDA has been bombarded by a proton particle. Results confirmed that electron charge motions mostly occur within the glycine molecule (intramolecular displacement), although a weak intermolecular displacement is observed between the glycine molecule and the ice environment. Induced electron displacement was also captured during the ion-PAH collision study. In both cases, an ebb-and-flow effect was captured.

This PhD research has contributed to the elaboration of new computational tools, allowing to address large size systems and ultrafast electron dynamics. Regarding the applications, this thesis work broadened the knowledge of glycine conformational space and the related glycine's transformations. In addition, the electronic response of these systems submitted to collisions with stellar winds or cosmic rays particles have been simulated. On the basis of the new theoretical tools that have been implemented and their extensions, new perspectives are opened for theoretical studies of the irradiation induced chemical evolution of COMs or their ionization phenomena.

Résumé

Les molécules complexes présentes dans le milieu interstellaire, telles que les molécules organiques complexes (COM pour *Complex Organic Molecules*) et les hydrocarbures aromatiques polycycliques (PAH pour *Polycyclic aromatic hydrocarbon*), interessent la communauté scientifique depuis longtemps. La compréhension de leur formation, de leur comportement dans/sur les glaces interstellaires ou de leur évolution sous l'influence d'irradiations des rayonnement cosmique, a pris une grande importance car on pense qu'elles jouent un rôle clé dans l'apparition de la vie sur Terre.

Cette thèse est principalement consacrée à l'étude de la molécule prébiotique de glycine immergée dans une glace interstellaire amorphe de faible densité (LDA pour *Low Density Amorphous*) et des PAH isolés, en utilisant des méthodes de chimie théorique. Il s'agit d'un travail complexe, compte tenu de la variété des conformations des COM et de leurs évolutions possibles dans différentes conditions interstellaires. Cette complexité a nécessité le développement d'outils théoriques qui ont été implémentés dans le logiciel deMonNano. Tout d'abord, afin de traiter les systèmes de grande taille, la combinaison des approches de la mécanique quantique (QM pour *Quantum Mechanics*) et de la mécanique moléculaire (MM) a été implémentée où le niveau de théorie DFTB (pour *Density Functional Tight-Binding*) est utilisé pour le calcul de l'énergie et des forces dans la partie QM et le potentiel des champs de force de classe 1 est utilisé dans la description de la MM. Ensuite,

pour l'étude des phénomènes ultrarapides, la méthode RT-TD-DFTB (pour *Real Time-Time Dependent-Density Functional Tight-Binding*) a été aussi implémentée. Les implémentations DFTB(QM)/MM et RT-TD-DFTB ont été combinées.

L'évolution dynamique de la molécule de glycine a d'abord été étudiée en intégrant la glycine dans la glace LDA et en utilisant l'approche QM/MM. Les questions de l'espace conformationnel de la glycine et des transformations associées ont été abordées pour des conditions astrophysiques pertinentes. Les résultats ont confirmé que le vaste réseau d'eau autour de la glycine et que les divers isomères de la glycine doivent être explicitement pris en compte afin de comprendre les mécanismes de tautomérisation. Du point de vue de l'astrochimie, cette recherche suggère que la glycine dans les glaces interstellaires froides devrait rester sous sa forme neutre jusqu'à au moins 250 K.

Enfin, les études préliminaires de la dynamique électronique ultrarapide induite par la collision ionique ont été réalisées à l'aide de la méthode RT-TD-DFTB, où la molécule de N-Glycine dans le LDA a été bombardée par un proton. Les résultats ont confirmé que les migrations de charges électroniques se produisent principalement à l'intérieur de la molécule de glycine (migration intramoléculaire), bien qu'une faible migration intermoléculaire soit observée entre la molécule de glycine et l'environnement de glace. La migration des électrons a également été observée lors de l'étude de la collision ion-PAH. Dans les deux cas, un effet de flux et de reflux a été observé.

En conclusion, cette thèse a contribué à l'élaboration de nouveaux outils de calcul, permettant d'"etudier des systèmes de grande taille et la dynamique électronique ultrarapide. En ce qui concerne les applications, ce travail de thèse a permis d'élargir les connaissances relatives à l'espace conformationnel de la glycine et aux transformations qui y sont liées. De plus, la réponse électronique de ces systèmes soumis à des collisions avec des vents stellaires ou des particules de rayons cosmiques a été simulée. Grâce aux nouveaux outils théoriques qui ont été implémentés et de leurs extensions, de nouvelles perspectives sont ouvertes pour l'étude théorique de l'évolution chimique des COMs induite par l'irradiation et des phénomènes d'ionisation.

Contents

Glosssary

A

ADFT	Auxiliary-Density Functional Theory
AIBs	Aromatic Infrared Bands
ALMA	The Atacama Large Millimeter/submillimeter Array
AMBER-FF99SB	Set of parameters for AMBER Force Field Potential
AOs	Atomic orbitals
aug-cc-PVTZ	Augmented correlation consistent Triple - Zeta basis set

B

BIO	mio DFTB Sets of Parameters
BFGS	Broyden, Fletcher, Goldfarb, Shanno Optimization Frame
BO	Born-Oppenheimer Approximation
BOMD	Born-Oppenheimer Molecular Dynamics

C

CC	Coupled Cluster methods
CCSD(T)	Coupled Cluster Single Double (Triple) excitations
CI	Configuration Interaction methods
CM3	Class IV/Charge Model 3 charge representation
COMs	Complex Organic Molecules

CRs	Cosmis Rays
cc	cis cis (angles)
ct	cis trans (angles)

D

deMon2k	Quantum Chemistry package based on ADFT method
deMonNano	Quantum Chemistry package based on DFTB method
DIBs	Diffuse Interstellar Bands
DFT	Density Functional Theory
DFTB	Density Functional based Tight-Binding Theory
DFTB0	Zero order DFTB method
DFTB2	Second order DFTB method
DFTB3	Third order DFTB method
DFT/MM	DFT/Molecular Mechanics hybrid method
DFTB/MM	DFTB/Molecular Mechanics hybrid method
DZVP	LDA Double - Zeta Polarization Basis Set

E

ESP	Electronic Stopping Power

F

FF	Force Field approach

G

GGA	Generalized Gradient Approximation

H

HDA	High Density Amorphous ice
HF	Hartree-Fock approach

I

IR	Infrared
ISM	Interstellar Medium

J

JWST	James Webb Space Telescope

L

LCAO	Linear Combination of Atomic Orbitals
LDA	Low Density Amorphous ice
LDA*	Local Density Approximations
LJ	Lennard - Jones potential
LR-TD-DFT	Linear Response - Time Dependent - DFT
LR-TD-DFT	Linear Response - Time Dependent - DFTB
LvN	Liouville-von Neumann
LSO	Löwdin Symmetric Orthogonalization Method

M

MAT	matsci[1] DFTB set of parameters
MD	Molecular Dynamics
MM	Molecular Mechanics approach
MOs	Molecular Orbitals
MP	Magnus Propagator
MP*	Møller - Plesset approach

N

NAMD	Non - Adiabatic Molecular Dynamics
N-Gly	Neutral glycine
NVE	Microcanonical ensemble
NVT	Canonical Ensemble under constant Temperature conditions
NVP	Canonical Ensemble under constant Pressure conditions

O

OPLS-AA	Set of parameters for Force Field potential OPLS
SCRF	Onsager Self - Consistent Reaction Field Method

P

PAHs	Polycyclic Aromatic Hydrocarbons
PBE96	Hybrid Functional for the Exchange - Correlation in DFT
PES	Potential Energy Surface

Q

QM	Quantum Mechanics Approach
QM/MM	Quantum Mechanics/Molecular Mechanics hybrid method

QM/MM-pol	Polarizable Force Fields
QMD	Quantum Molecular Dynamics

R

RMSQ	Forces Maximum Root Mean Square
RKS	Restricted Kohn-Sham SCF tolerance
RT-TD-DFT	Real Time - Time Dependent - DFT
RT-TD-DFTB	Real Time - Time Dependent - DFTB

S

SCC-DFTB	Self - Consistent - Charge - DFTB
SCF	Self - Consistent - Field
SDC	Steepest Descent method
SOMP	Second-Order Magnus propagator
STO	Slater Type Orbitals

T

TD-Schrödinger	Time Dependent - Schrödinger equation
TI-Schrödinger	Time Independent - Schrödinger equation
TIP3P	3-Site Rigid water model
tt	trans trans (angles)
tc	trans cis (angles)

U

UFF	Universal Force Field
UV	Ultraviolet Radiation

V

VUV	Vacuum Ultraviolet Radiation

X

XC	Exchange - Correlation Energy in DFT
XUV	Extreme Ultraviolet Radiation
***xy*-Gly-nH_2O**	Glycine isomer immerse in water environment
***xy*-Gly-LDA**	Glycine isomer immerse in LDA ice environment

Z

Z-Gly	Zwitterionic glycine
Z_D-Gly	Zwitterionic glycine from direct proton transfer mechanism
Z_C-Gly	Zwitterionic glycine from concerted proton transfer mechanism

CHAPTER 1

Complex Molecules in the Interstellar Medium

Understanding the observable Universe and its composition has been one of the greatest goals in Astrophysics. Advances in astronomical observations, experimental and theoretical laboratory studies have given an insight regarding the astronomical compounds, their formations and their possible evolution. In the past decades, researches regarding the formation, distribution and behaviour of complex molecules in space have been gaining importance as they are believed to play a key role in the appearance of life on Earth. The main hypothesis lies in the idea that icy-grains trapping complex molecules agglomerate into large-sized ices. These interstellar ices, ultimately evolve into comets in the forming-planetary systems and arrived on the early Earth constituting **the precursors of the origins of life on Earth**.

In this PhD work, we will focus on the understanding of complex molecules immersed in interstellar ices or in their isolated forms. This thesis will be mostly devoted to the study of **glycine prebiotic molecule immersed in an interstellar ice**. Glycine is one of the simplest amino acid present in several biological systems and has already been detected twice in the coma of two different comets.

In addition, due to their major presence in the Interstellar Medium (ISM), the polycyclic aromatic hydrocarbons (PAH) in their isolated form are also going to be investigated here.

In this introduction chapter, a brief astrochemical background is reported. Details on the interstellar ices, glycine prebiotic molecule and PAH compounds are presented. Furthermore, a brief introduction on the high energy collision processes occurring in the ISM is also reported. This phenomena will be later modelled in this work. The objectives and the outline of this thesis are finally presented.

1.1 The Galactic Matter Cycle

In the known Universe, a great part of the visible matter is constituted by galaxies where 90% of the mass is localized in stars. The remaining 10% is the so-called **Interstellar Medium (ISM)** which is the region between stellar systems made of gas, dust, and ices.[2]

Two regions can be distinguished within the ISM:

- The **diffuse ISM**, consisting of hot and low dense clouds with highly ionized gas.
- The **dense ISM** formed by the eventual contraction of diffuse clouds. These dense clouds are also called **condensed molecular clouds**. They are made of **dust** (silicates, iron and carbon-rich materials) and **gas** (helium and hydrogen molecules). Dust may clump to form, **grains**. Due to the decrease in temperature, these grains eventually will be covered by a thin layer of ice, constituting **icy-grains**.
  In these regions, small **complex molecules** may exist in gas phase or trapped in the icy-grains, in particular polycyclic aromatic hydrocarbons **(PAH)** and **Complex Organic Molecules (COMs)** (such as CO, H_2O, CN).

Here it should be noted that different COMs' definitions can be found in literature. In the context of this thesis, the astrobiology definition was chosen, establishing that COMs are any molecule constituted by carbon, nitrogen and oxygen atoms with sizes in between 2 to 13 atoms.[3–5]

More details about COMs and PAHs can be found in Section 1.3.

It is also important to mention that in these ISM regions high energy irradiation processes such as **comic rays (CRs)** exist. These processes are expected to play an important role in the chemistry of formation and evolution of the previously mentioned molecular systems. A brief review of irradiation processes is detailed in Section 1.5.

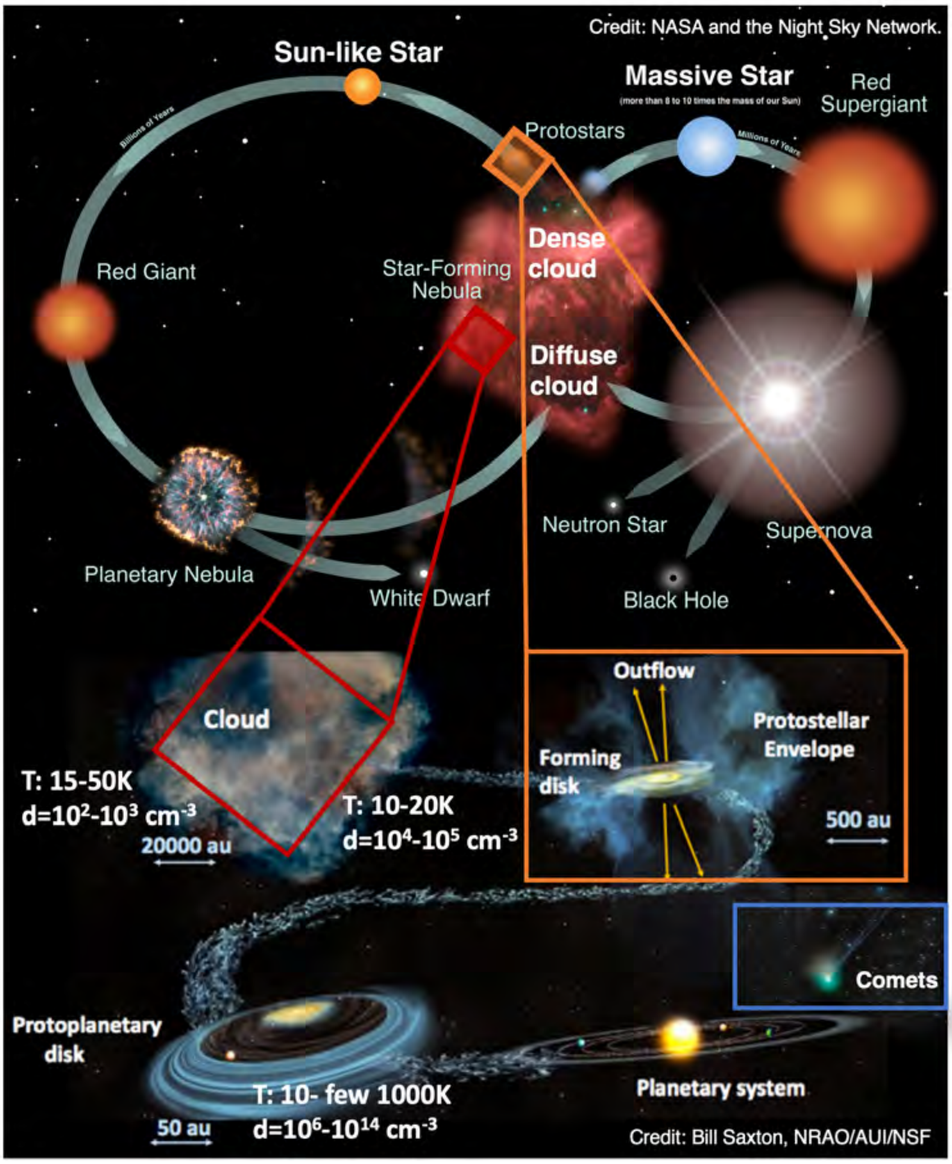


Figure 1.1: Scheme of various stages in the star life cycle combined with the planetary formation sketch. The dense and diffuse ISM clouds are illustrated, along with the physical conditions (temperature and density) in several space regions. Density is express in hydrogen atoms per cm^{-3} (Upper panel image-Credit: NASA and the Night Sky Network. and bottom panel-Credit: Bill Saxton, NRAO/AUI/NSF with modifications)[6]

The eventual collapsing of condense molecular clouds yields to **star formation**. At the same time, the residual material spiraling around the core of the star flattens, forming an acceleration disk (or circumstellar disk). This young star gets nourished from the swirling gas, dust and icy-grains, and reshapes into a newborn star named **protostar** and a diffuse disk called **protostellar envelope**[7] (see Figure 1.1-bottom panel). The young star will continue to absorb the matter in the disk while in the outside disk envelope, where the temperatures are lower, icy-grains continue to agglomerate forming bigger entities of **ices**.[8]

The protostar will continue to sustain it-self by the matter in the disk, while the dust particles and ices may continue to clump in the outside disk envelopes. Thus, **protoplanets in a protoplanetary disk** are formed (see Figure 1.1-bottom panel).[7] Moreover, in the lower temperature side of the protoplanetary disk (10 K to few 1000 K) ices can be shaped covered by a thicker mantle of several nm mainly constituted by water[7](see Figure 1.1-bottom panel). In this ice mantles **larger COMs** can be formed such as CO_2, NH_3, CH_3OH, etc.

From the protoplanetary disk, protoplanets will continue to evolve and will establish their orbits. The star will stop flaring up, reaching the final stages of the planetary system formation (such as our Solar System). Furthermore, ices, dust and rocks still agglomerate forming the so-called **comets**[7,8] (see Figure 1.1-bottom panel). It is important to note that comets can contain previously mentioned molecular systems such as COMs. There exists one main hypothesis that lies in the idea that COMs trapped in ices, ultimately evolve into comets and could arrive to the early planet constituting **the precursors of the origins of life on Earth**. This hypothesis sets the COMs as possible key precursors of prebiotic compounds.

In all previously mentioned stellar environements, energy irradiation processes such as **stellar winds** can appear contributing to the chemical evolution of interstellar ices and complex molecules in space (see more details in Section 1.5).

Finally, the star life cycle will continue to evolve until the star's death. At this last stage, matter (molecules, dust, and interstellar ices) will eventually feed the diffuse ISM. This might diffuse and dense molecular clouds that may collapse and form new stars (see Figure 1.1-upper left panel). Hence, the star and the matter life cycle gets restarted.

In the following, both icy-grains (in the dense ISM molecular cloud) and big ices (in the protoplanetary disk) will be referred as **interstellar ices**. They exist in different sizes, compositions and structures. More details are reported in Section 1.4.

Observations, laboratory experiments and modeling of complex molecules in space are rapidly evolving topics in Astrochemistry. In the past several decades, the idea that the origin of life can be traced back to the star formation has seen its importance growing.[9] Many efforts have been devoted to the better understanding of complex molecules immerse in interstellar ices or in their isolated forms. **One of the main motivation of this thesis relies on searching to learn more about this interstellar ices trapping complex molecules and to understand their evolution in the astrophysical conditions.**

1.2 Observational Astrochemistry

In the last decades, great advances in observatory detection of molecules have been achieved thanks to the space telescopes such as Hubble (HST), FUSE, Spitzer, Herschel, ISO and the latest one, the James Webb Space Telescope (JWST).[10] Each device operates at different or overlaping wavelengths providing absorption and emission spectra covering a spectral range from the Far-UV to the Far-Infrared spectra. On the other hand, observations can also be done remotely with Earth telescopes, such as the NOEMA, IRAM or the recent revolutionary telescope, The Atacama Large Millimeter/submillimeter Array (ALMA).

From the first detection of molecules in space (CH, CN and CH^+)[11–13] in 1937, a long list from simple to more complex molecules have been reported. From NH_3,[14] H_2O and formaldehyde first detections,[15,16] up to 256 complex molecules have been detected in the ISM. In the last decade, the access to the cold dark molecular clouds was possible, thanks to ALMA observations. In TMC-1 (a cold pre-stellar core), a carbon-rich chemistry has been detected. It is believed that carbon-chains, radicals, ions and neutral cyanopolyynes could be formed in this region.[17,18] A complete review of identified complex molecules can be found in the reference.[19]

Last observations with JWST confirmed the presence of CO, OH, CH^+ and what is thought to be the characteristic of CO_2 absorbed in a interstellar ice.[20,21] Furthermore, the **presence of CH_3^+ cation** in a protoplanetary disk in the Orion bar was reported for the first time.[22] This is the first detection ever done on this important molecule, thought to be one of the very first precursors to trigger gas phase organic chemistry. Moreover, it is believed to be one of **the building blocks of amino acids such as glycine**[23–26] (more detail in Section 1.3.1). From this detection, it was suggested that this gas phase organic chemistry may be activated by the UV irradiation from a massive star close to the Orion bar.[22]

Finally, many **detections of molecules in comets** have been done, using absorption spectroscopy (*in situ* and remotely), as well as mass spectroscopy techniques. This is the case of the Rosetta mission,[27,28] aiming the detection of amino acids in comets. Up to 60 different COMs were observed and **the first detection of glycine molecule** was achieved.[27] Another great detection from JWST is the presence of water vapor around the Comet 238P in the Solar System's asteroid belt. This latter indicates that the water ices of the primordial solar system can be preserved in that region.[29]

In following sections, specific details on space observations focusing on glycine prebiotic molecule, PAH and interstellar ices are provided. Along with space observations, a support from the development and assessment of astronomical systems is enable, thanks to the experimental and theoretical studies.

1.3 Complex Molecules in the ISM: Glycine and PAHs

Complex molecules in the ISM such as COMs or PAHs, have been the long-sought astronomical species. Still many questions remain to be answered regarding their formation, their behaviour in/on interstellar ices or their evolution under astronomical sources of irradiations. **This thesis aiming to understand the behaviour of interstellar ices trapping complex molecules and their evolution in the ISM will target two specific complex molecular systems: the glycine prebiotic molecule and the PAHs molecular systems.**

1.3.1 Glycine: A Prebiotic COM

Several evidences have shown that the formation of **glycine (NH_2CH_2COOH)**, the simplest amino acid, could occur in space. Indeed, it has been discovered in the Murchison meteorite,[30] **detected in samples of cometary dust** from comet 81P/Wild 2 returned by the Stardust mission,[31] observed in the coma of the Churyumov- Gerasimenko comet by the ROSINA mass spectrometer (Rosetta mission)[32] and very recently detected in the Ryugu samples collected by the Hayabusa2 mission.[33] Gas-phase glycine has been widely searched in interstellar sources.[34] A first detection was claimed in 2003[35] but it was refuted in 2005[36] and up to now, glycine has not been detected in the ISM. However, several precursors of glycine molecule have been already detected such as methylamine,[37] formamide,[38,39] glycolonitrile,[40] aminoacetonitrile[41] and cyanomethanimine.[42] In addition, a very recent work reports the detection of **the first isomer of glycine in the ISM**, the glycolamide ($NH_2C(O)CH_2OH$).[43]

From the standpoint of astrochemical modelling, the formation of glycine in space has been investigated considering both gas phase[44] and solid state chemistry.[45] In the laboratory, many studies on glycine in gas phase or embedded in pure or mixed water ices (as models of interstellar ices) have been undertaken. Most of them were devoted to either the formation of glycine or to the investigation of its stability with respect to external sources of radiation.[26,37,46–52] However, only few of them focused on the conformations of the formed glycine.[26,37,53]

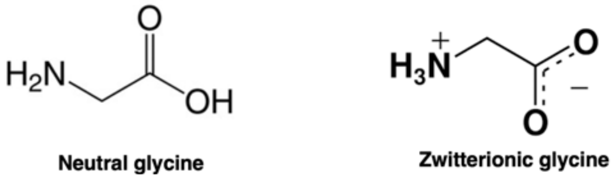


Figure 1.2: Formulation for neutral and zwitterionic glycine

Glycine is known to exist in its **neutral form (N-Gly) (NH_2CH_2COOH)** (Figure 1.2-left panel) in gas-phase or in low temperature inert matrices[53,54] and in its **zwitterionic form (Z-Gly) ($NH_3^+CH_2COO^-$)** (Figure 1.2-right panel) in water solution or in glycine crystals. Following the work of Maté *et al.*,[53] the transformation from neutral to zwitterionic glycine is restricted in non-polar environment, whereas the embedding of neutral glycine in polar (water) medium favors its conversion to Z-Gly even at low temperature of 25 K.[53] Holtom *et al.*[37] reported the formation of zwitterionic glycine in methylamine–carbon dioxide ices at T=10 K and anionic isomers ($NH_2CH_2COO^-$) upon heating to 300 K and cooling down to 10 K. In a more recent study,[26] the formation of N-Gly in a water-rich ice at 13 K without external energetic stimuli has been confirmed by in-situ reflection absorption infrared (RAIR) spectroscopy. Furthermore, upon heating to 230 - 240 K, N-Gly is converted only partially to its Z-Gly form. These different transformation behaviours of glycine in water-rich icy environment at low temperature could, therefore, depend not only on the specific ice composition and temperature,[53] but also on the formation mechanism that may preferentially stabilise particular glycine conformer(s) over others.

Neutral glycine conformers in gas-phase are defined with respect to the three internal rotational degrees of freedom around glycine C-C, C-O and C-N bonds, resulting in eight conformers (see Figure 1.3).[55] Among them the lowest in energy is the trans-trans-trans (*ttt*) conformer[56] followed by the second most stable cis-cis-cis (*ccc*) conformer, higher in energy by only 1.4 kcal/mol than the ttt conformer.[57] In water medium and at room temperature, the small relative energy differences between the conformers are overcomed, resulting in an ensemble of conformations.[58]

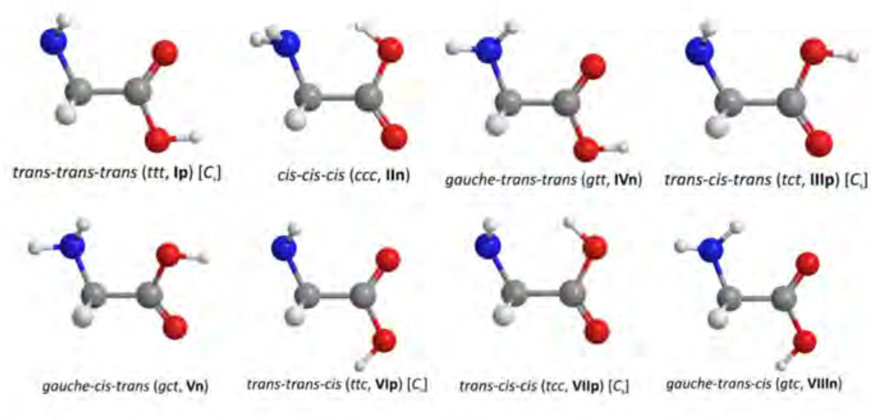


Figure 1.3: Illustration of glycine conformers respect to the three internal rotational degrees of freedom.[55]

To conclude, from the experimental investigations on the relative energy stability and the mechanism of tautomerization from neutral to zwitterionic conformers, quantum mechanics (QM) chemical methods appears as a great complementary tool. Several references of theoretical studies have been reported in literature[55, 58–60, 60–70] and will be presented in detail in Chapter 4.

1.3.2 Polycyclic Aromatic Hydrocarbons (PAH)

Polycyclic Aromatic Hydrocarbons (PAHs) are complex compounds with a carbonaceous skeleton made of a combination of 6-member carbon atom rings terminated at their borders by hydrogen atoms (see examples in Figure 1.4). The presence of a delocalised π resonant system renders PAHs molecules special stability enforcing the structures to be planar. Sometimes they can be called hydrogenated graphene nanoflakes[71] as they can be seen as small pieces of graphene layer.

PAHs molecular systems detections have been widely studied since past decades. Allamandola *et al.*[72] and Léger and Puget's[73] attributed the Aromatic Infrared

Bands (AIBs) (between 3.3 - 20 μm) to PAHs vibrational spectra. JWST improves the spatial/spectral resolution in particular regarding the 3.3 μm neutral PAHs grain emission band.[21,74]

The broadening of the AIBs bands are probably related to the existence of a large number of species. In addition, thermal effects and anharmonicity are also expected to be at the origin of this band broadening.[75–79] This broadening makes it challenging to identify a single precise molecule.[80] Nevertheless, several detections have been done in IR vibrational or rotational spectra, such as the benzene molecule in the enveloppe of the C-rich CRL 618 planetary nebulae.[81] The cyano-naphthalene ($C_{10}H_7$–CN)[82] and indene (C_9H_8)[83,84] have been also observed both in the TMC-1 molecular cloud. Furthermore, in primitive meteorites small and medium size PAHs have been detected.[85,86] A family closely related to PAHs compounds, the so-called fullerenes such as C_{60},[87,88] C_{70},[87] C_{60}^{+}[88,89] have been observed and it is believed that PAHs could be the possible precursors of the latter compounds.[90,91]

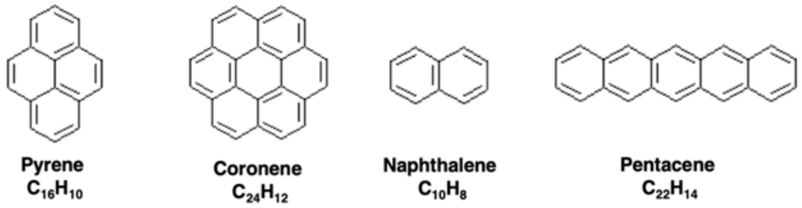


Figure 1.4: Illustration of several PAHs size and conformations.

Many efforts have been devoted to the understanding of formation and evolution of PAHs compounds.[92] It has been suggested that the circumstellar envelopes around the C-rich forming star could be one of the regions were PAH start to be formed.[93]

Several investigations regarding the formation, growth and behaviour of PAHs under irradiation processes are reported in the following references.[19,94,95] It is stressed in these studies, the remarkable stability of PAHs and their supposed survival under irradiation. Experimentally, many studies have been motivated to understand the static and dynamical[96–100] properties of PAHs in intestellar conditions. Furthermore, many experimental infrared and ultraviolet spectra have been measured for neutral and charged PAH compounds as reviewed in references.[101,102] In addition, a complete theoretical study investigating the spectral properties of PAHs have been carried out, collecting all results in the database developed by Malloci. *et al.*[103]

To conclude, along with the detection and experimental works, many modellisation studies of PAHs have been done.[104–107] In the MAD group from Université Toulouse III - Paul Sabatier and CNRS, the theoretical study of PAHs and related compounds have been widely addressed: from the study of anharmonic infrared spectrocopy,[108,109] researches in the isomerisation pathways[77,100,110–112] or the dissociation and fragmentation investigations.[113,114]

1.4 Interstellar Ices

As introduced in Section 1.1, interstellar ices are believed to be formed in the dense ISM clouds and in the cold areas of the protoplanetary disks. Interstellar ices can exist in many different compositions, sizes and structures. Indeed, the ice mantle recovering the grains in the ISM dense clouds are mainly formed by H_2O, as well as CO, CO_2, H_2CO, CH_3OH and NH_3 species (see Figure 1.5). Despite the challenging task that is the identification of ices' compositions, several detection have given a hint on the molecular abundances in ices.[115–117] As expected, water was found to be the predominant species in the ice mantle. Furthermore, JWST detection had shown the presence of water vapor around the Comet 238P in the Solar System's asteroid belt, indicating that the water ices are preserved in this region.

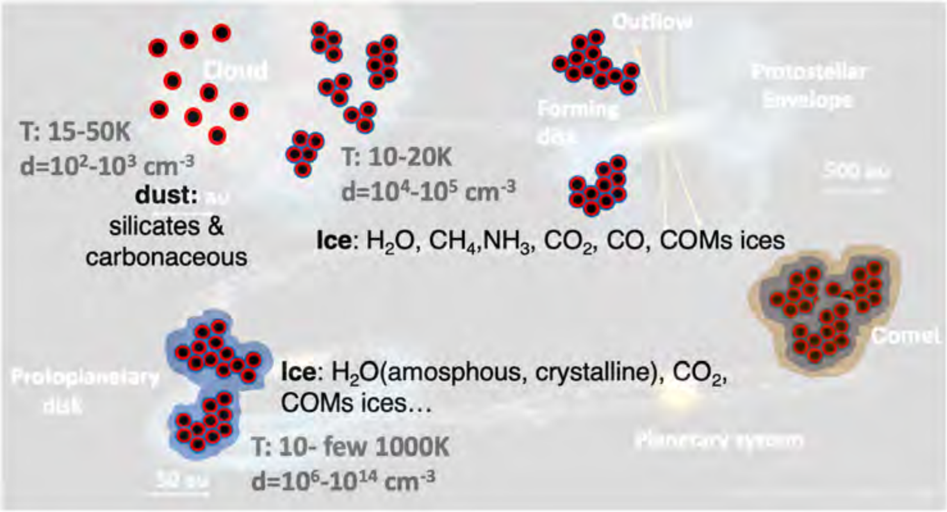


Figure 1.5: Illustration of interstellar ice model. Density is express in hydrogen atoms per cm^{-3}

In this thesis work, we will focus only on the water type interstellar ices and therefore only details about this latter are reported in the following.

Experimental studies have been carried out in order to properly understand ices' formation and structures.[118,119] Three main types of interstellar water ices exist regarding the organization of the water molecules and the temperature (only low pressure ice are considered):[120]

- **Hexagonal ice** is a crystalline ice stabilized at temperatures higher than 170 K.
- **Cubic ice** is a crystalline ice stabilized between 130 and 170 K.[121]
- **Amorphous ice** where two ices phases exist. The **Low Density Amorphous Ice (LDA)** and **High Density Amorphous Ice (HDA)**. The first LDA structure is only obtained at temperatures lower than 60 K. It is thought that for higher temperatures, porous structures will start to collapse constituting the HDA ice.[122] From temperatures higher than ~170 K the LDA ice organization is not maintained.[123]

The conversion from amorphous to crystalline ice appears with increasing temperature. Their Far-IR profiles are characteristic and unique. In addition, the IR studies provide information about the composition of this condense phase. **In the context of this thesis, the model of interstellar ice in the ISM will be restricted to water type ice with a LDA ice organization.**

It is important to recall the catalytic character of ices in the chemistry of the ISM. As it was previously mentioned, the photochemical or thermal processes might enhance the reactions and growth of complex molecules in the ISM. A detailed review about the prebiotic astrochemistry in interstellar ices can be found in reference.[124]

1.5 High Energy Irradiation Processes in the ISM

As previously mentioned, irradiation processes in space play an important role in the formation, chemistry and survival (fragmentation) of complex molecules. Let us first make a distinction between radiation and irradiation process. The term **radiation** refers to a process where a certain energy is radiated from a certain source. The term **irradiation** refers to any process where the energy radiated incidents on/in a certain object. This latter process will be referred as **collision (or bombardment) processes** hereafter. During collision, the colliding particle (consider to be any matter particle or photons) will deposit some energy in the targeted object. The amount of **deposited energy** depends on the charge, velocity and composition of the colliding particle, as well as the composition and dimension of the bombarded object. Lastly, the term **energy stopping power** can be found also in literature referring to the ability of any object to slow down the particle traveling through itself.[125]

From the different multiwavelength observations and models, a list of several radiation fields (with different colliding energy values) have been reported: cosmic rays ($\sim 10^3$ eV to 10^{21} eV), vacuum ultraviolet radiation (VUV: 6.2 - 12.4 eV), extreme ultraviolet radiation (XUV: 12.4-124 eV), X-rays ($>$ 124 eV), γ-rays (100k eV - 10^{12} eV) and stellar winds ($\sim 10^3$ eV up to 10^6 eV).[5,126] Each of the mentioned radiation fields have different compositions and different formation sources. For example, in the case of **cosmis rays (CRs)** (90% protons and 10% heavy nuclei), the sources of formations are believed to be the star-formation region, the circumstellar disks, the core-collapse supernovae and/or the strong fast winds in a young massive star. Eventually, these high-energy CRs will penetrate into the ISM molecular clouds providing ionization and heating processes. During these processes they will decelerate and decrease their kinetic energy value. This latter results in the so-called **low-energy CRs**.[127] Moreover, UV radiation is believe to appear in the ISM due to CRs secondary effects.[5] Another example of irradiation are **stellar winds**, consisting of electrons and protons initially ejected from the upper atmosphere of the stars. The irradiation energies of these stellar winds will depend on the star's life time, *e.g* $\sim 10^3$eV in the case of the solar winds (Sun's stellar winds)[126] or much higher in the case of young stars.

It is obvious that astrophysical dust, grains, ices and complex molecular compounds presented before (see Sections 1.3 and 1.4) are exposed to all kind of the previously mentioned irradiations. These collision processes play an important role in **the space chemistry (astrobiochemistry)**,[126] *e.g* COMs and amino acid formation in grains or interstellar ices, COMs' energy stopping power, extreme UV excitation of PAH or fragmentation processes. Note that in any of these mentioned processes, highly excited molecules may be formed.

Experimental and theoretical studies on COMs irradiation processes have been reported. For instance, the gas-phase chemistry in the ISM enhanced by high energy collision processes have been reviewed.[5,128–130] Radiolysis studies in/on interstellar ices driven by the effect of CRs or solar winds have also been reported.[126,128,131] Furthermore, recent studies captured the ultrafast ($\sim$as or 1 fs) relaxation processes of complex molecules under high energy collision processes.[51,52] Maclot *et al.*,[51] combined an experimental and theoretical study where a doubly ionized glycine amino acid was bombarded by a highly charged Xe^{25+} ion (387.5 keV), monitoring the ultrafast intramolecular hydrogen migration.

In the case of PAHs, many experimental and theoretical researches showing for instance the photoexcitation or photoionization processes in PAHs[132–137] have been performed. Furthermore, the study of PAHs behaviour under collisions processes by VUV, XUV, stellar winds or CRs, in circumstellar and dense cloud environments, have gained a great interest.[99, 138–144] For example, the ultrafast relaxation of PAHs under a high XUV energy collision processes were monitored thanks to the advances on the femtosecond pump-probe experiments.[99, 143, 144]

The study of collision processes is expected to be one of the key elements to understand the formation and evolution of astrophysically relevant complex molecules. **In this thesis, we address the evolution of glycine and PAHs under the effect of stellar winds and CRs irradiation.**

1.6 Theoretical Considerations

From the previous sections, it can be concluded that a collaborative work between observations, experimental laboratory and theoretical models, can provide a better understanding on the physics and chemistry of astrophysical environments. Observations give a wide list of molecules detected in space. The experimental research gives a great insight on the possible reactions paths and the product formation. Finally, theoretical works appear as complementary tools for the understanding of those reactions at the atomistic scale and their molecular properties unavailable by experiments.

Quantum Chemistry techniques have been rapidly developed in past several decades. A large number of chemical properties can be derived from solving the time dependent- and time independent- Schrödinger equation. Nevertheless, the resolution of these many-body problem for systems bigger than the diatomic hydrogen quantum system, becomes challenging and demands great computational resources. In these quantum leap, one breakthorugh in the field was reached, thanks to the development of the Hartree-Fock (HF) approach. This latter simplifies the exact many-body problem in a many one-electron problem.

From the HF and other several approximations three main families of computational methods exists: wave function based methods, electron density based methods and force field type methods. Depending on the size of the systems, the computational resources and the chemical properties that are expected to be studied, a specific computational method must be chosen. For example, the glycine molecule has a total of 10 atoms or 30 valence electrons. The calculation of some of the integrals by means of the wave function based methods becomes very complex if not unfeasible due to their scaling factor (between $O(n^5) - O(n^8)$ n being the number of electrons[145]). In all upcoming studies, all calculation will be performed by electron density based methods, in particular, the Density Functional Theory (DFT) and the Density Functional based Tight-Binding Theory (DFTB).

Furthermore, in order to overcome several computational limitations due to the size of the molecular systems treated and the studied processes, several code implementations have been carried out.

Note that, in this section, we mentioned several quantum electronic methods. A complete overview is reported in detail in Chapter 2.

1.7 Objectives and Outlines

COMs formation and behaviour in the ISM (in particular glycine prebiotic molecule) are believed to play a key role in the appearance of life on Earth. In this PhD work, we aim to understand the behaviour of glycine prebiotic molecule immerse in an interstellar ice under different conditions relevant for the ISM. The reported experimental and theoretical works have contributed to valuable advances on the stability and tautomerization of glycine conformations in water medium. However the accumulated knowledge is still incomplete (see Section 1.3.1) and it does not apply directly to the glycine conformational behaviour in the interstellar ice media. Hitherto, glycine tautomerization has been studied exclusively in view of its biological relevance at room temperature.[58,63,68,146,147]

Note that we will not study the formation of glycine molecule. We will assume that glycine has been already formed in the dense ISM clouds or in the protoplanetary disk (following the COMs chemistry suggested in Section 1.3.1). **In this thesis work, we address the questions of glycine conformational space and the related transformations, including tautomerization mechanisms, in ice-water media and different temperatures relevant with respect to the interstellar ice conditions.**

Does the glycine molecule mantains its original structure under the effect of the water-ice (LDA) environment and the effect of temperature?

The relevance of the high energy collision processes on the astrochemsitry of the ISM has been highlighted also in the previous paragraphs. Despite the experimental and theoretical studies presented in Section 1.5, the modeling of the ultrafast phenomena remains challenging and the knowledge remains insufficient, in particular, for glycine molecule under interstellar media conditions. **In this thesis, we address the questions of the explicit ultrafast electron dynamics of glycine molecule embedded in an interstellar ice under high energy solar winds collision.**

What is the electronic evolution of glycine molecule induce by a collision?

In the case of PAHs species, a similar study is performed with the goal to understand **the ultrafast electron dynamics of PAHs under high energy particle collision.** Note that we will study PAHs in their gas phase as they are believed to be found in the ISM, protostellar or cirscumstellar regions mostly in their isolated forms. We will choose a set of linear PAHs and studying their behaviour under CRs collisions.

In this thesis we will use DFT and DFTB levels of theory employing deMon2k[148] and deMonNano[149] software packages respectively. Note that due to the computational limitations, DFT calculations in some occasions were replaced by DFTB ones. In order to achieve the goals mentioned above, two new implementations were needed in the deMonNano package. In first place, the modelisation of glycine molecule trapped in an interstellar ice becomes challenging due to the large number of atoms present in this system.

For these reason, DFTB quantum method will be combined with class 1 molecular mechanics approached in a QM/MM hybrid scheme.

Furthermore, the simulation of the ultrafast processes induced by high energy collisions requires to go beyond time independent - Schrödinger problem and to model the explicit electron dynamics.

For this reason, the Real Time - Time Dependent - DFTB approach (RT-TD-DFTB) will be implemented in deMonNano in combination with the DFTB/MM class 1 approach.

This latter combination of DFTB/MM class 1 and RT-TD-DFTB implementations fulfill the last goal of this thesis, understanding the ultrafast evolution process of glycine molecule trapped in an interstellar ice (large size system) under collision. This attains also one of the objectives of the ANR-RUBI project (*Reactivité Ultrarapide des Biomolecules sous Irradiation*) which this thesis takes part in.

To conclude, we presente the outline of the chapters in this manuscript:

- **Chapter 2**. A detailed report on the methods and computational tools in Quantum Chemistry are presented.
- **Chapter 3**. The new DFTB/MM class 1 implementation is reported along with several benchmark studies.
- **Chapter 4**. The conformational study of glycine molecule trapped in a water-ice environment is detailed.
- **Chapter 5**. The Real Time -Time Dependent -DFTB implementation along with several benchmark studies are reported. The first applications of this model to simulate high energy collisions on glycine molecular systems and PAH are detailed.
- **Conclusions and perspectives** are listed regarding the implementations and the applications.

Complementary to this outline, Figure 1.6 points out the time scales addressed in each of the chapters.

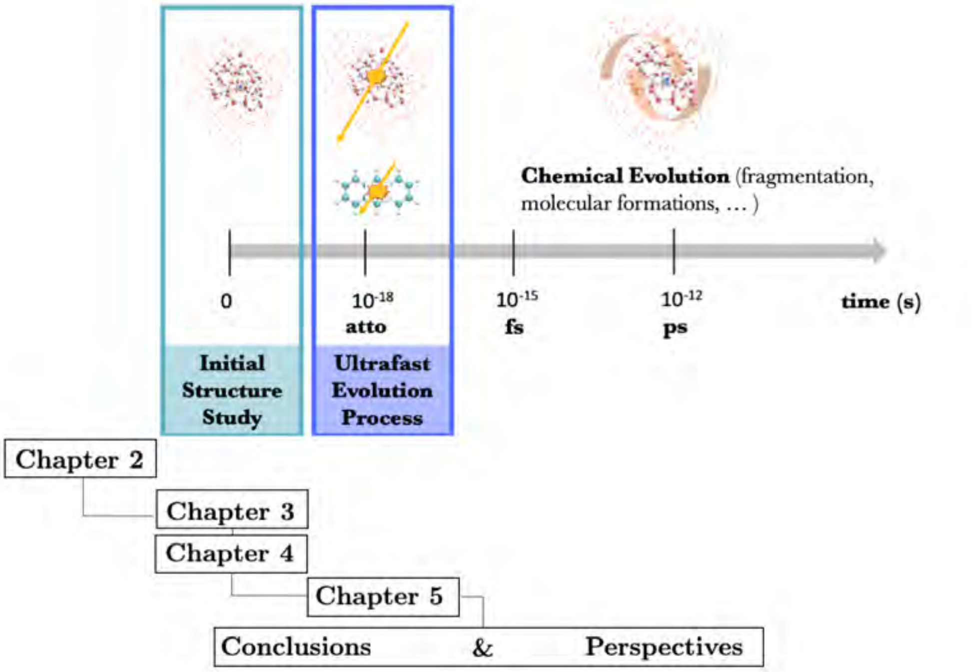


Figure 1.6: Chapters outline with the support of the time scales and molecular systems studied in this thesis work. Note that each chapter tag refers to the process situated in the upper part.

CHAPTER 2

Methods and Computational Tools

Quantum mechanics bursts in the XXs century from a need of a proper description of atomic and subatomic particles' behaviour beyond its classical mechanics description. Planck's study of light emission quantization, Mileva Marić mathematical development and Einstein's statement of light behaviour's both particle-like and wave-like, marked the beginning of quantum mechanics. Furthermore, de Broglie's suggestion of the wave aspect of electron motion, and Heisenberg uncertainty principle set the foundation of quantum mechanics.[150] Quantum chemistry aims to learn about the chemical properties of a system making use of quantum mechanics. Nowadays, computational chemistry make use of computational resources to solve quantum chemistry problems.

An introduction to the Schrödinger equation and its handy Born-Oppenheimer approximation are presented in this chapter. In addition, an overview of the main electronic structure computational methods aiming to solve this quantum problem are listed. Further details on the exploration of system's potential energy surfaces and the simulation of their dynamical evolution in time are described.

2.1 Schrödinger equation and Born-Oppenheimer approximation

From classical mechanics, the motion of a particle is governed by Newton's second law ($\mathbf{F} = m\mathbf{a} = md^2\mathbf{R}/dt^2$), holding on the idea of the **state** of a system (reassembling the specifications of position and velocity of each particle at a certain instant time, plus the forces acting on it). From this law, it is possible to find the future state of a classical-mechanical system from the knowledge of its present state.

Similarly in quantum mechanics, a **state function** or **wave function $\mathbf{\Psi}$** can be defined as a function of particles' coordinates and time $\mathbf{\Psi}(\mathbf{r}, \mathbf{t})$. It is important to point out that, while in classical mechanics a state function is defined locally, in quantum mechanics, the state function is defined everywhere in space. The evolution in time of this wave function is determined by the **time dependent Schrödinger equation**:[150]

$$i\hbar\frac{\partial\mathbf{\Psi}(\mathbf{r}, t)}{\partial t} = \hat{H}\mathbf{\Psi}(\mathbf{r}, t) \tag{2.1}$$

The constant $\hbar$ refers to $\hbar = h/2\pi$ with h being Planck's constant ($h = 6.62607015 \cdot 10^{-34}$ J s) and $\hat{H}$ is the Hamiltonian operator, associated to the system's total energy. For an atomic or molecular system, consisting of nuclei and electrons, defined by $\mathbf{R}$ and $\mathbf{r}$ coordinates respectively, the previous equation can be rewritten as:

$$i\hbar\frac{\partial\mathbf{\Psi}(\mathbf{R}, \mathbf{r}, t)}{\partial t} = \hat{H}\mathbf{\Psi}(\mathbf{R}, \mathbf{r}, t) \tag{2.2}$$

In the case of a **stationary state system**, *i.e* when properties do not vary in time, the spatial and time variables in the Schrödinger equation can be treated independently, leading to the **time- independent Schrödinger equation** :

$$\hat{H}\Psi(\mathbf{R}, \mathbf{r}) = E\Psi(\mathbf{R}, \mathbf{r}) \tag{2.3}$$

Where E corresponds to the total energy (**eigenvalue**) of the system.

The evolution of the state wave function (**eigenfunction**) is given by:

$$\mathbf{\Psi}(\mathbf{R}, \mathbf{r}, t) = e^{-i\frac{Et}{\hbar}}\Psi(\mathbf{R}, \mathbf{r}) \tag{2.4}$$

Neglecting any spin-orbit coupling or any other relativistic effect, the $\hat{H}$ molecular

Hamiltonian is :

$$\begin{aligned}\hat{H} &= \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ne} + \hat{V}_{ee} \\ &= -\frac{\hbar^2}{2}\sum_a \frac{1}{m_a}\nabla_a^2 - \frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 \\ &+ \sum_a \sum_{b>a} \frac{Z_a Z_b e^2}{4\pi\varepsilon_0 r_{ab}} - \sum_a \sum_i \frac{Z_a e^2}{4\pi\varepsilon_0 r_{ia}} + \sum_j \sum_{i>j} \frac{e^2}{4\pi\varepsilon_0 r_{ij}}\end{aligned} \tag{2.5}$$

a and b refers to the nuclei and i and j to the electrons. The first two terms in Eq. 2.5 are the nuceli and electron kinetic energy operators respectively, with m_a the nuclei's mass and m_e the electron mass. ∇^2 is the Laplacian operator refering to the second differentiation with respect to the corresponding nuclei or electron coordinates. The third term is the nuclear repulsion potential energy, where Z_a and Z_b are a and b nuclear atomic numbers, e the electron charge, and r_{ab} the distance between a and b nuclei. The furth term is the nuclear-electron attraction potential energy, expressed as a function of r_{ia}, the distance between electron i and nuclei a. Lastly, the fifth term is the potential energy for the electron-electron Coulomb repulsion where r_{ij} is the distance between i and j electrons. In the following, the Hamiltonian terms in Eq.2.5 will be expressed in atomic units ($\hbar = m_e = 1/4\pi\epsilon_0 = e = 1$):

$$\hat{H} = -\frac{1}{2}\sum_a \frac{1}{m_a}\nabla_a^2 - \frac{1}{2}\sum_i \nabla_i^2 + \sum_a \sum_{b>a} \frac{Z_a Z_b}{r_{ab}} - \sum_a \sum_i \frac{Z_a}{r_{ia}} + \sum_j \sum_{i>j} \frac{1}{r_{ij}} \tag{2.6}$$

Born-Oppenheimer Approximation

It is well known that the complexity to solve the time-independent Schrödinger equation (Eq. 2.3) increases with the number of degrees of freedom of the system. It is then convenient to reformulate the total Hamiltonian expression (Eq. 2.6) as a sum of two independent Hamiltonian operators:

$$\begin{aligned}\hat{H} &= \hat{H}_e + \hat{H}_n \\ &= \hat{H}_e + \hat{T}_n + \hat{V}_{nn}\end{aligned} \tag{2.7}$$

All terms depending purely on the nuclear coordinates are gathered in **the nuclear Hamiltonian** $\hat{H}_n$, while the operator terms depending on the electronic and mixed electron and nuclear coordinates are all gathered in **the electronic Hamiltonian** $\hat{H}_e$. The latter one is given by:

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_a \sum_i \frac{Z_a}{r_{ia}} + \sum_j \sum_{i>j} \frac{1}{r_{ij}} \tag{2.8}$$

In many cases, due to the differences in masses ($m_a >> m_e$), electrons are considered to move faster than nuclei. At the electronic timescales, electrons are considered to move in a field of fixed nuclei having the time to adapt to this field, reaching a stationary state. Accordingly, all nuclear kinetic ($\hat{T}_n$) and Coulomb interaction ($\hat{V}_{nn}$) terms in Eq. 2.7 can be left apart and the electronic problem can be solved independently. The possible **stationary electronic states (k)** can be obtained by solving the time independent- Schrödinguer equation (2.3) for the electronic system only:

$$\hat{H}_e \psi_e^k(\mathbf{r}_i; \mathbf{R}_a) = E_e^k(\mathbf{R_a}) \psi_e^k(\mathbf{r}_i; \mathbf{R}_a) \tag{2.9}$$

For a given nuclear configuration ($\mathbf{R_a}$), this electronic Schrödinger equation can be solved, obtaining **all the stationary electronic state k, the corresponding electronic energy $\mathbf{E_e^k}(\mathbf{R}_a)$ and the electronic wave function $\psi_e^k(\mathbf{r}_i;\mathbf{R}_a)$.**[150] Both, will consequently depend parametrically on the nuclear coordinates ($\mathbf{R_a}$), meaning that for each nuclear arrangement, there will exist an electronic set of E_e^k and ψ_e^k. For the sake of simplicity, in this text $E_e^k(\mathbf{R}_a)$ and $\psi_e^k(\mathbf{r}_i;\mathbf{R}_a)$ will be referred as E_e^k and $\psi_e^k(\mathbf{r}_i)$.

We note that the **total wave function** $\Psi(\mathbf{r}, \mathbf{R})$ of a quantum system can always be expressed as the sum over all electronic states k of the product of electronic eigenfunctions $\psi_e^k(\mathbf{r}_i)$ and the nuclear wave functions $\chi_n^k(\mathbf{R}_a)$

$$\Psi_{\text{tot}}\left(\mathbf{r}_i, \mathbf{R}_a\right) = \sum_k \psi_e^k\left(\mathbf{r}_i\right) \chi_n^k\left(\mathbf{R}_a\right) \tag{2.10}$$

An approximated total wave function solution is obtained by keeping only one electronic state contribution (commonly the electronic ground state contribution $k = 0$). This is the **Born-Oppenheimer (BO) approximation**. The total wave function for the **electronic ground state level** reads :

$$\Psi_{\text{tot}}\left(\mathbf{r}_i, \mathbf{R}_a\right) = \psi_e^{k=0}\left(\mathbf{r}_i\right) \chi_n^{k=0}\left(\mathbf{R}_a\right) = \psi_e\left(\mathbf{r}_i\right) \chi_n\left(\mathbf{R}_a\right) \tag{2.11}$$

Lastly, **the total energy** for a quantum system **in the BO approximation** is given by an effective Hamiltonian:

$$\hat{H}_{\text{efft}} = \hat{T}_n + U(\mathbf{R_a}) \tag{2.12}$$

where

$$U(\mathbf{R_a}) = \hat{V}_{nn} + E_e(\mathbf{R_a}) \tag{2.13}$$

From the effective potential definition $U(\mathbf{R_a})$ in Eq. 2.13 it is assumed that, at each nuclear movement, electrons are adapting themselves **adiabatically and instantaneously** to the electronic ground state. The nuclei moving on a **potential energy surface (PES)** can be obtained by solving the electronic problem ($\hat{H}_e$ and E_e). It has to be clarify that Born-Oppenheimer approximation is only used for **electronic states energies that are non degenerate**. In the cases where electronic states are close in energy values, other approaches that are more sophisticated must be used to solve the Schrödinger equation.

In the upcoming chapters, unless precised otherwise, only the stationary ground state electron problem is taken in to consideration, thus the electronic ground state energy E_e needs to be calculated. Any subindex e referring to the electronic part, or k, will be drop out. Only at the latter Chapter 5, the electronic state value will be upgraded to the **electronic excited states ($k > 0$)** and to the time dependent frame (beyond the Born-Oppenheimer approximation).

2.2 Variational method, Spin Orbital and Antisymmetry Principle

The Variational method

The solution for the time-independent Schrödinger equation comes through the the obtention of the proper wave function (Ψ). From the **variational principle** stating that *any approximate wave function has an energy above the exact one*, the solution for the energy problem is held by tracking the proper wave function that minimizes the energy of the system[151] .

The energy of a *well-behaved* wave function can be obtained as an expectation value of the Hamiltonian operator :

$$E_e = \frac{< \Psi|\hat{H}_e|\Psi >}{< \Psi|\Psi >} \geq E_0 \tag{2.14}$$

Therefore, obtaining the energy E_e from any system is held by collecting the "best" wave function and its corresponding parameters.[151] The equality between electronic and ground state energy (E_0) holds only for the exact wave function.

Spin Orbital

For a complete description of the electron in Eq. 2.8, there is a need to include not only the spatial coordinates ($\mathbf{r}$) but also to specify the **spin** coordinates (σ). Since an electron is a fermion, its spin quantum number value corresponds to $\pm 1/2$. In a non-relativistic frame, two **spin functions** can be defined: $\alpha(\sigma)$ *spin up* ($\sigma = +1/2$) and $\beta(\sigma)$ *spin down* ($\sigma = -1/2$). It has to be mentioned that spin functions are complete and orthonormal. Therewith, an electron is described by its four coordinates:[152]

$$\mathbf{x} = \mathbf{x}(\mathbf{r}, \sigma) \tag{2.15}$$

An **orbital** refers to any wave function for a single electron particle. From both, spatial and spin electron coordinates, two types of orbitals can be defined. In one hand, the **spatial orbitals** ($\phi(\mathbf{r})$) describe the spatial distribution of an electron as a function of the space vector $\mathbf{r}$. On the other hand, the **spin orbitals** (**one-electron wave function**) can be written as a product of a spatial orbital and a spin function (α or β) describing the electron i :

$$\varphi_i\left(\mathbf{x_i}\right) = \begin{cases} \phi_i\left(\mathbf{r_i}\right)\alpha\left(\sigma_i\right) \\ \phi_i\left(\mathbf{r_i}\right)\beta\left(\sigma_i\right) \end{cases} \tag{2.16}$$

The spatial orbitals are normally considered to be orthonormal (and so for the spin ones):

$$\int d\mathbf{r}\ \phi_i^*(\mathbf{r})\phi_j(\mathbf{r}) =< \varphi_i|\varphi_j > \delta_{ij} \tag{2.17}$$

Antisymmetry Principle and N-electron wave function

Extrapolating from the one-electron ($\mathbf{x_1}$) to the N-electron system, the wave function $\psi_e^k(\mathbf{r}_i)$ is rewritten as: $\Psi = \Psi(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N})$. In an N-electron wave system, the **antisymmetry principle** states that: "*A N-electron wave function must be antisymmetric with respect to the interchange of the coordinates* **x** *(both space and spin) of any two electrons*".[152]

$$\Psi(\mathbf{x_1}, ..., \mathbf{x_i}, ..., \mathbf{x_j}, ..., \mathbf{x_N}) = -\Psi(\mathbf{x_1}, ..., \mathbf{x_j}, ..., \mathbf{x_i}, ..., \mathbf{x_N}) \tag{2.18}$$

A direct consequence from the antisymmetric principle is the **Pauli exclusion principle**: "*Two electrons cannot have all quantum numbers equal*".

When extending the problem to the N-electron systems, the ***N*-electron wave function** is given by the product of many one-electron wave functions (also spin orbitals) ensuring the antisymmetry principle. N electrons $\Psi(\mathbf{x_i})$, is given by the following expression known as the **N-electron Slater determinant**:[152]

$$\Psi\left(\mathbf{x_1},\mathbf{x_2},\ldots\mathbf{x_N}\right) = (N!)^{-1/2}\begin{vmatrix} \varphi_1\left(\mathbf{x_1}\right) & \varphi_2\left(\mathbf{x_1}\right) & \ldots & \varphi_N\left(\mathbf{x_1}\right) \\ \varphi_1\left(\mathbf{x_2}\right) & \varphi_2\left(\mathbf{x_2}\right) & \ldots & \varphi_N\left(\mathbf{x_2}\right) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1\left(\mathbf{x_N}\right) & \varphi_2\left(\mathbf{x_N}\right) & \ldots & \varphi_N\left(\mathbf{x_N}\right) \end{vmatrix} \tag{2.19}$$

Where $(N!)^{-1/2}$ is a normalization factor. The N-electron Slater determinant 2.19 has N electrons occupying N spin orbitals $\varphi_i\left(\mathbf{x_i}\right)$ (also called **molecular orbitals MOs**). The columns in Eq. 2.19 are labeled by the spin orbitals and the rows by the N electrons in the system.

2.3 Electronic Structure Methods

From the above section, it was shown that the electronic energy eigenvalue (E_e) is held by the obtention of the electronic wave function Ψ and the resolution of the Schrödinger equation problem Eq. 2.9. For the case of a diatomic hydrogen quantum system, the Schrödinger problem can still be solved analytically, contrary large polyatomic molecular systems. Depending on the approximations, as well as the type of properties that are looked for to be studied, different kind of approximated **electronic structure methods** exist. Three main families exist eahc of them encompassing several computational methods: **1) Wave function based methods**, **2) Electron density based methods** and **3) Force Field type methods**. Wave function and electron density based methods are often refered as ***ab initio* methods**, although in the latter case some of them include empirical parameters, *e.g.* semi-empirical exchange-correlation functionals. Figure 2.1 depicts the listed three groups, including the principal electronic structure methods belonging to each family. Furthermore, a comparison between the computational simulation time versus the quantum system size (number of atoms) for each group is done.

This figure shows that a compromise has to be done between the quantum system size, the computational simulation time and the level of accuracy. For this reason, the choice of one or another computational method will depend on the problem to be covered. As an example, when investigating the electronic structure properties for a system (*e.g* magnetic or optical properties of a molecule), the wave function based computational method will be preferably employed. On the other

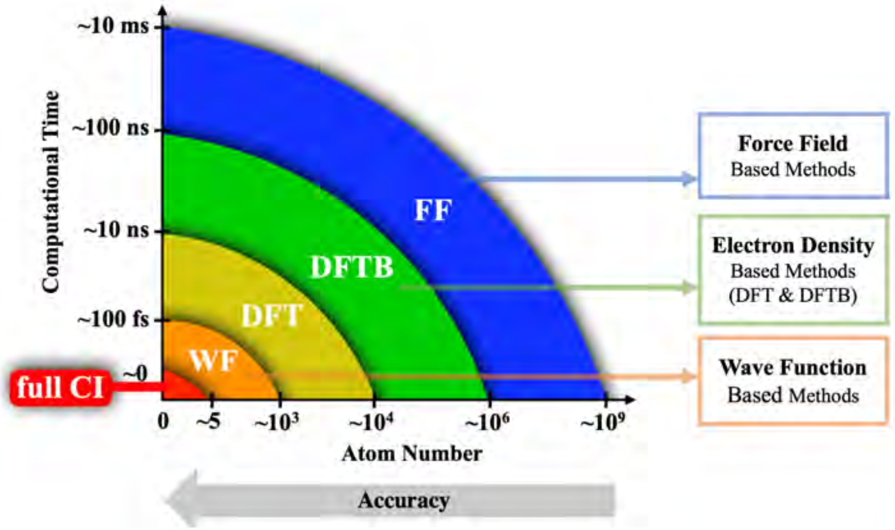


Figure 2.1: Electronic structure computational methods classification. Comparative between the computational time simulations and quantum system size. The *x-axis* refers to the maximum number of atoms accessible for a single point calculation type. The *y-axis* count for the maximal computational time needed approximately for a molecular dynamics calculation for an average size quantum system. From Louisnard[153] with modifications.

hand, when studying macroscopic properties (*e.g* thermodynamical stability or heat capacity of a system), requiring a whole overview of the system's potential energy surface, less accurate methods such as the electron density based or the force field type will be prioritized.

2.3.1 Wave function based methods

Hartree-Fock Method

D.R. Hartree[154] in 1948 gave a first wave function approximation, stating that *the many-electron wave function can be approximated as a product of the one-electron wave functions* (Eq. 2.20), *assuming that each electron feels other electrons as an average (mean-field) potential.*. This leads to the Hartree product:

$$\Psi\left(\mathbf{x_1},\mathbf{x_2},\ldots\mathbf{x_N}\right)=\prod_{i=1}^{N}\varphi_i\left(x\right) \tag{2.20}$$

V. Fock[155] improved Hartree's wave function product by using the proper antisymmetric wave function representation. The **Hartree-Fock (HF) equations** are derived from the application of the variational method to a single Slater determinant (ground state) by the principle of energy minimization. The eigenvalue Hartree-Fock equations are now represented in Eq. 2.21.[152]

$$\hat{F}\varphi_i(\mathbf{x}) = \epsilon_i \varphi_i(\mathbf{x}) \tag{2.21}$$

Where $\hat{F}$ is the Fock operator given by:

$$\begin{aligned} \hat{F} &= \hat{T}_e + \hat{V}_{ne} + v^{HF} \\ &= -\frac{\hbar^2}{2m_e}\nabla^2 - \sum_a \frac{Z_a e^2}{4\pi\varepsilon_0 r_a} + \sum_{j=1}^{N}\left(2\hat{J}_j - \hat{K}_j\right) \end{aligned} \tag{2.22}$$

With Hartree-Fock approximation, the electron-electron repulsion is considered as an average effect. $\hat{J}_j$ and $\hat{K}_j$ are the *Coulomb* and *exchange* operators:

$$\begin{aligned} \hat{J}_j &= \int \varphi_j^*(\mathbf{x}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\varphi_j(\mathbf{x}')\, d\mathbf{x}' \\ \hat{K}_j &= \int \varphi_j^*(\mathbf{x}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\varphi_j(\mathbf{x}')\, d\mathbf{x}' \end{aligned} \tag{2.23}$$

The Hartree-Fock eigen problem (Eq. 2.21) allows to determine the wave function Ψ (or set of MOs) that gives the lowest energy value, with the constraint of the MOs remaining orthonomal. This is a constrained optimization that can be solved by means of **Lagrange multipliers**. Because v^{HF} for an electron i depends on the other electron molecular orbitals (spin orbitals), the Hartree-Fock equations are not linear. For that reason, this equation problem is solved iteratively. This procedure is know as **self-consistent-field (SCF)**. The total minimal energy solutions can be written as:

$$\begin{aligned} E^{HF} &= \sum_i^N \epsilon_i - \frac{1}{2}\sum_{i,j}^N (J_{ij} - K_{ij}) \\ \epsilon_i &= \left\langle \varphi_i(\mathbf{x}) \middle| \hat{F} \middle| \varphi_i(\mathbf{x}) \right\rangle = \left\langle \varphi_i(\mathbf{x}) \middle| \hat{T}_e + \hat{V}_{ne} \middle| \varphi_i(\mathbf{x}) \right\rangle + \sum_j^N (J_{ij} - K_{ij}) \end{aligned} \tag{2.24}$$

with

$$\begin{aligned} \hat{J}_{ij} &= \int \varphi_i(\mathbf{x})\varphi_i^*(\mathbf{x})\frac{1}{|\mathbf{r}-\mathbf{r}'|}\varphi_j^*(\mathbf{x}')\,\varphi_j(\mathbf{x}')\, d\mathbf{x}d\mathbf{x}' \\ \hat{K}_{ij} &= \int \varphi_j(\mathbf{x})\varphi_i^*(\mathbf{x})\frac{1}{|\mathbf{r}-\mathbf{r}'|}\varphi_j^*(\mathbf{x}')\,\varphi_i(\mathbf{x}')\, d\mathbf{x}d\mathbf{x}' \end{aligned} \tag{2.25}$$

A latter approximation is done in order to solve computationaly the wave function problem. Quantum chemistry codes preferably make use of an expansion of MOs ($\varphi_i(\mathbf{x})$) in terms of atomic orbitals basis functions (χ_μ). This is the **Linear**

Combination of Atomic Orbitals (LCAO):

$$\varphi_i = \sum_{\mu}^{M_{\text{basis}}} c_{\mu i} \chi_\mu \tag{2.26}$$

Where χ_μ are usually Gaussian or Slater Type Orbitals (STO) functions centered on the atomic position. Therefore, solving the Hartree-Fock problem consists of finding the coefficients $c_{\mu i}$.

Post **Hartree-Fock methods**

The difference between the exact energy of a quantum chemistry system and its Hartree-Fock energy (Eq.2.24) is the so-called **correlation energy** $E_{corr} = E_{exact} - E^{HF}$. ***Post*** **Hartree-Fock** methods make use of the Hartree-Fock wave function as a starting point, and several approximations are applied to estimate the correlation energy. Three main groups exists depending on how E_{corr} is determined: **1) Møller-Plesset (MP) methods**[156] where the calculation of E_{corr} is based on a perturbative treatment at the second (MP2) or higher order levels;[157,158] **2) Coupled Cluster methods (CC)** making use of an ansatz exponential operator including several Slater determinants for different electron configurations (single CCS, double CCSD and even triple CCSD(T) excitations are treated by a perturbative approach);[159–162] **3) Configuration Interaction methods** describing the wave function of a system as a linear combination of Slater determinants. The exact solution for the Schrödinger equation is given by the *full CI* method, while the *CASSCF* scheme solves the exact Schrödinger equation in a chosen limited orbital space. In addition, the energy calculation can be improved by the use of *CASPT2/NEVPT2* methods, where the energy solution is calculated by a perturbative MP2 type method.

The wave function based methods are often among the most accurate ones, nevertheless they are heavy in terms of computational time. **Electron density based methods** aim to obtain a proper value for the energy correlation, making use of new approximations to minimize the computational time.

2.3.2 Density Functional Theory (DFT)

The **electron density** for an electronic system is defined as *the number of electrons per unit of volume.*

$$\rho(\mathbf{x}) = N\int d_{\mathbf{x}}^{N-1}\Psi^*(\mathbf{x},\mathbf{x}_2,\ldots.\mathbf{x_N})\Psi(\mathbf{x},\mathbf{x}_2,\ldots.\mathbf{x_N}) = N\int d_{\mathbf{x}}^{N-1}|\Psi(\mathbf{x},\mathbf{x}_2,\ldots.\mathbf{x_N})|^2 \tag{2.27}$$

Where $\rho(\mathbf{x})$ integrates to the total number of electrons of the system $\int\rho(\mathbf{x})d\mathbf{x} = N$. In the following, only close shell systems are considered, and so the spin coordinates (index) will be ignored.

Hohenberg and Kohn theorems[163] are considered one of the greatest breakthrough in Quantum Chemistry, setting the foundation of the Density Functional Theory (DFT). They stated that the electronic energy of a system is a function of its electronic density only, leading to the reformulation of the electronic Schrödinger problem as:

$$\begin{aligned}\hat{H}_e &= -\frac{1}{2}\sum_i\nabla_i^2 - \sum_a\sum_i\frac{Z_a}{r_{ia}} + \sum_j\sum_{i>j}\frac{1}{r_{ij}}\\ &= \hat{T}_e[\rho(\mathbf{r})] + \hat{V}_{ne}[\rho(\mathbf{r})] + \int d\mathbf{r}\rho(\mathbf{r})\ v_{ext}(\mathbf{r})\end{aligned} \tag{2.28}$$

The **first Hohenberg and Kohn theorem** (1964) states that the external potential v_{ext} is determined completely by electron density $\rho(\mathbf{r})$. Furthermore, the **second Hohenberg and Kohn theorem** enlarges the first one, proving that the exact ground state density $\rho_0(\mathbf{r})$ minimizes the energy of a system[163](extension of the variation principle in the DFT theory):

$$E[\rho_T(\mathbf{r})] \geq E[\rho_0(\mathbf{r})] = E_0 \tag{2.29}$$

Where $\rho_T(\mathbf{r})$ can be any trial density. With this mapping, the Schrödinger equation problem is translated from a wave function problem to an electron density one.

The Kohn-Sham approach

The essential Kohn-Sham (KS) theory establishes that the ground state density $\rho_0(\mathbf{r})$ of an **interacting system** (submitted to an external potential $v(\mathbf{r})$) can be precisely reproduced in a **non interacting system** submitted to a $v_{KS}(\mathbf{r})$ potential. With this approach, the electronic energy of a non interacting system will be calculated from an auxiliary set of orbitals Φ^{KS}. The total electronic energy of a system $E^{DFT}[\rho(\mathbf{r})]$ is given by:

$$\begin{aligned} E^{DFT}[\rho(\mathbf{r})] &= \sum_{i=1}^{N} -\frac{1}{2}\int d\mathbf{r}\phi_i^{KS*}(\mathbf{r})\nabla^2\phi_i^{KS}(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &+ E_{xc}[\rho] + \int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r}) \\ &= T_{KS}[\rho(\mathbf{r})] + E_H[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] + \int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r}) \end{aligned} \tag{2.30}$$

with

$$\rho(\mathbf{r}) = \sum_i n_i\phi_i^*(\mathbf{r})\phi_i(\mathbf{r}) = \sum_i n_i\phi_i^{KS}(\mathbf{r})\phi_i^{KS}(\mathbf{r}) \tag{2.31}$$

It has to be noted that in the KS approach, the electron density in the first term in Eq. 2.30 is given in terms of Kohn Sham orbitals. $E_H[\rho(\mathbf{r})]$ is the Hartree energy (second term in Eq. 2.30) and $v(\mathbf{r})$ is an external potential. Lastly, $E_{xc}[\rho]$ is **the exchange-correlation energy** containing the differences between the real electron interacting system and the non interacting one.

$$E_{xc}[\rho(\mathbf{r})] = (\hat{T}[\rho(\mathbf{r})] - \hat{T}_{KS}[\rho(\mathbf{r})]) + (\hat{V}_{ee}[\rho(\mathbf{r})] - E_H[\rho(\mathbf{r})]) \tag{2.32}$$

It is important to note that Eq. 2.30 is an exact expression for the total energy. Unfortunately, the exchange-correlation energy functional is an only unknown term in the total DFT energy expression. Founding the proper expression of this functional remains one of the main challenges in quantum chemistry. Several approaches to estimate E_{xc} have been developed, remarking here the four principal groups: **1) Local Density Approximations (LDA)** treating the electron density localy as a uniform electron gas; **2) Generalized Gradient Approximation (GGA)** adding the first derivative of the density; **3) Meta-GGA** including the laplacian of the density besides the first derivative, and **4) Hybrid functionals** containing a percentage of HF exchange correlation and an approximation (semi or total empirical) to the exchange correlation functional in DFT, *e.g.* PBE0,[164] B3LYP,[165–168] etc.

From the general expression of the DFT energy and the E_{xc} approximations, finding the **electronic ground state DFT energy** is done by minimizing $E^{DFT}[\rho(\mathbf{r})]$ (Eq.2.30) making use of Lagrange multipliers ensuring the constraint of orthogonal orbitals (or density matrix idempotency). The energy problem is now reduced to an eigenvalue problem, solving the Kohn-Sham equations:

$$\hat{H}_{KS}\phi_i^{KS}(\mathbf{r}) = \epsilon_i\phi_i^{KS}(\mathbf{r}) \tag{2.33}$$

$$\hat{H}_{KS} = \left(-\frac{1}{2}\nabla_i^2 + \hat{V}_{KS}[\rho(\mathbf{r})] \right) \tag{2.34}$$

$$\begin{aligned} \hat{V}_{KS}[\rho(\mathbf{r})] &= \frac{\delta E_{KS}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} = \hat{V}(\mathbf{r}) + \hat{V}_H[\rho(\mathbf{r})] + \hat{V}_{xc}[\rho(\mathbf{r})] \\ &= v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \end{aligned} \tag{2.35}$$

The potential operator $\hat{V}(\mathbf{r})$ corresponding to the attraction interaction between nuclei and electrons can be also referred as the **external potential** operator. It has to be noted that the eigenvalue Kohn-Sham problem has to be solved self-consistently as the electron density depends on the Kohn-Sham orbitals ($\rho(\mathbf{r}) = \sum_i n_i |\phi_i^{KS}(\mathbf{r})|^2$). The ground state DFT energy reads:

$$E^{DFT}[\rho(\mathbf{r})] = \sum_{i=1}^{N} n_i \left\langle \phi_i^{KS}(\mathbf{r}) \middle| -\frac{1}{2}\nabla_i^2 + V(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \middle| \phi_i^{KS}(\mathbf{r}) \right\rangle + E_{xc}[\rho(\mathbf{r})] \tag{2.36}$$

In practice, the Khon-Sham orbitals are often expressed as a Linear Combination of Atomic Orbitals (LCAO) (Eq.2.26) and the Eq. 2.33 can be rewritten:

$$\sum_a c_{\mu i} \left(\left\langle \chi_\mu \middle| \hat{H}_{KS} \middle| \chi_\nu \right\rangle - \epsilon_i \left\langle \chi_\mu | \chi_\nu \right\rangle \right) = 0 \ ; \quad \forall \nu, i \tag{2.37}$$

$$H_{\mu\nu} = \left\langle \chi_\mu \middle| \hat{H}_{KS} \middle| \chi_\nu \right\rangle \tag{2.38}$$

$$S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle \tag{2.39}$$

$H_{\mu\nu}$ and $S_{\mu\nu}$ are **the Kohn Sham operator and overlap matrix elements** used to solved the Kohn-Sham eigenvalue problem. At present, many DFT-based softwares are available. In this thesis, **deMon2k software package**[169] was employed. deMon2k code uses the Linear Combination of Gaussian type orbitals (LCGTO) and, as a particularity, it eludes the calculation of the four-center electron repulsion integrals by the use of an **auxiliary basis function** for the variational fitting of Coulomb potential.[170–172] This approximation is named as the **Auxiliary Density Functional Theory (ADFT)**.

Time Dependent - DFT (TD-DFT)

Further, in addition to the electronic ground state stationary case considered so far, the access to dynamical properties and the excited states of a quantum system are now achievable at the **time depencendy variable** into the Schrödinger equation problem. The **time dependent Schrödinger equation** (see Eq.2.1 in Section 2.1), for fixed nuclei, can be reformulated in terms of the electron wave function (Eq. 2.40). A time dependent formulation of DFT is established, where the ground stationary state is perturbed and the external potential is modified.[173]

$$i\hbar\frac{\partial\psi_e(\mathbf{r_i},t)}{\partial t} = \hat{H}_e\psi_e(\mathbf{r_i},t) \tag{2.40}$$

Analogous to the Hohenberg Kohn theorems, a similar and unique mapping between **time dependent electron densities** ($\rho(\mathbf{r},t)$) and **time dependent potentials** ($v(\mathbf{r},t)$) was proved by Runge and Gross in 1984.[174] This leads to a new **time dependent Khon-Sham equation** (Eq.2.41) for the MO's:

$$i\frac{\partial\phi_i^{KS}(\mathbf{r},t)}{\partial t} = \left[-\frac{1}{2}\nabla_i^2 + \hat{V}_{KS}[\rho(\mathbf{r},t)]\right]\phi_i^{KS}(\mathbf{r},t) \tag{2.41}$$

In this framework, the time dependent density reads $\rho(\mathbf{r},t) = \sum_i n_i|\phi_i^{KS}(\mathbf{r},t)|^2$ and V_{KS} is a time dependent function from adiabatic approximation.[175,176]

Solving the TD-DFT problem needs some additional approximations. One of the main tools that can be used to simulate the time evolution of the electronic system is the **Real Time - TD -DFT** (see Chapter 5). This is very useful in the case of time dependent atomistic simulations. On the other hand, for a weak electromagnetic oscillating field perturbation, the **Linear Response (LR) approximation** can be applied[173] . The frequency-dependent response properties can be studied. The LR approximation only takes in consideration the **linear components of the density fluctuation**:

$$\delta\rho(\mathbf{r},w) = \int \chi(\mathbf{r},\mathbf{r}',w)\delta\hat{V}_{KS}(\mathbf{r}',w)d\mathbf{r}' \tag{2.42}$$

$$\chi(\mathbf{r},\mathbf{r}',w) = \lim_{\eta\to 0^-}\sum_{k=1}^{N}\sum_{l=1}^{N}(f_k - f_l)\frac{\phi_l(\mathbf{r})\phi_k(\mathbf{r})\phi_l(\mathbf{r}')\phi_k(\mathbf{r}')}{w - (\varepsilon_l - \varepsilon_k) + i\eta}. \tag{2.43}$$

Where $\hat{V}_{KS}(\mathbf{r}',w)$ is the linearized time dependent KS potential and χ is the response function of the non-interacting KS system to the perturbation expressed as a function of the stationary KS orbitals ($\phi_i^{KS}(\mathbf{r})$) (for simplicity no complex conjugation is included in the equation). **M. Casida**[177] proposed a practical solution to solve the above equation problem.

In deMon2k software package, the Linear Response approximation is implemented as well as the RT-TD-DFT approach.[178]

2.3.3 Density Functional based Tight-Binding Theory (DFTB)

The limitations of the *ab initio* wave function and DFT levels of theory are related to the simulated system sizes and the computational requirements. When the systems constituted by a large number of atoms, or where a lot of single point energies are intended to be done, approaches like the **Density Functional based Tight-Binding (DFTB) method** become useful (see Figure 2.1). Like other Tight-Binding methods, DFTB is based on the reduced atomic orbital basis set representation,[179] besides of an amount of parametrized-elements liable for the reduction of the computational time (DFTB i up to ∼3 orders of magnitude faster than DFT).

The **four main approximation** of the DFTB method are listed below, laying out the development of DFTB approach, as well as the guideline of the upcoming subsections:

1. A Taylor expansion perturbation is applied to the E^{DFT} energy equation with $\rho(\mathbf{r}) = \rho_0(\mathbf{r}) + \delta\rho$ (with $\rho_0(\mathbf{r})$ being a reference density).
2. Only valence electrons are treated explicitly.
3. MOs are expressed as a linear combination of atomic valence orbitals.
4. Only one and two-centered integrals are taken in to consideration.

Therefore, from the **first approximation** ① listed above *i.e.* Taylor expansion truncated to the second order with respect to a density perturbation, $\delta\rho = \rho(\mathbf{r}) - \rho_0(\mathbf{r})$, the **total DFT energy** is approximated as:

$$E^{DFT}[\rho_0(\mathbf{r})+\delta\rho] \approx E^{DFT}[\rho_0(\mathbf{r})]+\int \left.\frac{\delta E^{DFT}[\rho_0(\mathbf{r})]}{\delta\rho}\right|_{\rho_0} \delta\rho+\frac{1}{2}\int\int \left.\frac{\delta^2 E^{DFT}[\rho_0(\mathbf{r})]}{\delta\rho\delta\rho'}\right|_{\rho_0} \delta\rho\delta\rho' \tag{2.44}$$

From the combination of the DFT energy expression (Eq. 2.30) in Eq. 2.44, with the definition of Kohn-Sham operator in Eq. 2.35 (the partial derivation of the functional of the DFT energy with respect to the density) and recalling Eq. 2.31, the above equation can be rewritten as:

$$
\left.\begin{aligned}
E = &\int d\mathbf{r} v(\mathbf{r})\rho_0(\mathbf{r}) + E_{xc}[\rho_0] - \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' \frac{\rho_0(\mathbf{r})\rho_0\left(\mathbf{r}'\right)}{|\mathbf{r}-\mathbf{r}'|} \\
&+ \sum_{i=1}^{N} n_i \left\langle \phi_i(\mathbf{r}) \middle| \hat{H}_{KS} \middle| \phi_i(\mathbf{r}) \right\rangle
\end{aligned}\right\} \mathbf{E^{DFTB0}}
$$

$$
\left.+\frac{1}{2}\int\int \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \left.\frac{\delta^2 E_{xc}[\rho(\mathbf{r})]}{\delta\rho\delta\rho'}\right|_{\rho=\rho_0} \right) \delta\rho\delta\rho' \right\} \mathbf{E^{2nd}} \tag{2.45}
$$

The total energy E, contains different energy contributions. In the first version of DFTB,[179] second order and higher order terms are neglected. This is the **zero order DFTB method (DFTB0)**. Later in 1998, Elstner *et al.*[180] introduced the last line in Eq. 2.45 in order to improve the charge balance in heteronuclear systems, yielding to the **Second order DFTB method (DFTB2)**. More details about the derivations of each energy expression are going to be provided in the following paragraphs.

Zero order DFTB approach (DFTB)

The formulation of the first DFTB energy contribution (DFTB) starts with a rearrangement of the nuclear-nuclear repulsion potential and the terms depending on $\rho_0(\mathbf{r})$ only inside a **short-range and repulsive energy term**, *i.e* all terms in the first line in Eq.2.45 are gathered in a energy contribution term named **DFTB energy repulsion** $E_{rep}[\rho_0(\mathbf{r})]$.

In order to express the second line in Eq. 2.45, the Linear Combination of Atomic Orbitals is used. The atomic orbitals in DFTB are obtained from atomic DFT calculations limiting the basis to a valence shell[181] (**approximations ② and ③**):

$$
\phi_i = \sum_{\mu}^{M_{\text{basis}}} c_{\mu i}\chi_\mu \tag{2.46}
$$

where $\chi_\mu = \chi_\mu\left(\mathbf{r} - \mathbf{R_a}\right)$ are atomic orbitals. The DFTB energy expression is written as:

$$
E^{DFTB0} = E_{rep}[\rho_0(\mathbf{r})] + \sum_{i=1}^{N} n_i \sum_{\mu\nu}^{M} c_{\mu i}c_{\nu i} \underbrace{\left\langle \chi_\mu \middle| \hat{H}_{KS} \middle| \chi_\nu \right\rangle}_{H^0_{\mu\nu}} \tag{2.47}
$$

Finding **the minimal $\mathbf{E^{DFTB0}}$** energy leads to a secular equation problem, that is solved making use of Lagrangian multipliers (deriving with respect to the coefficients

c_i) with the orthonormal constraint for the molecular orbitals:

$$\sum_{\nu} c_{\nu i} \left(H^0_{\mu\nu} - \epsilon_i S_{\mu\nu} \right) = 0 \ ; \quad \forall \mu, i \tag{2.48}$$

The **ground state DFTB0 energy** is given by Eq.2.47 as:

$$E^{DFTB0} = E_{rep}[\rho_0(\mathbf{r})] + \sum_{i=1}^{N} n_i \epsilon_i \tag{2.49}$$

Now that the ground state E^{DFTB0} energy expression has been introduced, the parametrization problem is addressed. We find define the reference density as a frozen atoms density $\rho_0 = \rho^a_0 + \rho^b_0 + ...$ with ρ^a_0 the density calculated for the frozen a atom and ρ^b_0 the frozen atom density for b atom. $\hat{H}^0_{\mu\nu}[\rho_0]$ and $E_{rep}[\rho_0]$ are then computed using the Tight-Binding approach[182] (**approximation ④**) :

1. For $\mu \in a$, $\nu \in a \rightarrow H^0_{\mu\nu} =< \chi^a_\mu \left| \hat{H}_{KS}[\rho^a_0 + \rho^b_0 + ...] \right| \chi^a_\nu >$
   $H^0_{\mu\nu} \approx< \chi^a_\mu \left| \hat{H}_{KS}[\rho^a_0] \right| \chi^a_\nu >= \epsilon^a_\mu \delta_{\mu\nu}$ (**One-body Parm.**)
   ϵ^a_μ atomic energy (DFT calculations at isolated atom a)
   $S_{\mu\nu} = \delta_{\mu\nu}$
2. For $\mu \in a$, $\nu \in b \rightarrow H^0_{\mu\nu} =< \chi^a_\mu \left| \hat{H}_{KS}[\rho^a_0 + \rho^b_0 + ...] \right| \chi^a_\nu >$
   $H^0_{\mu\nu} \approx< \chi^a_\mu \left| \hat{H}_{KS}[\rho^a_0 + \rho^b_0] \right| \chi^b_\nu >= f(\mathbf{R}_a - \mathbf{R}_b)\delta_{\mu\nu}$
   $S_{\mu\nu} = f(\mathbf{R}_a - \mathbf{R}_b)\delta_{\mu\nu}$ (**Two-body Param.**)
3. $E_{rep}[\rho_0(\mathbf{r})] = \frac{1}{2}\sum_{ab}^{K} V^{ab}_{rep}(\mathbf{R}_a - \mathbf{R}_b)$ $V^{ab}_{rep} \equiv$ repulsion potential for a b atoms.

All terms are parametrized from DFT calculations on the isolated pair of atoms (usually LDA or PBE functionals). Note that $E_{rep}[\rho_0(\mathbf{r})]$ is expressed as a sum of diatomic potentials. All energy values and the one and two-center integrals parameters are stored in a external data file (**Slater-Koster tables**) that are posteriorly used for the DFTB energy calculations. The formulation of the DFTB in terms of minimal atomic basis along with the above parametrization recipe are the main keys for the drastic reduction of the computational expenses of the DFTB approach .

Second-Order (Self-Consitent Charge) DFTB (DFTB2)

The second order energy contribution term $\mathbf{E^{2nd}}$ (third line in Eq. 2.45) was introduced by Elstner *et al.*[180] in 1998, to improve the charge balance between atoms.

To compute this term, the formulation of the density perturbation $\delta\rho$ is expressed as a superposition of atomic contributions $\delta\rho^a$ ($\delta\rho = \sum_a \delta\rho^a$). These atomic contributions are expressed as a sequence of radial and angular fluctuations truncated up to the monopole term:[180]

$$\delta\rho^a(\mathbf{r}) \approx \Delta q_a F_{00}^a(|\mathbf{r} - \mathbf{R}_a|)Y_{00} \approx \Delta q_a \frac{F_{00}^a(|\mathbf{r} - \mathbf{R}_a|)}{\sqrt{4\pi}} \tag{2.50}$$

Where F_{00}^a are the normalized radial dependence of the fluctuation charge Δq_a for $m = l = 0$ quantum numbers. Now, the **Second order energy contribution** (**$\mathbf{E^{2nd}}$**) is written as :

$$\begin{aligned} E^{2nd} &= \frac{1}{2}\int\int \overbrace{\left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} + \left.\frac{\delta^2 E_{xc}[\rho(\mathbf{r})]}{\delta\rho\delta\rho'}\right|_{\rho=\rho_0}\right)}^{\Gamma[\mathbf{r},\mathbf{r}\prime,\rho_0]} \delta\rho\delta\rho' \\ &= \frac{1}{2}\sum_{a=1}^{K}\sum_{b=1}^{K}\int\int \Gamma[\mathbf{r},\mathbf{r}\prime,\rho_0]\delta\rho_a(\mathbf{r})\delta\rho_b(\mathbf{r})d\mathbf{r}d\mathbf{r}\prime \\ &= \frac{1}{2}\sum_{a=1}^{K}\sum_{b=1}^{K}\Delta q_a\gamma_{ab}\Delta q_b \end{aligned} \tag{2.51}$$

where

$$\gamma_{ab} = \int\int \Gamma[\mathbf{r},\mathbf{r}\prime,\rho_0]\frac{F_{00}^a(|\mathbf{r} - \mathbf{R_a}|)F_{00}^b(|\mathbf{r}\prime - \mathbf{R_b}|)}{4\pi}d\mathbf{r}d\mathbf{r}\prime \tag{2.52}$$

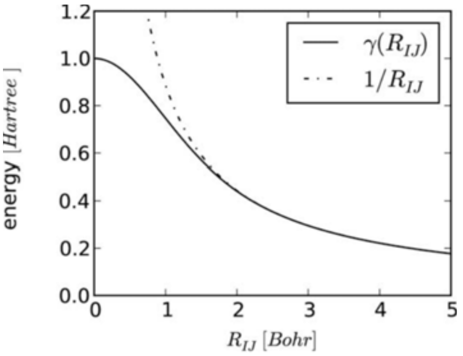


Figure 2.2: Interaction energy for two point charges(I and J are a and b in the text).[183]

When the charges are localized on the same atom, (γ^{aa}) is identified to the **Hubbard parameters ($\mathbf{U_a}$)** ($\gamma_{aa} \approx$ IP$_a$− EA$_a$ $\approx U_a$).[184] For the diatomic term γ^{ab} value will converge to $1/R_{ab}$, in the limit of large interatomic distances, while for short/intermediate interatomic distances, γ^{ab} must converge to the Hubbard parameter. Figure 2.2, illustrates the evolution of the γ term.

Finally, from DFTB energy expression Eq. 2.47 and Eq. 2.51 and 2.52, the expression for **the Second-Order DFTB energy ($\mathbf{E^{DFTB2}}$)** is:

$$E^{DFTB2} = E_{rep}[\rho_0(\mathbf{r})] + \sum_{i=1}^{N} n_i \sum_{\mu\nu}^{M} c_{\mu i}c_{\nu i}H_{\mu\nu}^0 + \frac{1}{2}\sum_{a=1}^{K}\sum_{b=1}^{K}\Delta q_a\gamma_{ab}\Delta q_b \tag{2.53}$$

The charge fluctuations ($\Delta q_a = q_a - Z_a$) are evaluated in terms of the **Mulliken charges approximation**[181] :

$$q_a = \frac{1}{2}\sum_{i=1}^{N} n_i \sum_{\mu\in a}\sum_{b=1}^{K}\sum_{\nu\in b}\left(c^*_{\mu i}c_{\nu i}S_{\mu\nu} + c_{\mu i}c^*_{\nu i}S_{\nu\mu}\right) \tag{2.54}$$

with the total charge of the system given by $\sum_a \Delta q_a = \int \delta\rho(\mathbf{r})d\mathbf{r}$. In the following, charge fluctuations Δq_a will be replaced by the Mulliken charges q_a. From the combination of Eq. 2.54 with Eq. 2.53, followed by a derivation by means of Lagrangian multipliers (similar to the zero order DFTB approach), the secular equation to obtain **the ground state $\mathbf{E^{DFTB2}}$ energy** reads:

$$\sum_a c_{\nu i}\left(H_{\mu\nu} - \epsilon_i S_{\mu\nu}\right) = 0\ ;\quad \forall\mu, i \tag{2.55}$$

$$H_{\mu\nu} = \left\langle \chi_\mu \left| \hat{H}_{KS} \right| \chi_\nu \right\rangle - \frac{1}{2}S_{\mu\nu}\sum_{\xi=1}^{K}\left(\gamma_{a\xi} + \gamma_{b\xi}\right)q_\xi = H^0_{\mu\nu} + H^1_{\mu\nu} \tag{2.56}$$

$$S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle \quad \forall\mu \in a;\ \ \forall\nu \in b \tag{2.57}$$

Due to the DFTB operator dependence on the atomic charges and latter dependency on the MO's coefficients, the new eigenvalue problem cannot be solved only by a single diagonalization (as done for the zero-order DFTB procedure). For this reason, solving the ground state DFTB2 energy problem is done by a **Self-Consistent Charge** approach (standing for the SCC acronym). From a trial set of atomic charges q_a, $H^1_{\mu\nu}$ and $H^0_{\mu\nu}$ are computed and latter diagonalized in order to get a new set of coefficients $c_{i\mu}$ and charges . This is done iteratively until convergency. Finally, the value of the total energy E^{DFTB2} is given by Eq. 2.53. In the following, the second order DFTB method will be referred as DFTB2 although in literature can be also find as **SCC-DFTB**.

DFTB0 and DFTB2 energy gradients

Due to their importance in optimization, dynamic simulations and further implementations in Chapter 3, the analytical expression for the atomic **DFTB0 and DFTB2 forces** are now introduced. They can be computed as the partial derivatives of E^{DFTB0} and E^{DFTB2} with respect to the atoms coordinates ($\mathbf{R_a}$).[180] The

DFTB0 energy gradient is given by:

$$\mathbf{F_a} = -\frac{\partial E^{DFTB0}}{\partial \mathbf{R_a}} = -\sum_{a\neq b}^{K} \frac{\partial V_{rep}^{ab}(\mathbf{R}_a - \mathbf{R}_b)}{\partial \mathbf{R_a}} + \sum_{i=1}^{N} n_i \sum_{\mu\nu}^{M} c_{\mu i} c_{\nu i} \left(\frac{H_{\mu\nu}^{0}}{\partial \mathbf{R_a}} - \epsilon_i \frac{S_{\mu\nu}}{\partial \mathbf{R_a}} \right) \tag{2.58}$$

The last term $(-\epsilon_i S_{\mu\nu}/\partial \mathbf{R_a})$ in Eq.2.58, appears as a direct application of Hellman-Feynman approach for non-orthonormal basis (also known as Pulay stress or force term).

Similarly, the **DFTB2 atomic forces** are given by a slightly more complex expression:[181]

$$\begin{aligned} \mathbf{F_a} = -\frac{\partial E^{SCC-DFTB}}{\partial \mathbf{R_a}} &= -\sum_{a\neq b}^{K} \frac{\partial V_{rep}^{ab}(\mathbf{R}_a - \mathbf{R}_b)}{\partial \mathbf{R_a}} - q_a \sum_{\xi=1}^{K} \frac{\partial \gamma_{a\xi}}{\partial \mathbf{R_a}} q_\xi \\ &+ \sum_{i=1}^{N} n_i \sum_{\mu\nu}^{M} c_{\mu i} c_{\nu i} \left[\frac{H_{\mu\nu}^{0}}{\partial \mathbf{R_a}} - \left(\epsilon_i - \frac{H_{\mu\nu}^{1}}{S_{\mu\nu}} \right) \frac{S_{\mu\nu}}{\partial \mathbf{R_a}} \right] \end{aligned} \tag{2.59}$$

Many different software packages are available to perform DFTB calculations (*e.g* DFTB+,[185] ADF,[186] Amber,[187] Gromacs,[188] Gaussian,[189] DFTBaby,[190] CP2K[191]). The one employed in this thesis is the **deMonNano code**.[149] The set of DFTB parameters can be downloaded from *www.dftb.org*. Two main sets of parameters exists, **matsci**[1] (in this work will be named **MAT**) developed mainly for Material Science applications and **mio**[180](in this work called **BIO**) firstly parametrized for biological systems.

Some limitations and extensions of the conventional DFTB approach

Due to the drawbacks already present in DFT method and to several limitations intrinsic to the DFTB approach, new extensions and improvements in the conventional DFTB method have been done :

- A poor representation of the weak electron interaction is observed with the conventional DFT exchange-correlations functionals. This can be solved with an **empirical dispersion energy (London forces) correction**. In the case of deMonNano package, the E^{disp} correction is implemented as:[192]

$$E^{disp} = -\sum_{a}^{K} \sum_{b\neq a}^{K} f_{damp}(R_{ab}) \frac{C_6^{ab}}{R_6^{ab}} \tag{2.60}$$

where f_{damp} is a cut-off function avoiding divergences at short distance.

- DFTB method is build on the Mulliken charge representation, known by its limitations in the proper description of atomic charges. Other definitions of atomic charges are present in literature as the NBO,[193] Bader,[194] EPF[195,196] and the CM3 charge representation. The latter one (from Kalinowski *et al.*[197]) gives a better description of the bond electric dipole and the electrostatic potential. The CM3 expression for charges is given by:

  $$q_a^{CM3} = q_a^{Mull} + \sum_{b \neq a}^{K} (D_{t_a t_b} B_{ab} + C_{t_a t_b} B_{ab}^2) \tag{2.61}$$

  where q_a^{Mull} are the Mulliken charges, $D_{t_a t_b}$ and $C_{t_a t_b}$ are empirical parameters determined for each atomic-type pair of non identical atoms and B_{ab} is the Mayer's bond order.[198]

  Another alternative is the weighted Mulliken (WMull) atomic charge approach introducing a bias in the sharing of the interatomic density matrix elements:[199]

  $$\chi_\mu(r)\chi_\nu(r) \simeq \frac{1}{2} S_{\mu\nu} \left((1 + t_{\mu\nu}) \left|\chi_\mu(r)\right|^2 + (1 - t_{\mu\nu}) \left|\chi_\nu(r)\right|^2 \right) \tag{2.62}$$

  where $t_{\mu\nu}$ is an empirical value between -1 and 1, recovering the Mulliken charge at $t_{\mu\nu}$ =0.

- As a direct extension of the SCC-DFTB method, **third order DFTB (DFTB3)** approach arises when upgrading the Taylor expansion for the total DFT energy perturbation (Eq. 2.44) up to the third order.[181] The new terms in the total energy expression introduce fluctuations in the Hubbard parameters with respect to the charges.

- The excited states properties for the systems (*i.e* energies, dipole transitions, forces, etc) can be simulated by the **Linear Response - TD - DFTB and Real Time - TD - DFTB**.[200,201] The already mentioned Casida's work[173] in LR-TD-DFT, has been extended by Niehaus *et al.*[202] leading to the LR-TD-DFTB. More details regarding the RT-TD-DFTB approach can be found in Chapter 5.

- DFT suffers from a self-interaction electron problem related with the double counting of electron electron interactions. They are also present in the DFTB approach. This can be fixed by the separation of the Coulomb interaction in short and long-range contributions respectively and adding certain corrections. Few corrections on the long-range interaction has been done in Humeniuk and

Mitrić[203] and Lutsker *et al.*[204] In the case of charge resonance in molecular clusters, self-interaction electron proble can be cured by DFTB-CI scheme.

Force Matching DFTB potentials

Traditionally, $V_{rep}^{ab}(\mathbf{R}_a - \mathbf{R}_b)$ repulsion is parametrized from diatomic dissociation cure. the Based on the work of M. P. Kroonblawd *et al.*,[205] a new recent extension has been implemented in deMonNano package with the aim of parametrizing the $V_{rep}^{ab}(\mathbf{R}_a - \mathbf{R}_b)$ repulsion from DFT.

The total DFTB energy is separated in two terms, the repulsive one and the other contributions collected in a non-repulsive term: $\mathrm{E}^{\mathrm{DFTB/DFTB2}}= E_{rep}[\rho_0(\mathbf{r})] + \mathrm{E}^{\text{NO-REP}}$. The $\mathrm{E}^{\text{NO-REP}}$ energy term are taken from the original matsci or mio DFTB parameters, whereas the $E_{rep}[\rho_0(\mathbf{r})]$ (containing the diatomic contribution V_{rep}^{ab}) is fitted from DFT training sets. In practice, $E_{rep}[\rho_0(\mathbf{r})]$ is parameterized to approach the target forces $\mathbf{F}^*_{\mathbf{rep}}$[205] :

$$\mathbf{F}^*_{\mathbf{rep}} = \mathbf{F}^{\mathrm{DFT}} - \mathbf{F}^{\text{NO-REP}} \tag{2.63}$$

Where the $\mathbf{F}^{\mathrm{DFT}}$ forces come from DFT calculations and $\mathbf{F}^{\text{NO-REP}}$ from the DFTB calculations without including the repulsive contribution for a same set of q trial snapshots. The V_{rep}^{ab} repulsion potential pair is expressed as a power of series of pairwise interactions given by:[205]

$$V_{rep}^{ab}(r_{kl}) = \begin{cases} \sum_{n=2}^{K} c_n^{t_a t_b} \left(r_c^{t_a t_b} - r_{ab})^n\right) & r_{ab} \leq r_c^{t_a t_b} \\ 0 \quad \text{otherwise} & \end{cases} \tag{2.64}$$

where a and b are atoms of type t_a and t_b, $r_c^{t_a t_b}$ is a cut-off distance. $c_n^{t_a t_b}$ are the parameters to be adjusted. Therefore, the problem of parametrizing the V_{rep}^{ab} repulsion energy is now to minimized the difference :

$$\left| \frac{\partial V_{rep}^{ab}(r_{kl})}{\partial \mathbf{R}} - \mathbf{F}^*_{\mathbf{rep}} \right| \tag{2.65}$$

This minimization is carried out by the single value decomposition (SVD) approach.

2.3.4 Force Field type methods

The last computational approach listed in this chapter, corresponds to classical mechanics methods where atoms are treated explicitly but not the electrons. For systems with a large number of atoms ($\sim$ <1000 atoms) or for large simulations time ($\sim <$ 1ns), even the semi-empirical DFTB method can be computational time demanding. For this reason, **Force Field (FF) or Molecular Mechanics approaches (MM)** are used. In this method, **interatomic interactions** are modelled by more or less simple mathematical expressions. In most of cases, the general expression for a FF potential reads:[206]

$$V_{MM} = \underbrace{\sum_{l}^{N_{bonds}} V_l^{bonds} + \sum_{p}^{N_{angles}} V_p^{angles} + \sum_{q}^{N_{torsion}} V_q^{torsion}}_{\textbf{Covalent(interacting) bonding terms}} + \underbrace{\sum_{a}^{N_{MM}} \sum_{b>a}^{N_{MM}} V_{ab}^{Coul} + \sum_{a}^{N_{MM}} \sum_{b>a}^{N_{MM}} V_{ab}^{LJ}}_{\textbf{Non-covalent(non-interacting) bonding terms}} \tag{2.66}$$

Where N_{MM} are the total number of atoms in the system. There are many type of FF representations in literature, with different levels of complexity. In this thesis, the employed ones are the **first class force field potentials** consisting of a first order linear expression for to the energy potentials $V_l^{bonds}, V_p^{angles}, ...$ Bond and angle potentials are usually described as high frequency modes and are modelled by harmonic functions. The **covalent bonding terms** in Eq. 2.66 can be rewritten as:[207]

$$\begin{aligned} V_l^{bonds} &= \sum_{l}^{N_{bonds}} \frac{1}{2} k_l (r - r_0)^2 \quad ; \quad V_q^{torsion} = \sum_{q}^{N_{torsion}} \frac{V_n}{2} [1 + cos(n\beta + \delta)] \\ V_p^{angles} &= \sum_{p}^{N_{angles}} \frac{1}{2} k_p (\theta - \theta_0)^2 \end{aligned} \tag{2.67}$$

The **bond stretching term** V_l^{bond} involves a k_l spring constant usually estimated from Raman or infrared spectra, an equilibrium covalent bond distance r_0 evaluated by X-ray diffraction experiments and a bond distance (r) which varies during simulation. Similarly, the **angle bending term** (V_p^{angle}) consists of a constant k_p and an equilibrium bending angle θ_0. On the other hand, torsion motions are less stiff compared to bonds or angles ones. They are commonly represented by periodic functions. In Eq. 2.67, this $V_q^{torsion}$ term depends on a torsion angle β, a phase δ and n, *i.e.* the number of of minima and maxima between 0 and 2π. V_n determines the height of the rotational potential barrier[207] . The $V_q^{torsion}$ expression is used indistinctly for both proper and improper torsion. Note that a proper torsion refers to atoms that are successively connected while, an improper torsion, atoms connectivity is not mandatory. Figure 2.3 (a,b,c,d) illustrates the different FF potential modes formulated in equations 2.66 and 2.67. In some codes, as is the case of de-

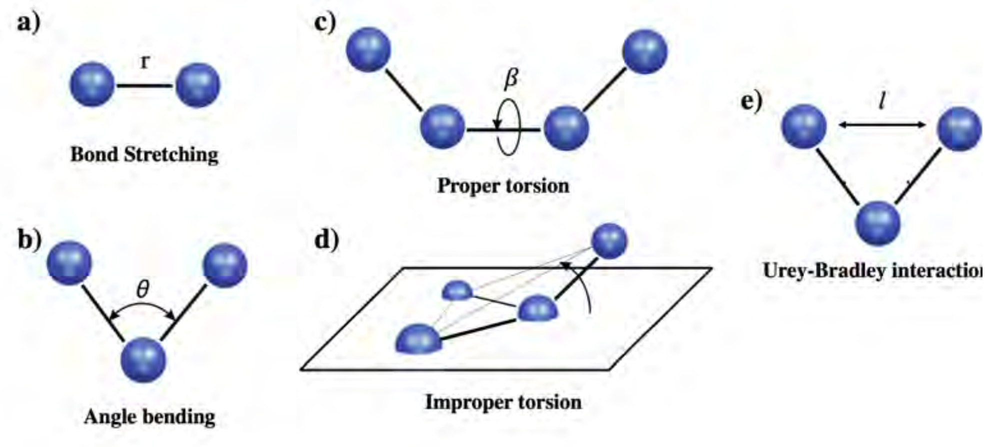


Figure 2.3: Schematic representation for covalent term in Eq.2.67. a) Bond bending mode V_a^{bonds} for a bond distance r, b) angle bending mode V_b^{angles} for a bending angle θ, c) and d) for the proper and improper torsion modes. In addition, the extra Urey-Bradley term incorporated in the general expression FF is depicted in e). Image from thesis work[208] with modifications.

Mon2k, extra terms are added to the general expression of FF in order to optimise the vibrational fitting approximations. This is the case of the **Urey-Bradley term** potential interaction very similar to the stretching bond term, despite that r_0 and r are replaced by l_0 and l. l_0 and l are the equilibrium and bond distances between two atoms, both being at the end of an angle[169] (Figure 2.3.e):

$$V_{UB} = \sum_{b}^{N_{bonds}} \frac{1}{2} k_{UV} (l - l_0)^2 \tag{2.68}$$

Regarding the **non-covalent bonding terms**, the two terms in Eq. 2.66 correspond to the **Coulomb pairwise electrostatic interactions** and the **Van der Waals interactions**. The Coulomb potential between atoms a and b is given by:

$$V_{ab}^{Coul} = \sum_{a} \sum_{b} \frac{q_a q_b e^2}{4\pi\varepsilon_0 r_{ab}} \tag{2.69}$$

with r_{ab} being the interatomic distances and ε_0 the dielectric constant.

Van der Waals interactions are described as a combination of a short-range Pauli repulsion interaction and a long-range dispersion attraction usually modelled by a **Lennard-Jonnes potential**:

$$V_{ab}^{LJ} = 4 \sum_{a>b}^{N_{MM}} \epsilon_{ab} \left[\left(\frac{C_{12}^{ab}}{R_{ab}} \right)^{12} - \left(\frac{C_{6}^{ab}}{R_{ab}} \right)^{6} \right] \tag{2.70}$$

with C_{12}^{ab} and C_{6}^{ab} being respectively the repulsion and attraction parameters depending on the type of atoms a and b. The ϵ_{ab} diatomic parameter can be obtained as an average (arithmetic or geometric) of atomic-pair attraction interactions (well depth parameter).[169]

The constants (k_l, k_p, V_n, n, Q_a, C_{12}^{ab}, C_{6}^{ab} ...) or bond and angle values (r_0, θ_0, δ, ...) are all parametrized and collected in a set of **force field potential files**. So far, several types of FF potentials, *e.g.* OPLS-AA,[209] AMBER, AMBER-SF99,[206] etc, have been parametrised and can be found as a set of databases. In the case of the deMon2k code, the force field data sets are stored in an external (FFDS) file. These empirical parameters are usually valid for a particular electronic ground state at a given covalent structure. This is the primarily limitation of this kind of methods, where the simulation of chemical reactions, *e.g* bond breaking or bond formation, may become challenging. The appearance of hybrid **Quantum Mechanics (QM) / Molecular Mechanics (MM) hybrids methods** allow to describe properly the chemical processes (by QM) while dealing with a large number of atoms (treated by MM). A more detailed explanation of QM/MM hybrid method can be found in Chapter 3.

2.4 Potential Energy Surface (PES) exploration

As mentioned in Section 2.1, when Born-Oppenheimer approximation can be applied, nuclei are stated to move on a multi-dimensional PES, this energy being the electronic energy computed from the time independent-Schödinger equation for electrons. Figure 2.4 shows a schematic picture of a PES energy surface as a function of the nuclear coordinates ($E = f(\mathbf{R})$). It exists many forms and chemical quantities to express the nuclear coordinates, *e.g* interatomic distances, angles, coordination number, etc. In this thesis, the coordinates $\mathbf{R}$ are represented in the form of Cartesian coordinates. For a N atoms finite size system, the number of degrees of freedom is $3N-6$, corresponding to 3 x, y, z Cartesian coordinates per atom minus 6 to remove global translational and rotational modes[210] .

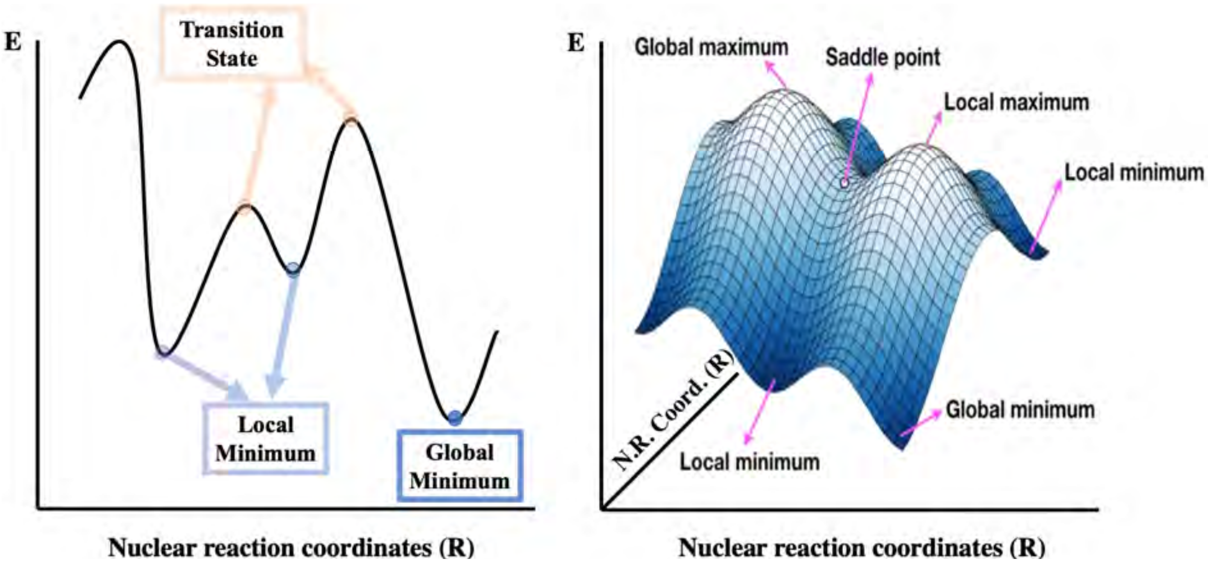


Figure 2.4: Schematic representation for a energy surface as a function of nuclear reaction coordinates for a N atoms system. The principal stationary points for the interest of the PES study are noted in the figure [From Zheng work with modiffications.[211]

2.4.1 Energy optimization techniques

Many problems and physico-chemical properties in computational chemistry are solved by the understanding of the PES function. Therefore, many informations about the physico-chemical properties, molecular structures, chemical reactivity or vibrational spectra can be extracted from the exploration of the PES and the identification of its **stationary points** (minimum, saddle points, ...). Different tools exist in computational chemistry for this purpose and are presented in the following paragraphs.

The stationary points are the sites of the PES where its energy first derivatives with respect to the nuclear coordinates are equal to zero ($\partial E/\partial \mathbf{R} = 0$). Usually the most searched stationary points in the PES exploration are the **local minima**, corresponding to a first order energy gradient equal to zero and second order energy derivatives greater than zero ($\partial E/\partial \mathbf{R} = 0$ and $\partial^2 E/\partial \mathbf{R}^2 > 0$). These PES sites give **energetically stable molecular configurations**. Many local minima can exist on a PES, but only one exists that gives the lowest energy value, and therefore, the most stable molecular configuration. This is the **global minimum**.

The transition between two local minima is given by a **saddle point or transition state** for which the first energy gradient equals to zero and at least one of the second derivatives is negative ($\partial E/\partial \mathbf{R} = 0$ and at least one R_i so that $\partial^2 E/\partial \mathbf{R_i}^2 < 0$).

In computational chemistry, the main tools to explore locally the PES are called the **local optimization techniques**, where the stationary points are targeted and the energy gradients values analysed, looking for the minimum closest to an initial point. Many local optimization techniques exists *e.g.* steepest descent, conjugated gradient methods, Newton-Raphson method.[210] The optimization techniques used in both deMon2k and deMonNano software packages are now briefly presented. In the case of deMon2k code, the optimization technique is based on quasi-Newton method, which computes an approximated Hessian matrix (matrix storing all second order energy derivatives' elements) updated at each step of the optimization by the Broyden, Fletcher, Goldfarb, Shanno (BFGS)[212–215] frame. This technique avoids several algebra steps during the Hessian matrix computation that can be computationally time demanding. In the deMonNano code, both Steepest Decent and Conjugated Gradients methods are implemented.[210] In both codes, a convergence criterion exist, based on the forces maximum root mean square (RMSQ).

For a more complete inspection of the PES, a **global exploration of the PES** can be performed providing an overview of the stationary points and targeting the global minimum. The main techniques used for the global PES exploration in deMonNano and deMon2k are based on **Monte Carlo approach**[216,217] or **Classical Molecular Dynamics and Parallel Tempering** extentions[218,219] .

2.4.2 Born-Oppenheimer Molecular Dynamics (BOMD)

When the aim of a simulation is to compute the evolution of the whole molecular system structure along time, other techniques like **Molecular Dynamics (MD)** are used. There exists different forms to propagate the nuclear coordinates in time, depending on the specific treatment of the nuclear particle nature. In the framework of this thesis, the performed dynamics simulations were carried out by the **Classical Molecular Dynamics** method, where **nuclei are not quantum anymore but are treated as classical particles** within the Born-Oppenheimer framework. In the following, unless precised otherwise, only Classical Molecular Dynamics will be used. In Section 2.1, it was stated that within the Born-Oppenheimer approximation, nuclei move on a unique potential energy surface. BOMD methods propagate a set of nuclear coordinates $\mathbf{R}$ and velocities on the electronic PES E($\mathbf{R}$)(**trajectory**). The propagation of coordinates and velocities is done by the Newton's second law:[210]

$$\mathbf{F_a} = m_a \mathbf{a_a} - \frac{dE}{d\mathbf{R_a}} = m_a \frac{d^2\mathbf{R_a}}{dt^2} \tag{2.71}$$

where m_a, $\mathbf{a_a}$ and $\mathbf{F_a}$ are the mass, acceleration and forces for the atom a respectively and $\mathbf{R_a}$ are the vector coordinates. Several approaches to solve this set of equations exist, *e.g.* Euler's algorithm[220] , Runge-Kutta method[221,222] , predictor-corrector approach[223] , Verlet simple algorithm,[224] Leapfrog Verlet,[225] and Velocity Verlet algorithm[226] . The latter **Velocity Verlet algorithm** is implemented in both the deMon2k and deMonNano codes.

Giving a set of atoms with positions $\mathbf{R_a(t)}$ at a time t, the propagation of these atoms with a **time step** δt is given by the algorithmic integration:

$$\mathbf{R_a}(t+\delta t) = \mathbf{R_a}(t) + \frac{\delta \mathbf{R_a}(t)}{\delta t}\delta t + \frac{1}{2}\frac{\delta^2 \mathbf{R_a}(t)}{\delta t^2}(\delta t)^2 + \frac{1}{6}\frac{\delta^3 \mathbf{R_a}(t)}{\delta t^3}(\delta t)^3 + ... \tag{2.72}$$

$$\mathbf{R_a}(t+\delta t) = \mathbf{R_a}(t) + \mathbf{v_a}\delta t + \frac{1}{2}\mathbf{a_a}(\delta t)^2 \tag{2.73}$$

$$\mathbf{v_a}(t+\delta t) = \mathbf{v_a}(t) + \frac{\mathbf{a_a}(t) + \mathbf{a_a}(t+\delta t)}{2}\delta t \tag{2.74}$$

where Eq. 2.72 is the Taylor expansion, truncated to the second order and rearranged in terms of velocities and accelerations (Eq. 2.73 and Eq. 2.74). One of the key steps in the Velocity Verlet algorithm is the propagation of velocity $\mathbf{v_a}(t+\delta t)$ as shown in Eq. 2.74. These set of equations are solved in four the steps:

1. Velocities are calculated at the $t+\delta t$ instant:

$$\mathbf{v_a}\left(t+\frac{1}{2}\delta t\right) = \mathbf{v_a}(t) + \frac{1}{2}\mathbf{a_a}(t)\delta t$$

2. From the new velocities, the positions at $\mathbf{R_a}(t+\delta t)$ are calculated as:

$$\mathbf{R_a}(t+\delta t) = \mathbf{R_a}(t) + \mathbf{v_a}\left(t+\frac{1}{2}\delta t\right)\delta t$$

3. From this new coordinates, the new energy potential is calculated and thus the acceleration $\mathbf{a_a}(t+\delta t)$

$$\mathbf{a_a}(t+\delta t) = -\frac{1}{m_a}\frac{dE}{d\mathbf{R_a}} = -\frac{1}{m_a}\frac{d^2\mathbf{R_a}(t)}{dt^2}$$

4. Finally velocities at $t+\delta t$ is given by:

$$\mathbf{v_a}(t+\delta t) = \mathbf{v_a}\left(t+\frac{1}{2}\delta t\right) + \frac{1}{2}\mathbf{a_a}(t+\delta t)\delta t$$

From the algorithm presented above, there is an access to the positions, velocities and forces ($\mathbf{R_a}, \mathbf{v_k}$, $\mathbf{a_k}$). Furthermore, depending on the MD conditions, different types of ensembles can be simulated: **(i) NVE ensemble** where the number of particles, the volume and the energy of the system are kept constant along the whole simulation and the **(ii) NVT ensemble** where the number of particles, the volume and the temperature are kept constant during the calculations. In the latter ensemble, **thermostats** can be employed to maintain the temperature, and therefore maintained the correct velocity distribution. Many types of thermostats exits in literature *e.g.* Langevin thermostat, Stochastic Velocity Rescale, the Nosé-Hoover thermostat, etc. In this work, the employed thermostat was a **Nosé-Hoover** chain, which introduced an extra term in the Hamiltonian to simulate the heat bath.

Although it is out of the scope of this thesis, many alternatives to the classical BOMD approach exists. When nuclei are not treated as classical but as quantum particles, dynamic simulations can be performed by **Quantum Molecular Dynamics (QMD)** methods.[227] In addition, if nuclei cannot be treated separately from the electrons, the BO approximation is not valid anymore, and is replaced by **Non Adiabatic Molecular Dynamics (NAMD)**.[228, 229]

CHAPTER 3

Quantum Mechanics/Molecular Mechanics (QM/MM) Hybrid Method Implementation in deMonNano and Benchmark Study

The popular QM/MM hybrid schemes emerged in the 70s. The combination of Quantum Chemistry methods (QM) and Molecular Mechanics approaches (MM) (also called force fields type methods), based on Warshel and Karplus[230] and Warshel and Levitt[231] works. These approaches hunts an accurate balance between the reliability of calculations and the application to extended systems, with reasonable computational time consumption. The QM methods are employed to give a proper description of a chemical process taking place within a molecular system, the MM approaches integrates the possible environmental effects on the molecular system.

Within the QM approach, any type of electronic structure computational method can in principle be used, reminding that the more accurate the method is (wave function or DFT based methods) the more computational time will be. For this reason and due to the large systems' size studied in this thesis, a new QM/MM incode implementation has been performed in the deMonNano code and is reported in this chapter. This approach is based on the combination of density functional based tight binding method with class 1 force fields DFTB/MM (DFTB0-DFTB2-DFTB3/MM). In the following, when no specifications are done, DFTB can refer to either DFTB0, DFTB2 or DFTB3 indistinctly.

Firstly, a general introduction to QM/MM theory is presented, followed by a more technical section collecting the derivation of DFTB/MM equations and their implementations in deMonNano. Lastly, a complete benchmark study is detailed, based on the investigation of harmonic IR-spectra and binding energies of simple molecules in gas phase and in water clusters computed by the DFTB/MM new implementation, and compared to DFT/MM calculations.

3.1 QM/MM hybrid method general introduction

The hybrid QM/MM approach depicts **an explicit electron representation at the reactive center** performed by quantum methods (QM), embedded in **the electrostatic field of an environment** treated by force field type methods (MM). In the investigated system, two different regions can be distinguished. Firstly, **the QM region**, also referred as the *reaction center* where electrons and nuclei are treated explicitly. Secondly, **the MM region**, also known as the *spectator region*, where the atoms (treated as punctual particles and not as electrons and nuclei) do not participate directly to the reaction.[206,232] In the following, all $\mathbf{N^{QM}}$ atoms contained in the QM region will be referred as **QM atoms** and the $\mathbf{N^{MM}}$ atoms comprised in the MM region will be named as **MM atoms**. The total number of atoms in the system is given by $N=N^{QM}+N^{MM}$. Figure 3.1 depicts the two mentioned regions for a molecular system example.[206] In practice, in the context of this thesis, the QM and MM regions were defined by labeling the QM or MM atoms.

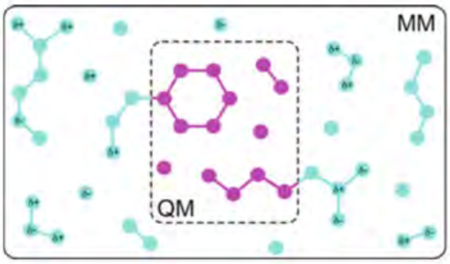


Figure 3.1: Illustration for the QM and MM regions in a general example of molecular system. Figure taken from G. Groenhof work.[206]

As an example, when investigating biological systems as enzymes, the catalytic process is known to be limited to an active site located in a precise area of a protein. The

QM region will be limited to the active site, whereas the protein is restricted to the MM area. Another example is the investigation of an interstellar ice embedding an amino acid molecule. In this case, the amino acid and the closest solvation water shells are involved in the main chemical processes, constituting the QM site. On the other hand, the bulk solvent is represented by the MM region, as only weak interactions will have an effect on the main chemical processes.

In the following paragraphs, references to *"treated by QM methods"* expression, will concern to the use of **DFT** or **DFTB** methods, exclusively for QM atoms (N^{QM}). On the other hand, expressions such as "*treated by MM methods*", will relate to force field class 1 method (Eq. 2.66 in previous chapter).

A brief illustration on the energy nomenclature used in the following is done using a general expression for the energy term $\mathbf{E^{Meth}}(\mathbf{N^{Inter}})$. In this energy expression, the upper-index **Meth** will denote the level of theory employed in the energy calculations. *Meth* can be either, QM, MM or QM-MM index (more details will be given in following section). On the other hand, $\mathbf{N^{Inter}}$ refers to the atomic interactions. N^{Inter} writes either as:

- N^{QM} refers to interactions between QM atoms. Similarly, N^{MM} concerns interactions between MM atoms.
- $N^{QM} + N^{MM}$ accounts for all interactions between every atom.
- $N^{MM} \rightleftarrows N^{QM}$ refers to the interactions between QM and MM atoms.

3.1.1 QM/MM coupling schemes

In previous chapter (see Chapter 2), different computational methods were presented for the calculation of individual E^{QM} and E^{MM} energies. However, for the latter $E^{QM\text{-}MM}$ energy term, some difficulties are encounter. Several approaches exist in order to determined the $E^{QM\text{-}MM}$ energy agglomerated in two different families[206] : **subtracting** and **additive QM/MM coupling schemes**. In this section we explain the two different ways to determine the total energy of the system $E^{QM/MM}$, respectively for the both couplings schemes.

Substractive QM/MM coupling

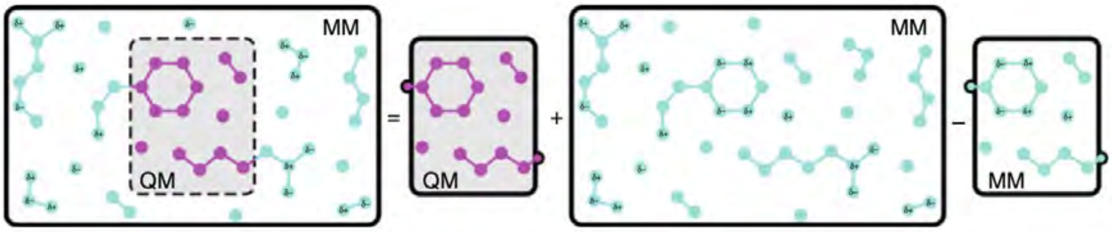
Three steps are followed to compute the total energy $\mathbf{E^{QM/MM}}$ when employing **the substractive QM/MM coupling scheme**[206, 233–235] :

1. Computation of the energy term $\mathbf{E^{MM}(N^{QM}+N^{MM})}$ for the total molecular system, comprising the QM and MM atoms in both regions. This energy term is evaluated by force fields type methods.
2. Calculation of the energy contribution for **the isolated QM subsystem** $\mathbf{E^{QM}}(\mathbf{N^{QM}})$ employing quantum methods and evaluated only on the QM atoms. This energy is then **added** to the former $\mathrm{E^{MM}(N^{QM}+N^{MM})}$ energy.
3. Lastly, the energy for the same QM region is evaluated **by a molecular mechanics method** $\mathbf{E^{MM}}(\mathbf{N^{QM}})$ and then **subtracted**. This last operation rectifies the double counting on the interactions within the QM subregion.

The **total energy** $\mathbf{E^{QM/MM}}$ is given by the expression:

$$\mathbf{E^{QM/MM}} = \mathrm{E^{MM}(N^{QM} + N^{MM}) + E^{QM}(N^{QM}) - E^{MM}(N^{QM})} \tag{3.1}$$

Figure 3.2 depicts the subtractive coupling within QM and MM subspaces, the level of theories employed and their evaluation on $\mathrm{N^{QM}}$ or $\mathrm{N^{MM}}$ atoms.



$$\mathbf{E^{QM/MM}} = \mathbf{E^{QM}}(\mathbf{N^{QM}}) + \mathbf{E^{MM}}(\mathbf{N^{QM}} + \mathbf{N^{MM}}) - \mathbf{E^{MM}}(\mathbf{N^{QM}})$$

Figure 3.2: Schematic QM/MM subtractive coupling representation. The total energy $\mathrm{E^{QM/MM}}$ (left hand side of the equation) is represented as the sum of $\mathrm{E^{MM}(N^{QM}+N^{MM})}$ plus $\mathrm{E^{QM}(N^{QM})}$ and then subtracted $\mathrm{E^{MM}(N^{QM})}$. Figure from G. Groenhof work.[206]

The most popular subtractive coupling approach is the ONIOM method,[233,235] although some other approaches exist, *e.g* ComQum.[234]

When translating this coupling scheme in terms of computational implementation, a great advantage is **the autonomy of QM and MM procedures**. In other words, no communication between the quantum and the molecular mechanics routines is needed. On the other hand, the lack of interchange between routines, turns into a poor description of certain interactions that play a major role in the chemistry of the system. This is the example of the **polarization effect** on the QM atoms by the MM region. Furthermore, due to the energy evaluation in the QM region by terms of MM methods (third energy term in Eq.3.1), very flexible and sophisticated force fields parametrizations are needed, in order to well describe the reaction center in the QM subsystem.

For this reason, a more realistic QM/MM representation is presented, where the interaction between MM and QM regions are more faithfully represented.

Additive QM/MM coupling

In **the additive QM/MM coupling** scheme, the total energy $\mathrm{E}^{\mathrm{QM/MM}}$ of a molecular system is built as follows:[206,236,237]

1. Firstly, the energy for **the QM subsystem is evaluated by quantum methods**. This is **the quantum energy** $\mathbf{E}^{\mathbf{QM}}(\mathbf{N}^{\mathbf{QM}})$.
2. Secondly, the energy **exclusively for the MM subsystem** is computed by a force field approach. This is the classical energy $\mathbf{E}^{\mathbf{MM}}(\mathbf{N}^{\mathbf{MM}})$ **added** to the previous energy value.
3. The last term, computes the interactions between QM and MM atoms. These interactions notably consist of the **electrostatic Coulomb type and long-range Van der Waal type interactions**, although some other covalent bonding interactions can be added to the energy representation (see Section 3.1.3). This energy $\mathbf{E}^{\mathbf{QM-MM}}(\mathbf{N}^{\mathbf{QM}} \rightleftarrows \mathbf{N}^{\mathbf{MM}})$ is evaluated by force fields methods and **added** to the previous two energies.

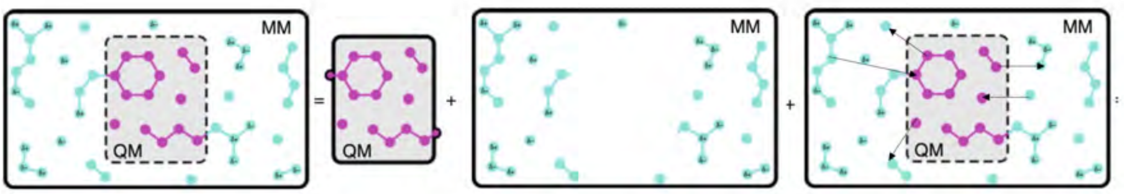
The reader must pay attention to the particular nomenclature in the energy term $\mathrm{E}^{\text{QM-MM}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows N^{\mathrm{MM}})$. The *upper-index* QM - MM highlights that this energy term evaluates the atomic interactions using both, quantum and molecular mechanics approaches.

The total energy expression for a QM/MM hybrid system is now given by the following expression[206] :

$$\mathbf{E}^{\mathbf{QM/MM}} = \mathrm{E}^{\mathrm{QM}}(\mathrm{N}^{\mathrm{QM}}) + \mathrm{E}^{\mathrm{MM}}(\mathrm{N}^{\mathrm{MM}}) + \mathrm{E}^{\text{QM-MM}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows \mathrm{N}^{\mathrm{MM}}) \tag{3.2}$$

In Figure 3.3, a graphical representation for the additive QM/MM coupling scheme is shown:

The additive QM/MM coupling is the scheme structure used in the new DFTB/MM implementations for this thesis. More details, regarding routines and encoding are given in Section 3.2.



$$\mathbf{E^{QM/MM}} = \mathbf{E^{QM}(N^{QM})} + \mathbf{E^{MM}(N^{MM})} + \mathbf{E^{QM-MM}(N^{QM} \rightleftarrows N^{MM})}$$

Figure 3.3: Schematic QM/MM additive coupling representation. The total energy $\mathrm{E^{QM/MM}}$ (left hand side of the equation) is given by the sum of $\mathrm{E^{QM}(N^{QM})}$ plus $\mathrm{E^{MM}(N^{NM})}$ plus the interaction energy $\mathrm{E^{QM\text{-}MM}(N^{QM}} \rightleftarrows N^{\mathrm{MM}})$. Arrows represent the corresponding electrostatic and long-range interactions between QM and MM atoms. Figure from G. Groenhof work with modifications.[206]

Different classes of embedding schemes within the additive QM/MM coupling frame exists, depending on the sophistication of $\mathrm{E^{QM\text{-}MM}(N^{QM}} \rightleftarrows N^{\mathrm{MM}})$ energy computation. In the following subsection, the two main **mechanical and electrostatic embedding** approaches are presented.

3.1.2 QM/MM embedding schemes

Based on the QM/MM region division, two possible scenarios can be depicted: **(i)** each molecule is seen as a *block* molecule, belonging to QM or MM region independently. Any division between atoms in a same molecule is not allowed. **(ii)** A molecule can be splitted, where few atoms are placed in the QM region, and the rest in the MM one. This introduces the possibility of sharing bonds, angles and dihedral angles between QM and MM atoms (see Figure 3.4 a, b and c), and therefore, incorporating covalent bonding energy terms in $\mathrm{E^{QM\text{-}MM}(N^{QM}} \rightleftarrows N^{\mathrm{MM}})$ expression (recall Eq.2.66) .

In the framework of this thesis, only the first scenario is considered. **Molecules are only seen as *blocks***. The covalent bonding energy terms will be neglected and $\mathrm{E^{QM\text{-}MM}(N^{QM}} \rightleftarrows N^{\mathrm{MM}})$ **will contain only non-covalent interactions**. Figure 3.4 d and e, illustrates the Van der Waals and Coulomb interactions (see Eq.2.66). In addition, to define the "optimal" QM region a convergence criteria during calculations is chosen. The balance between a proper description of the system's chemistry and the computational time is targeted.

Mechanical Embedding

The main characteristic of this type of embedding is that **all interactions between QM and MM atoms are treated by force field methods**. Fur-

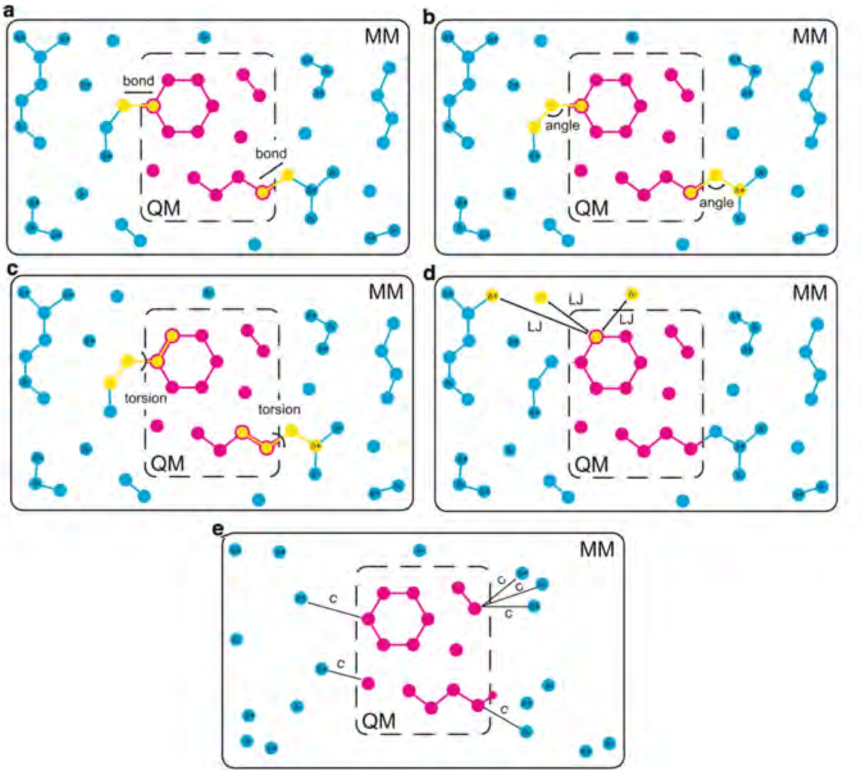


Figure 3.4: Schematic representation for chemical bond division. Images a, b and c represents the bond, angle and torsion elements for QM and MM atoms interaction. Panel d, depicts the Van der Waals interactions modelled by a Lennard-Jones potential. Panel e represents the Coulomb interaction for MM atoms charges represented as point charges. Figure from G. Groenhof work with modifications.[206]

thermore, in the case of the simplest mechanical embedding, only the short-range Van der Waals interactions are included in the $\mathrm{E}^{\text{QM-MM}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows N^{\mathrm{MM}})$ (Figure 3.4 d). The electronic wave function is evaluated for an isolated QM subsystem and consequently, **the electron density in QM region is not polarized by the MM environment**[206] . Van der Waals interactions are modeled by a Lennard-Jones potential (see Eq.2.70 in Chapter 2) and the energy term will be refered as $\mathbf{E}^{\textbf{QM-MM-Mech}}(\mathbf{N}^{\mathbf{QM}} \rightleftarrows \mathbf{N}^{\mathbf{MM}})$.

Due to the problem of capturing polarization in this kind of embedding, an upgraded on the embedding approach must be done, where the electrostatic effects are also taken in to consideration.

Electrostatic Embedding

This type of embedding is characterized by the computation of both non-covalent, Van der Waals and Coulomb interactions. Thus, the already included Van der Waals long-range interactions (mechanical embedding) are completed by the addition of the electrostatic Coulomb interactions between QM and MM regions. Now two energy terms will constitute the computation of $\mathrm{E}^{\text{QM-MM}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows N^{\mathrm{MM}})$ energy term:

$$\mathrm{E}^{\text{QM-MM}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows \mathrm{N}^{\mathrm{MM}}) = \mathrm{E}^{\text{QM-MM-Mech}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows \mathrm{N}^{\mathrm{MM}}) + \mathrm{E}^{\text{QM-MM-Elec}}(\mathrm{N}^{\mathrm{QM}} \rightleftarrows \mathrm{N}^{\mathrm{MM}}) \tag{3.3}$$

where $\mathbf{E}^{\textbf{QM-MM-Elec}}(\mathbf{N}^{\mathbf{QM}} \rightleftarrows \mathbf{N}^{\mathbf{MM}})$ refers to the Coulomb interactions between the subsystems QM and MM. The computation of this latter energy term is handle during the computation of the electronic wave function/electron density. Therefore, the MM atoms' charges enter the QM Hamiltonian as one-electron operators, and the total QM/MM operator is given by[206] :

$$\mathrm{h}_i^{\mathrm{QM/MM}}(\mathrm{N}^{\mathrm{QM}} + \mathrm{N}^{\mathrm{MM}}) = \mathrm{h}_i^{\mathrm{QM}}(\mathrm{N}^{\mathrm{QM}}) - \sum_{a}^{N^{MM}} \frac{e^2 Q_a}{4\pi\epsilon_0 \left|\mathbf{r_i} - \mathbf{R_a}\right|} + \mathrm{h}_i^{\mathrm{MM}}(\mathrm{N}^{\mathrm{MM}}) \tag{3.4}$$

where $\mathbf{r_i}$ and $\mathbf{R_a}$ correspond to the position of the QM i electrons and MM a atoms respectively. $\mathrm{h}_i^{\mathrm{QM}}(\mathrm{N}^{\mathrm{QM}})$ corresponds to the DFT or DFTB Hamiltonians followed by the extra term computing the Coulomb interaction between QM electrons and MM partial charges Q_a (Figure 3.4 e). Lastly, $\mathrm{h}_i^{\mathrm{MM}}(\mathrm{N}^{\mathrm{MM}})$ accounts for the classical molecular mechanics operator, being a constant during the SCF cycles.

Due to the second term in Eq.3.4, the code implementation of the electrostatic embedding model requires modifications not only in the MM routines but also in the quantum chemistry part of the code. More details will be given in Section 3.2.

Furthermore, from the latter inclusion of partial Q_a MM charges in the QM Hamiltonian, the effect of MM charges on the QM region is taken in consideration, but not vice versa. The electrostatic effect from QM atoms' charges on MM atoms is not computed in this embedding model. Therefore, only one-way of polarization effect is considered ($N^{\mathrm{MM}} \rightarrow N^{\mathrm{QM}}$).

In this thesis, both **mechanical and electrostatic embedding schemes have been implemented in deMonNano software**. More details are given the in upcoming sections.

Coupling and embedding model extensions

Although they are not used in this thesis, several coupling and embedding extensions are now reported. From the principal subtractive/additive couplings and mechanical/electrostatic embedding, new extensions in the QM/MM hybrid method approach exist:

- A way to enhance the polarization representation is given by the new **polarization embedding schemes**, where both QM and MM regions can mutually polarize each other.[206] Different approaches have been developed to model this mutual polarization as in: charge-on-a-spring model,[238] induced dipole model[239] and the fluctuating charge model.[240]

- In the cases where chemical bonds are shared by QM and MM atoms (recall the second scenario presented in Section 3.1.3), special care has to be taken when evaluating the QM wave function/electron density. Two main approaches exist, trying to well describe the open valence electron-pairs (appearing from the shared QM and MM bond representation).

  In one hand, the **link atoms approach** introduces monovalent link atoms at precise positions along the bond shared between QM and MM atoms. These link atoms are included during the QM calculations but are neglected during the MM ones.

  On the other hand, an alternative to the link atom representation, are the **localized orbitals**. The chemical shared bonds between QM and MM atoms are substituted by a doubly-occupied molecular orbital. Two main schemes can be found in literature, such as the **localised hybrid orbital method** introducing new orbitals in the QM atoms,[241] and the **generalized hybrid orbital approach** which places additional orbitals in the MM atoms[242] .

Several packages making use of DFT/MM schemes can be found in literature, such as, NWCHEM,[243] QCHEM,[244] Gaussian[245] and the one employed in this thesis *i.e* deMon2k.[148] Likewise, several softwares including DFTB/MM scheme approaches are accessible, such as CHARMM,[246] and the one used in this work, deMonNano.[149] At the present, other strategies can be also employed, considering the perturbation of a system described at the DFT or DFTB level by its environment, such as implicit models based on a continuous description of the medium[247,248] or models based on a coarse-graining strategy.[249] In addition, DFTB has also been included in QM/QM' schemes, acting as the lower level QM' method when combined with DFT.[250]

A state of art is now reported for the already implemented QM/MM hybrid methods in both deMon2k (DFT/MM) and deMonNano (DFTB/MM) codes.

3.1.3 QM/MM method in deMon2k and deMonNano. State of art

deMon2k QM/MM scheme implementations

In the description of QM/MM methodology implemented in deMon2k, two main developpements can be found based on an additive QM/MM coupling with an electrostatic embedding scheme.[251] Firstly, **the classical non-polarizable force fields** were implemented, including the electrostatic interaction between MM and QM atoms, but neglecting the polarization of MM atoms by QM ones. An upgraded level consisted in the inclusion of **the polarizable force fields (QM/MMpol)** in the previous mentioned implementations, where mutual polarization between QM and MM atoms are taken in consideration.[252,253]

Although in deMon2k, hybrid DFT/MM are not yet implemented for periodic boundary conditions, a good approximation is given by the modelling of long-range interactions **by the Onsager Self-Consistent Reaction Field (SCRF) method**[254] , to account for a polarizable continuum solvent medium. In addition, for cases were a covalent bond is found in the frontiers of QM and MM regions, **the link atom approach**[255,256] (see previous section) is implemented in this code.

A great interest of the QM/MM hybrid developments in deMon2k, is the possibility to include the Auxiliary-DFT (ADFT) method (see Section 2.3.2) also in the hybrid schemes. A reduction of the computational time from the use of both approaches is noted in simulations. Furthermore, all implementations can be employed in single point, optimizations and also molecular dynamics simulations .

Let us finally mention that, an important advantage of all QM/MM hybrid implementations in deMon2k comes from the fact that force field parameters are collected in an individual **dataset named *FFDS*** (based on Tinker[257] file program). This facilitates the use of any kind of force field potentials, as long as the user provides the corresponding force fields parameters. To this date, AMBERf99,[258] Amberff02,[259] OPLS[260] and CHARMM22[261] FFDS files are available in deMon2k.

deMonNano QM/MM scheme implementations

The previous DFTB/MM implementation in deMonNano package consisted of a **DFTB method coupled with the Universal Force Field (UFF)**. DFTB/UFF approach was based on a **substractive QM/MM coupling** scheme.[262] UFF is characterized by its unique definition of molecular mechanics parameters and it does not allow the change to other force field parameters. For instance, specification on the UFF water definitions were given only applying the TIP3P charges.[262] Furthermore, the implementation of electrostatic interaction in the latter development was not achieved, due to the use of the subtractive approach and therefore, restraining the performance of local optimization or MD simulations within the electrostatic scheme.

In addition, another **coupling of DFTB with a polarisable force field to account for argon matrix environment** was performed by Iftner *et al.*[263] The polarization of MM atoms by QM ones and dispersion effects are included in the coupling. However, the electrostatic effect from the individual MM charged atoms on the QM system are not taken in to consideration in this implementation, limiting its use to a rare gas inert environment.

3.2 Incode additive DFTB/MM Implementation

We present the first implementation work performed in this thesis which consists of a new coupling of DFTB quantum method with the class 1 force field potentials employing an additive QM/MM scheme, in both, mechanical and electrostatic embedding schemes developed in deMonNano.

It is important to mention that, in deMonNano software packages, DFTB/MM hybrid approach is implemented in a single code, as previously done also in some other softwares.[169, 246, 264, 265] This reduces the loss of computational efficiency compared to certain codes where an interface between QM and MM packages is needed, such as PUPIL[266] , Chemshell[267] , QMMM[268] or LICHEM[269] in the case of DFT/MM, and the combination of GROMACS+PLUMED[270] or DFTB+[201] for the case of DFTB/MM approach.

The improvements reached by this new DFTB/MM implementations, such as the use of FFDS dataset or the inclusion of the electrostatic scheme, are completed by their combination with the already existing deMonNano tools (CM3 charges, dispersion effects, see Sec.2.3.3). Furthermore, optimizations and MD simulations are possible due to the implementation of forces. In the following, DFTB/MM will only refer to the new implementation.

From the $E^{QM/MM}$ total energy term for an additive QM/MM coupling reported in Section 3.1.1, the energy expression can be as:

$$\mathbf{E}^{\mathbf{DFTB/MM}} = \mathrm{E}^{\mathrm{DFTB}}(\mathrm{N}^{\mathrm{DFTB}}) + \mathrm{E}^{\mathrm{MM}}(\mathrm{N}^{\mathrm{MM}}) + \mathrm{E}^{\mathrm{DFTB\text{-}MM}}(\mathrm{N}^{\mathrm{DFTB}} \rightleftarrows \mathrm{N}^{\mathrm{MM}}) \quad (3.5)$$

Recalling the energy expression for DFTB2 and the classical molecular mechanics equation (Sec. 2.3.3 and 2.3.4):

$$\mathbf{E}^{\mathbf{QM}}(\mathbf{N}^{\mathbf{QM}}) = \mathbf{E}^{\mathbf{DFTB2}} = \sum_{a,b} E_{rep}(R_{ab}) + \sum_{i}^{N} n_i \sum_{\mu,\nu}^{M} c_{i\mu} c_{i\nu} H^0_{\mu\nu} + \frac{1}{2} \sum_{a,b}^{N^{QM}} \gamma_{ab} q_a q_b \quad (3.6)$$

$$\begin{aligned} \mathbf{E}^{\mathbf{MM}}(\mathbf{N}^{\mathbf{MM}}) &= \sum_{l}^{N_{bond}} V_l^{bond} + \sum_{p}^{N_{angles}} V_p^{angles} + \sum_{q}^{N_{tors}} V_q^{tors} + \sum_{q}^{N_{imp}} V_q^{Imp} + \sum_{p}^{N_{angles}} V^{UB} \\ &+ \sum_{a}^{N^{MM}} \sum_{b>a}^{N^{MM}} V_{ab}^{Coul} + \sum_{a}^{N^{MM}} \sum_{b>a}^{N^{MM}} V_{ab}^{LJ} \end{aligned} \quad (3.7)$$

In Eq.3.7, covalent bonding and non-bonding terms are computed.

The the $\mathbf{E}^{\mathbf{MM}}(\mathbf{N}^{\mathbf{MM}})$ energy terms in this new implementation are computed from the parameter dataset file (FFDS). Making use of the deMon2k FFDS dataset representation, new force fields are available now in deMonNano such as: AMBERf99[258] , Amberff02[259] , OPLS[260] and CHARMM22.[261]

As a final appreciation, in MM approaches, the electronic degrees of freedom are not considered explicitly. As a consequence, the impact of the $\mathrm{E}^{\mathrm{MM}}(\mathrm{N}^{\mathrm{MM}})$ energy contribution on the secular equation can be, thus, ignored and its contribution to the total energy can be derived straightforwardly[232] (see Figure 3.5 and 3.6 right pannels).

The last step for the computation of the total energy (Eq.3.5) of a molecular system $\mathrm{E}^{\mathrm{DFTB/MM}}$ is to calculate the last energy term $\mathbf{E}^{\mathbf{DFTB-MM}}(\mathbf{N}^{\mathbf{QM}} \rightleftarrows \mathbf{N}^{\mathbf{MM}})$. In the following section, two different ways to calculate this contribution in the new DFTB/MM implementation are introduced, related to the embedding scheme type.

3.2.1 DFTB/MM Embedding Schemes Implementation

In this section the implementations of DFTB/MM mechanical and electrostatic embedding schemes in deMonNano are presented.

DFTB/MM Mechanical embedding implementation

As accounted in Section 3.1.3, the simplest mechanical embedding scheme comprises only the non-bonding interactions between QM and MM regions, the QM subsystem is not polarized by the MM subsystems and vice versa.[271] The $E^{\text{QM-MM}}$ energy interaction term between QM and MM atoms, referred throughout this work as $\mathbf{E^{QM-MM-Mech}}$, comprises only **the Van der Waals interactions**:

$$\mathbf{E^{QM-MM-Mech}} = \sum_{a \in QM}^{N^{QM}} \sum_{c \in MM}^{N^{MM}} \left(\frac{C_{12}^{ac}}{R_{ac}}\right)^{12} - \left(\frac{C_{6}^{ac}}{R_{ac}}\right)^{6} \tag{3.8}$$

Figure 3.5, illustrated the algorithm for the **the total energy $\mathbf{E^{QM/MM}}$ computation** for a molecular system. In this picture, it is shown how the DFTB ($E^{QM}(N^{QM})$) energy term is **computed independently** from the other two terms, $E^{MM}(N^{MM})$ and $E^{\text{QM-MM}}(N^{QM} \rightleftarrows N^{MM})$. The calculation of DFTB energy is done by the standard DFTB0/DFTB2/DFTB3 quantum methods. On the other hand, $E^{MM}(N^{MM})$ and $E^{\text{QM-MM}}(N^{QM} \rightleftarrows N^{MM})$ energies are solved respectively by the MM Eq.3.7 and Eq.3.8.

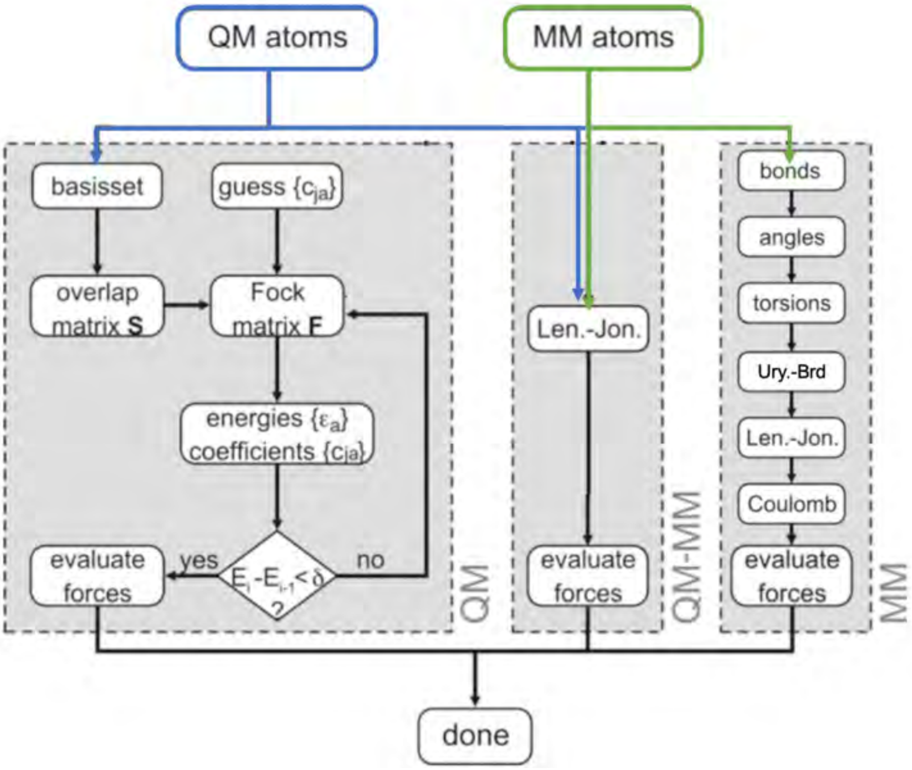


Figure 3.5: Total energy $E^{QM/MM}$ computation algorithm for new DFTB/MM class 1 implementations within a mechanical embedding scheme. QM panel refers to the computation of $E^{QM}(N^{QM})$ energy term. Similarly, MM and QM-MM panels, refers to $E^{MM}(N^{MM})$ and $E^{\text{QM-MM}}(N^{QM} \rightleftarrows N^{MM})$ respectively energy computations . Figure from G. Groenhof work with modifications.[206]

DFTB/MM Electrostatic embedding implementation

The Coulomb interaction between the MM point charges and the DFTB density should in principle involve Coulomb integrals over the DFTB density described in the atomic basis. However, following the DFTB approximation used to compute the Coulomb term in the second order energy correction (see Eq.3.6), the DFTB density is approximated for this term by atomic charges using the Mulliken scheme formalism to replace the electronic density by atomic monopoles. It leads to the following expression for the **QM-MM Coulomb potential energy** ($\mathbf{E^{QM-MM-Elec}}$):

$$\mathbf{E^{QM-MM-Elec}} = \sum_{a\in QM}^{N^{QM}} \sum_{c\in MM}^{N^{MM}} \frac{q_a q_c}{R_{ac}} \tag{3.9}$$

arising from the interaction between MM atomic charges (q_c) and DFTB atomic charges (q_a). The latter, derived from the electronic density in the DFTB region. The $\mathrm{E^{QM\text{-}MM}}$ interaction energy within an additive QM/MM coupling and an electrostatic embedding scheme (recall Eq. 3.2) is given by the following energy contri-

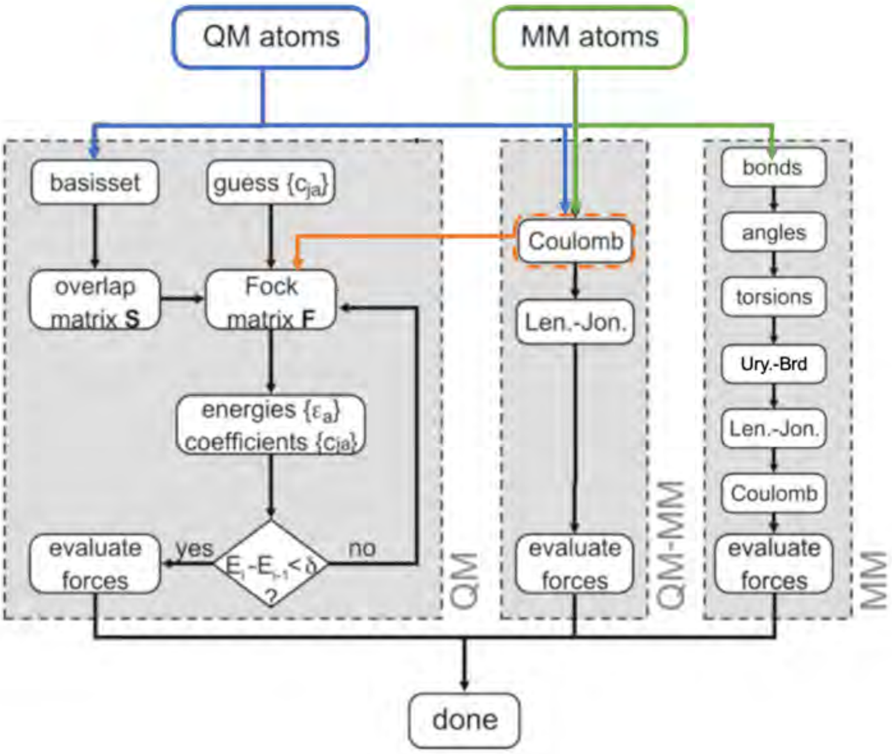


Figure 3.6: Total energy $\mathrm{E^{QM/MM}}$ computation algorithm for new DFTB/MM class 1 implementations within a electrostatic embedding scheme. QM panel refers to the computation of $\mathrm{E^{QM}(N^{QM})}$ energy term. Similarly, MM and QM-MM panels, refers to $\mathrm{E^{MM}(N^{MM})}$ and $\mathrm{E^{QM\text{-}MM}(N^{QM}+N^{MM})}$ respectively energy computations. Orange signal depict the Coulomb potentail for QM and MM atom interactions also involved in the $\mathrm{E^{QM}}$ energy computation. Figure from G. Groenhof work with modifications.[206]

butions:

$$\mathbf{E}^{\mathbf{QM-MM}}(\mathbf{N}^{\mathbf{QM}}+\mathbf{N}^{\mathbf{MM}}) = \sum_{\substack{a\in QM\\ c\in MM}}^{K}\left(\frac{C_{12}^{ac}}{R_{ac}}\right)^{12}-\left(\frac{C_{6}^{ac}}{R_{ac}}\right)^{6}+\sum_{\substack{a\in QM\\ c\in MM}}^{\mathbf{K}}\frac{\mathbf{q_a q_c}}{\mathbf{R_{ac}}} \tag{3.10}$$

Figure 3.6 captures **the total energy $\mathbf{E}^{\mathbf{QM/MM}}$** computation algorithm, where indeed the classical MM energy $\mathrm{E}^{\mathrm{MM}}(\mathrm{N}^{\mathrm{MM}})$ calculation is not modified. As the atomic charges from the QM region are optimised self-consistently, the DFTB operator used to solve the secular equation (recall Sec.2.3.3) now contains an additional term (in orange in Figure 3.6) leading to the following expression for the matrix elements in the atomic basis:

$$H_{\mu\nu} = H_{\mu\nu}^{DFTB} + H_{\mu\nu}^{\text{QM-MM-Elec}} + H_{\mu\nu}^{MM}$$

with $H_{\mu\nu}^{MM}$ being constant and

$$H_{\mu\in a;\nu\in b}^{\text{QM-MM-Elec}} = -\frac{1}{2}S_{\mu\nu}\sum_{\substack{a,b\in QM\\ c\in MM}} q_c\left(\frac{1}{R_{ac}}+\frac{1}{R_{bc}}\right) \tag{3.11}$$

The analytical gradients have been derived and implemented for both mechanical and electrostatic embedding schemes. A new term appears in the DFTB2 energy gradient expression within the electrostatic embedding scheme:

$$\begin{aligned}\mathbf{F_a} = -\frac{\partial E^{SCC-DFTB}}{\partial \mathbf{R_a}} &= -\sum_{a\neq b}^{K}\frac{\partial V_{rep}^{ab}(\mathbf{R}_a-\mathbf{R}_b)}{\partial \mathbf{R_a}} - q_a\sum_{\xi=1}^{K}\frac{\partial \gamma_{a\xi}}{\partial \mathbf{R_a}}q_\xi \\ &+\sum_{i=1}^{N} n_i\sum_{\mu\nu}^{M} c_{\mu i}c_{\nu i}\left[\frac{\partial H_{\mu\nu}^{0}}{\partial \mathbf{R_a}}-\left(\epsilon_i-\frac{H_{\mu\nu}^{1}}{S_{\mu\nu}}\right)\frac{\partial S_{\mu\nu}}{\partial \mathbf{R_a}}\right] \\ &+\sum_{\substack{a\in QM\\ c\in MM}}^{K} q_a q_c\frac{\mathbf{R_{ac}}}{|\mathbf{R_{ac}}|^3}\end{aligned} \tag{3.12}$$

The actual version of DFTB/MM hybrid method implementation is already available and can be downloaded in the deMonNano website.

3.3 Implementation Benchmark and Test Case Applications

In order to validate the implementation of the DFTB/MM hybrid method, a benchmark study was performed for the energy gradients implementation. Furthermore, a benchmark study was carried out comparing harmonic vibrational spectra and binding energy values obtained at the DFT/MM and DFTB/MM (class 1) levels for simple molecules, isolated or trapped in water clusters. .

All results are collected in the up coming sections. Firstly, the energy gradient benchmark is presented, followed by the initial test cases geometries and the computational strategy for both, DFT/MM and DFTB/MM simulations. Lastly, results and discussion are reported.

3.3.1 Energy Gradient Benchmark Analysis

This benchmark is done on an arbitrary system consisting of an ammonia molecule (QM level) and a single water molecule (MM level). The goal is to verify the well implementation of energy and gradients in the new DFTB/MM(class 1) approach. The energy gradient values with respect to atomic coordinates, for both analytical DFTB/MM implementation and by finite differences, are reported in Table 3.1. It confirms that the two approaches give almost the same values, as the remaining differences are smaller than 1.3×10^{-7} Hartree/Bohr. Furthermore, an additional benchmark is shown in Figure 3.7. It depicts the potential, kinetic and total energies during a molecular dynamics simulation for the ammonia molecule embedded in a 215 water cluster (test case system) in the micro-canonical ensemble. It can be observed that the total energy fluctuations (too small to be visible on the figure) remain much smaller than the kinetic and potential energies fluctuations. The latter result, validates the proper implementation of the energy gradient.

3.3.2 Test Case Molecular Systems and Computational Details

Four simple molecules in gas phase, ammonia (NH_3), *cis-cis*-glycine (NH_2CH_2COOH), pentane (C_5H_{12}) and a water molecule (H_2O), were chosen as ini-

Atom	**Coord.**	**DFTB/MM Deriv.**	**Numerical Deriv.**	**Diff.**
N	x_{QM}	-0.636E-01	-0.636E-01	-0.975E-07
	y_{QM}	0.875E-02	0.875E-02	-0.109E-07
	z_{QM}	0.131E-01	0.131E-01	-0.167E-07
H	x_{QM}	-0.368E-04	-0.369E-04	0.130E-06
	y_{QM}	0.143E-03	0.143E-03	0.113E-07
	z_{QM}	0.839E-04	0.839E-04	-0.109E-07
H	x_{QM}	0.210E-02	0.210E-02	0.177E-07
	y_{QM}	-0.467E-03	-0.467E-03	-0.487E-07
	z_{QM}	-0.120E-02	-0.120E-02	0.188E-07
H	x_{QM}	-0.107E-02	-0.107E-02	0.716E-08
	y_{QM}	0.908E-03	0.908E-03	0.520E-07
	z_{QM}	0.192E-02	0.192E-02	0.744E-08
O	x_{MM}	0.644E-01	0.644E-01	-0.825E-07
	y_{MM}	-0.116E-01	-0.116E-01	0.482E-07
	z_{MM}	-0.123E-01	-0.123E-01	-0.342E-08
H	x_{MM}	-0.479E-02	-0.479E-02	-0.317E-07
	y_{MM}	0.444E-02	0.444E-02	-0.411E-07
	z_{MM}	-0.106E-02	-0.106E-02	-0.152E-09
H	x_{MM}	0.297E-02	0.297E-02	0.326E-07
	y_{MM}	-0.214E-02	-0.214E-02	-0.414E-07
	z_{MM}	-0.453E-03	-0.453E-03	0.416E-10

Table 3.1: Comparison of the energy gradients (Hartree per Bohr) computed for an arbitrary geometry of NH_3/H_2O from the analytical expression and from finite differences calculations.

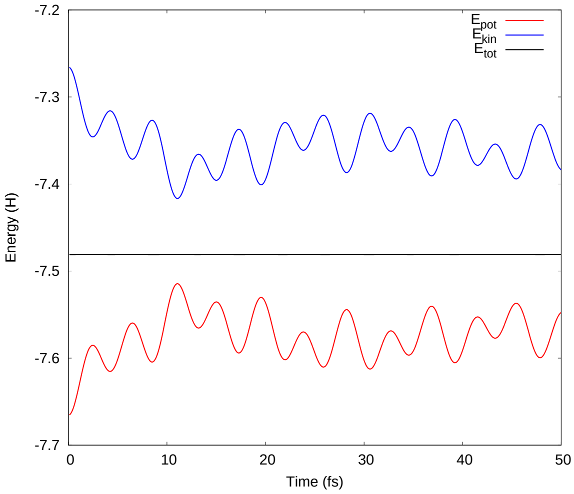


Figure 3.7: Kinetic, potential and total energies during a MD simulation of $NH_3(H_2O)_{215}$.

tial molecules. In addition to the isolated systems, latter molecules were trapped in water clusters of (~ 215 H_2O molecules) with the aim of **comparing the harmonic vibrational IR-spectra obtained by DFT/MM and DFTB/MM hybrid methods**. It has to be noted that no benchmark study was done between the DFTB/MM class 1 new implementation and the already implemented UFF QM/MM hybrid approach in deMonNano due to the lack of QM-MM embedding forces description of this latter.

Figure 3.8, depicts the four isolated and cluster molecular systems employed in this study. It must be noted that the geometrical configurations were arbitrarily generated, they do not exemplify any real system. A complete study of such systems would require global explorations of the complex potential energy surfaces, which is out of the scope of the present benchmark. Any analysis of the properties and characteristic of this geometries cannot be extracted.

Computational Strategy

The benchmark of the new DFTB/MM additive coupling with **electrostatic embedding scheme**, consisted of two steps. First, the selected small molecules (Figure 3.8) were optimized at both the DFT and DFTB3 levels in gas-phase. In the second step, each molecule was introduced in a water cluster. A short molecular dynamics (50 ps) was performed at the DFTB3/MM level in the canonical ensemble (temperature of 60 K maintained by a chain of 5 Nose-Hoover thermostats[272–274] with energy exchange frequency of 800 cm^{-1}). The final geometry was optimised at the DFTB3/MM level and the DFTB3/MM harmonic vibrational spectra was obtained. Finally, the latter geometries were further optimised at the DFT/MM level and similarly, DFT/MM harmonic IR-pectra was obtained. For these optimised geometries in gas-phase and in water clusters, the IR spectra were computed in the harmonic approximation, involving the diagonalization of the mass weighted Hessian matrix. In both, DFT/MM and DFTB3/MM implementations, only the dipole moments arising from the solute molecule, therefore only the QM dipole moments, were considered in the calculations of the IR intensities. The relevant information is therefore the solvation effects on the IR spectra of the solute molecules and not the spectrum of the full cluster.

All water cluster environments were treated at the MM level. OPLS-AA (TIP3P for water molecules) has been used for the MM potential in both, DFT/MM and DFTB3/MM cases. Solute molecules (NH_3, NH_2CH_2COOH, C_5H_{12} and H_2O) are hence treated by DFT or DFTB3 levels of theory. In the latter case, 3ob DFTB

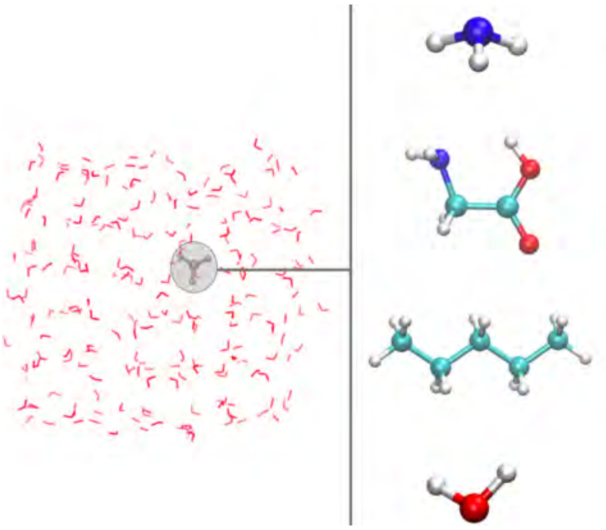


Figure 3.8: Test cases studies for ammonia (NH_3), *cis*-*cis*-glycine (NH_2CH_2COOH), pentane (C_5H_{12}) and a water molecule (H_2O). Each isolated molecules were trapped in the water cluster (left side of the image).

parameters were used.

As a last remark, DFT calculations were carried out using the PBE exchange correlation functional[275] in combination with the DZVP basis set.[276] The choice of this non-empirical exchange-correlation functional was driven due to reasonable accuracy over an extensive range of systems. In addition, the DFTB3 3OB is parametrized from PBE calculations. Automatically generated auxiliary basis sets were employed.[277] An adaptive grid for numerical quadrature was used. The SCF and gradient convergence tolerances were set to 10^{-7} and 10^{-5} a.u, respectively.

3.3.3 IR Spectra Analysis

IR spectra for each molecular system performed at DFT/MM and DFTB3/MM methods are reported in this section.

Ammonia (NH_3)

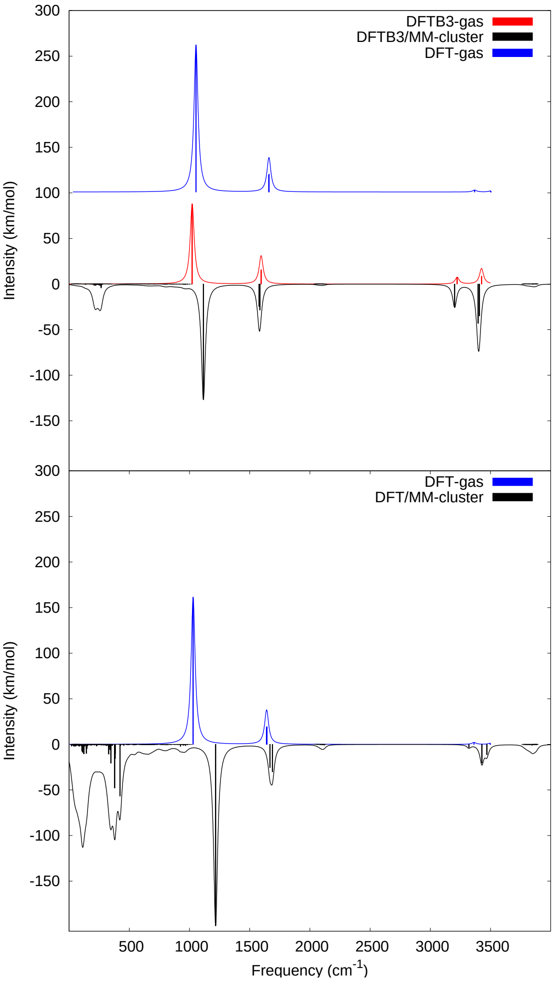


Figure 3.9: Harmonic IR spectra for ammonia. Top: gas phase molecule at the DFT and DFTB3 levels and solvated ammonia at DFTB3/MM levels; Bottom: gas phase and solvated ammonia at the DFT and DFT/MM levels. Harmonic stick spectra were convoluted with Lorentzian function (FWMH=20 cm^{-1}) to produce the line spectra.

Figure 3.9-top compares the ammonia IR spectra computed for the gas phase ammonia at the DFT and DFTB3 levels. At both the DFT and DFTB3 levels, the most intense band corresponds to N-H wagging vibrational mode and the frequencies are in good agreement (1022 cm^{-1} for DFTB3 vs 1030 cm^{-1} for DFT). A second band, by far less intense corresponds to H-N-H scissoring modes and the even less intense bands (almost no intensity at the DFT level) located around 3400 cm^{-1} result from N-H stretching modes.

When NH_3 is inside the cluster, the main band (wagging mode) computed at the DFT/MM level is shifted towards higher frequencies (Figure 3.9-bottom). This trend is also captured by the DFTB3/MM spectra (Figure 3.9-top), although the shift is smaller (93 cm^{-1} vs 187 cm^{-1} at the DFT/MM level). The intensity of this band is increased at both the DFT/MM and DFTB3/MM levels with respect to the isolated molecule spectrum. Very few changes are observed for the scissoring bands whatever the potential used. The intensities of the stretching modes are increased with either DFT/MM or DFTB3/MM. The last effects of the environment is the appearance of bands below 500 cm^{-1} visible in the

DFT/MM spectrum and present, although much less intense, in the DFTB3/MM spectrum. These frequencies correspond to soft intermolecular modes.

In some of them, the ammonia presents rotation and translation as a rigid body, whereas some others only correspond to intermolecular vibrations of water molecules without the involvement of displacements of ammonia atoms. In the latter modes, only the ammonia electronic density is perturbed.

A little contribution from water environment is also observed at both the DFTB3/MM and DFT/MM levels around 3855 cm^{-1} and corresponds to combinations of the water stretching modes, inducing ammonia dipole fluctuations. Similarly, at both levels of theory, the small feature around 2100 cm^{-1} is associated to combinations of bending modes of many water molecules.

Furthermore, an example of DFTB3/MM and DFT/MM harmonic spectra comparison for the soluted ammonia system (Figure 3.10). Analysis of both DFTB3/MM and DFT/MM hybrid methods, show the same trend of water environment effects' on ammonia molecule. Although the principal modes in DFTB3/MM spectra are red shifted compared to the DFT/MM modes, we observed in both cases the increase of intensities and the appearance of intermolecular water modes at low frequencies. Furthermore, the bending modes at 2100 cm^{-1} associated with many water molecules are found in both levels of theory.

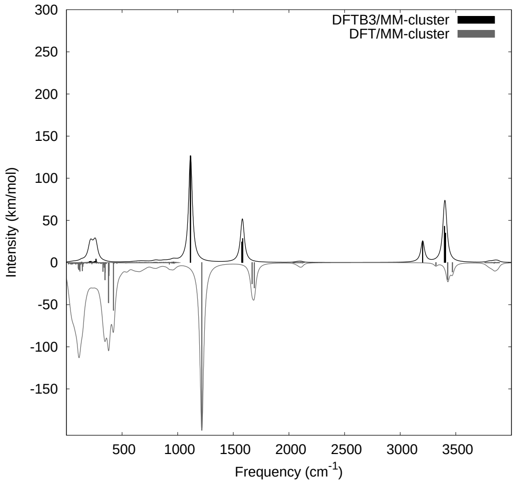


Figure 3.10: Harmonic IR Spectra solvated ammonia at the DFTB3/MM and DFT/MM level of theory.

Neutral (*cis-cis-cis*) glycine (NH_2CH_2COOH)

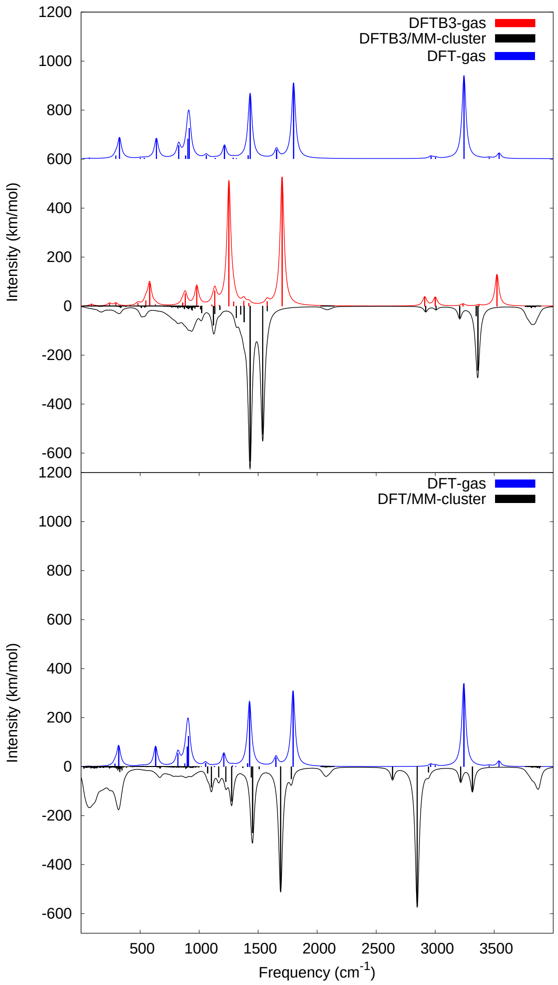


Figure 3.11: Harmonic IR spectra for neutral glycine. Top: gas phase molecule at the DFT and DFTB3 levels and solvated glycine at DFTB3/MM levels; Bottom: gas phase and solvated glycine at the DFT and DFT/MM levels. Harmonic stick spectra were convoluted with Lorentzian function (FWMH=20 cm^{-1}) to produce the line spectra.

The results obtained for the neutral glycine molecule are presented in Figure 3.11. Focusing first on the isolated molecule (top), it can be seen that the DFT spectrum presents an intense band at 3242.5 cm^{-1} assigned to the O-H stretching mode. This band is also among the most active ones at the DFTB3 level, although presenting a smaller intensity as well as a blueshift (3523.2 cm^{-1}) with respect to the DFT results. Apart from this band, the DFT neutral glycine spectrum is dominated by two modes corresponding mostly to the displacement of a carbon atom from the carboxyle group at 1795 cm^{-1} and to a C-O-H bending mode at 1427 cm^{-1}.

At the DFTB3 level, these two modes are the most active ones located at 1702 and 1252 cm^{-1}, respectively. Finally, several other modes are observed at lower frequencies in both spectra, the main difference being related to the modes around 900 cm^{-1} whose intensities are underestimated at the DFTB3 level in comparison to their DFT intensities.

When glycine is inside the water cluster, both methods lead to the same trends for the O-H streching mode, namely an increase of the intensity and a shift toward lower frequencies, the latter being stronger at the DFT/MM (395 cm^{-1}) level than at the DFTB3/MM level (162 cm^{-1}). The band at 1795 cm^{-1} (DFT value) is also red shifted in the cluster at both DFT/MM and DFTB3/MM levels, whereas a slight blue shift of the C-O-H bending mode is only observed at the DFTB3/MM level.

Finally, as reported in the case of ammonia, modes associated to either intermolecular vibrations, combinations of water bending or stretching modes are observed, in the low frequency domain and in the 2100 and 3850 cm^{-1} regions, respectively.

Pentane (C_5H_{12})

The harmonic IR spectra obtained for the pentane molecule are reported in Figure 3.12. For the isolated molecule, the two levels of theory agree on the main spectral features, namely the presence of two distinct active regions: (1) the [2900-3000] cm^{-1} region corresponding to various combinations of C-H stretching modes and (2) the [1300-1500] cm^{-1} region, corresponding to various combinations of C-H bending modes (essentially nodding and shearing modes). At lower frequencies, small contributions arise from modes associated to global deformations of the carbonaceous skeleton as well as some bending modes (torsion and swinging modes).

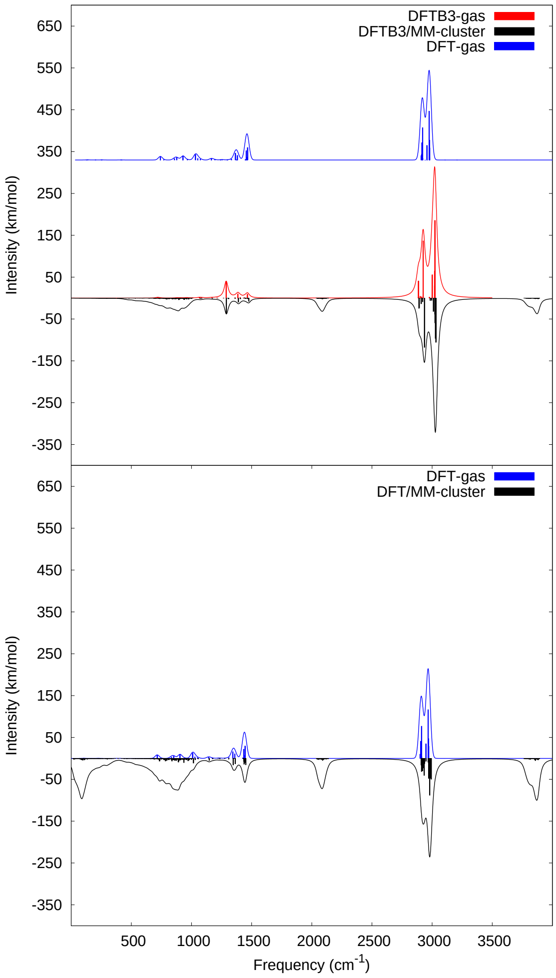


Figure 3.12: Harmonic IR spectra for pentane. Top: gas phase molecule at the DFT and DFTB3 levels and solvated pentane at DFTB3/MM levels; Bottom: gas phase and solvated pentane at the DFT and DFT/MM levels. Harmonic stick spectra were convoluted with Lorentzian function (FWMH=20 cm^{-1}) to produce the line spectra.

From Figure 3.12, in both top and bottom panels, it can be seen, in the opposite to what was observed in the case of ammonia and glycine, that the main features of the harmonic spectrum do not present significant changes due to the cluster environment, whatever the level of theory. This is well in line with the known hydrophobic nature of pentane,[278] which does not form strong H-bonds with the surrounding water molecules. The major observed changes are the appearance of frequencies in the 2100 and 3850 cm^{-1} regions, assigned to water H-O-H bending modes and O-H stretching vibrations, as well as in the low frequencies region, associated to intermolecular modes.

Water (H_20)

The last example, is the analysis of water spectra which are reported on Figure 3.13. Focusing first on the isolated molecule, the DFT spectrum indicates a bending mode at 1635 cm^{-1} and two stretching modes at 3721 and 3854 cm^{-1} for the symetric and antisymetric modes, respectively. These bands are also observed at the DFTB3 level at 1461, 3604 and 3877 cm^{-1}, respectively.

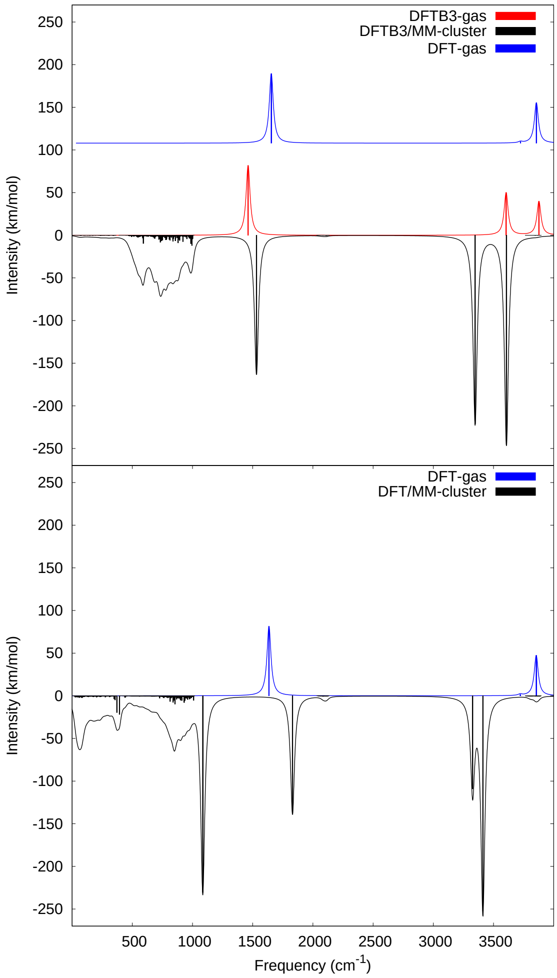


Figure 3.13: Harmonic IR spectra for water. Top: gas phase molecule at the DFT and DFTB3 levels and solvated water molecule at DFTB3/MM levels; Bottom: gas phase and solvated water molecule at the DFT and DFT/MM levels. Harmonic stick spectra were convoluted with Lorentzian function (FWMH=20 cm^{-1}) to produce the line spectra.

The intensities are in agreement with the exception of the symmetric stretching mode which is active in the DFTB3 spectrum but presents very low intensity in the DFT spectrum. Inside the water cluster, the investigated water molecule presents similar modifications when computed either at the DFT/MM or DFTB3/MM levels, namely an increase of the absorption intensities, in particular for the two stretching modes, a red-shift of the stretching modes and a blue-shift of the bending mode. Broad features below 1100 cm^{-1} originated from global intermolecular modes can be also observed in both spectra. The main difference is that, in the DFT/MM spectrum, one of these modes, located at 1085 cm^{-1} presents a strong intensity, whereas in the DFTB3/MM spectrum no intermolecular mode is observed with such a strong intensity.

In the following the influence of the size of the region treated at the QM level is discussed. To do so, an optimization and the computation of the harmonic spectrum was done for the water cluster, where all water molecules (*i.e.* those previously described either at the QM or at the MM level) were treated at the DFTB3 level. To be consistent with the previous calculations, only the dipole moment of the molecule previously described at the QM level is included in the intensity calculation. This spectrum is compared with the one of the isolated water molecule on Figure 3.14. As in the calculation performed at the DFTB3/MM level, the O-H stretching modes are red-shifted and their intensities are increased. It can be noted, that the intensity is now spread over many almost degenerated modes. The appearance of a broad pattern at low frequencies is also captured. The main difference relies on the bending modes, which were blue shifted in the DFTB3/MM spectrum. In the new simulation, they are almost not impacted by the environment, apart from the fact that its intensity is now spread over many modes with very similar frequencies. This can clearly be related to the fact that the frequency of the water bending mode is 1471 cm^{-1} when computed at the DFTB3 level and 2028 cm^{-1} when computed at the MM level. As a consequence, at the DFTB3/MM level the coupling of the bending mode of the QM molecule with the bending modes of the MM molecules at higher frequency are responsible for the observed blue shift.

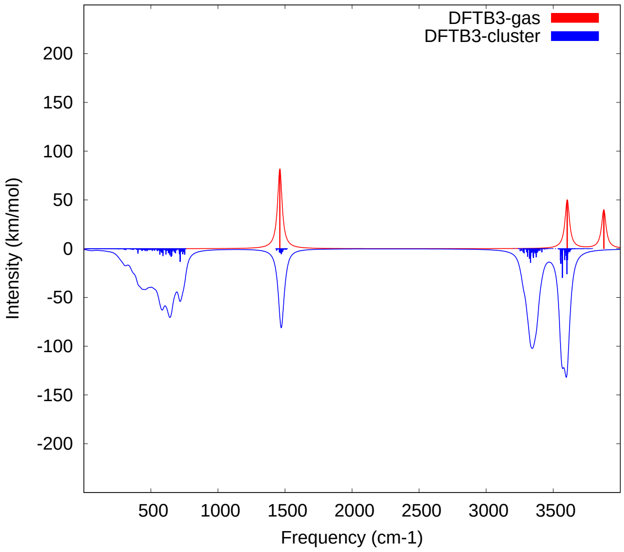


Figure 3.14: Harmonic spectra for water molecule in the gas phase and solvated in a water cluster computed at the DFTB3 level.

To conclude, it must be stressed that the size of the subsystem treated at the QM level should be large enough to limit such artifacts originated from the differences in the water molecules frequencies between the two levels of calculations. In addition, force fields of higher accuracy (polarisable force fields or TIP4P, or SPC, or TIP3F) could provide better descriptions of water frequencies.

3.3.4 Binding Energies Analysis

A comparison between binding energies values using both, DFT/MM and DFTB/MM methods was performed. The same test cases molecular systems from previous benchmark were used. Two subsystems, *i.e.* the studied molecules (NH_3, $C_2H_5NO_2$, C_5H_{12} or H_2O), treated at the QM level, and the surrounding water cluster, treated at the MM level were chosen. The binding energy is computed from the energies of the full system, minus the sum of the two isolated subsystems' energies:

$$\mathrm{E_{bind}} = \mathrm{E^{QM/MM}(N^{QM} + N^{MM}) - E^{QM}(N^{QM}) - E^{MM}(N^{MM})} \tag{3.13}$$

Before calculating the binding energies, geometries for the full (QM+MM) system were first optimized. No relaxation of geometries were performed for both isolated subsystems. Their coordinates were extracted from the full optimized system, firstly only for the QM subsystem, and then for the MM one. It must be precised that the optimizations of the full system were done using in both, DFT/MM and DFTB3/MM hybrid methods.

From the initial geometry structures optimized by DFT/MM hybrid techniques, only binding energies were calculated employing DFT/MM method. On the other hand, systems which were preliminarly locally optimized by DFTB3/MM approaches, binding energies were calculated simultaneusly employing DFT/MM, DFTB3/MM methods. Table 3.2 reports the different binding energiy values computed for the different systems.

Small differences (less than 0.13 eV) can be observed between the binding energies obtained at the DFT/MM level for the geometries resulting from local optimization at the DFTB3/MM (column 3) or DFT/MM (column2) levels. This indicates small structural relaxation when changing the computational scheme. The binding energies computed at the DFT/MM and DFTB3/MM levels for the same geometries (optimised at the DFTB3/MM level, columns 3 and 4) are in very good agreement for the glycine and pentane molecule, whereas for ammonia and water, the values obtained at the DFT/MM level (-0.649 and -1.380 eV, respectively) are larger than those obtained at the DFTB3/MM level (-0.394 and -1.001 eV, respectively).

$\mathbf{E_{Binding}}$ (eV)

Geometry Optimization	DFT/MM	DFTB3/MM		
Theory	DFT/MM	DFT/MM	DFTB3/MM	LJ contrib.
$\mathbf{NH_3}$	-0.778	-0.649	-0.394	0.065
$\mathbf{C_2H_5NO_2}$	-2.877	-2.759	-2.798	0.064
$\mathbf{C_5H_{12}}$	-0.639	-0.631	-0.640	-0.643
$\mathbf{H_2O}$	-1.498	-1.380	-1.001	0.317

Table 3.2: Binding energies computed at the DFT/MM or DFTB3/MM levels from Eq. 3.13 for the four investigated systems at the geometries optimized either at the DFT/MM or DFTB3/MM levels of theory. Lennard-Jones contributions to the binding energies are also reported.

In order to track the origin of this difference, we report the contribution of the LJ potential (column 5) to the coupling energy ($\mathrm{E}^{\text{QM-MM}}(\mathrm{N}^{\text{QM}} \rightleftarrows \mathrm{N}^{\text{MM}}) = V_{ab}^{LJ}$). Note that, this value is the same in both the DFT/MM and DFTB/MM schemes. It must be recalled, that even working within the electrostatic embedding (as it was the case in all simulations in this chapter), thanks to the incode implementations, individual values form Lennard-Jones and Coulomb interactions can be extracted from the calculations.

It appears that, apart from the pentane case where it almost equals the binding energy, the LJ contribution is repulsive. It means that most of the stabilising interaction arises from the interaction between the charges of the MM subsystem and electronic density of the QM subsystem. The difference could originate in the lack of polarization effect at the DFTB/MM level. In the DFTB and DFTB/MM operator expressions (Eq. 3.11), the atomic charges of the QM and MM regions induce couplings between the orbitals of different atoms but not between orbitals of the same atom, due to the fact that the overlap matrices elements $S_{\mu\nu}$ are zero if μ and ν correspond to two different orbitals of the same atom. As a consequence, an MM region would not induce polarization of an isolated QM atom. This neglect of atomic polarization energy could be, at least partially, at the origin of an underestimation of the binding energy at the DFTB3/MM level. Let us note in particular that the atomic polarisabilities of some elements,[279] like nitrogen (1.100 Å^3) or oxygen (0.802 Å^3), are significant with respect to molecular polarisabilities[280] of ammonia (2.103 Å^3) or water (1.501 Å^3) .

3.4 Conclusions and Perspectives

In this chapter, new implementations for a **hybrid DFTB(QM)/MM additive coupling approach, in both, mechanical and electrostatic embedding schemes** are reported. Different **DFTB** (DFTB0, DFTB2 and DFTB3) levels of theory are employed for the energy and forces computation in the QM part and various **class 1 force fields potentials** were included in the molecular mechanics description. At present, only the class 1 force fields AMBER-FF99SB[258] and OPLS-AA (OPLS)[260] have been used in the benchmark studies, but this new implementation allows to use any class 1 Force Field, as long as the user provides the corresponding parameters.

Within the additive approach in both, mechanical and electrostatic embedding schemes, the interactions between the MM and QM system subregions are computed from a Lenard-Jones potential. In the electrostatic coupling, an additional term describes the Coulomb interactions between the QM and MM atomic charges. The atomic charges from the MM domain remain constant but those of the QM region are optimised self-consistently within the DFTB variational scheme, including **the polarization of the QM region by the charges of the MM region**.

The **energy gradients**, mandatory to perform molecular dynamics or to compute the vibrational spectra, have also been implemented for this hybrid scheme. Benchmarks have been performed, comparing the energy gradients computed either from the analytical DFTB/MM or from finite differences. Furthermore, it was confirmed that the total energy was conserved during molecular dynamics simulations in a micro-canonical ensemble .

A more complete benchmark was also 0reported, for a set of selected small molecules (NH_3, NH_2CH_2COOH, C_5H_12 and H_2O) in water cluster environement ($\sim$ 215 H_2O molecules), comparing the harmonic IR spectra computed at the DFTB3/MM and DFT/MM levels. OPLS-AA (TIP3P model for water molecules) was employed for parametrizing the force field potential. From this benchmark, it is confirmed that **the main environment effects computed at the DFT/MM level are reproduced at the DFTB3/MM level**.

In the future, the combination of the present implementation with schemes similar to the one developed by Iftner *et al.*[263] to couple polarisable argon atoms description at the MM level with the DFTB scheme would allow us to mix DFTB with polarisable force fields similarily to what was done at the DFT level[169,252] .

CHAPTER 4

Temperature Driven Transformations of Neutral Glycine Molecule Embedded in Interstellar Ice

As already introduced in the first chapter, glycine amino acid and its formations on ice grains in space raises fundamental questions about glycine chemistry in interstellar media. In this work, we studied glycine conformational space and the related tautomerization mechanisms in water media in astrophysically relevant conditions by means of DFT/MM and DFTB/MM molecular dynamics simulations.

In this chapter, a brief state of art of glycine theoretical studies is reported, followed by a complete DFT/MM study of four glycine conformational isomers immersed in different water clusters and a low density amorphous (LDA) ice model. Lastly, a similar study performed with the newly implemented DFTB/MM class 1 FF method (see Chapter 3) is reported.

4.1 Introduction

The astrochemical context of glycine molecule and an overview of the laboratory experiments describing formation, tautomerization mechanisms and glycine's behaviour under external radiation sources were detailed in Chapter 1 (see Section 1.3.1).

Theoretical studies centered on the relative stability and **the tautomerization mechanisms** have also been reported in literature. These tautomerization processes from neutral (N-Gly) to zwitterionic (Z-Gly) glycine conformers have been predominantly investigated by quantum mechanics chemical methods, including wave function based methods (MP2, CCSD) and DFT.[55,58–67] Furthermore, QM/MM approaches have been used in order to explore glycine molecule immersed in a water environment.[60,68–70,281] In these latter, several ways of modeling the **water environment** have been used: either as a continuum solvent model;[65] by explicit representation considering small water clusters (from one to ten H_2O molecules),[59,64,67,282] or as periodic boxes.[58,63] All the studies agreed that water environment stabilizes the Z-Gly conformer, but disagree on the water-glycine interactions that favor the zwitterionic form. One water molecule and excess of one electron were found sufficient for N-Gly to Z-Gly tautomerization,[64] whereas in another studies of mono- and dihydrated clusters only N-Gly was identified.[59]

Molecular Dynamics (MD) simulations of zwitterionic glycine in a water box within the Car-Parinello scheme resulted in 4.4 water molecules in the first solvation shell.[63] The existence of numerous local minima were found by Bachrach *et al.*[282] in their DFT study on the potential energy surface of glycine-$(H_2O)_7$ cluster. The authors identified thirty-five N-Gly and sixty-one Z-Gly minimum-energy conformations under water influence. This suggests that a statistical averaging of glycine conformations is needed to describe the micro-properties of glycine-water systems. The stabilization of trans (*t*) N-Gly in water solvent (*ttc* and *gtc*) in comparison to N-Gly in gas-phase was obtained from QM/MM-MD simulations at T=300K,[68] therefore **emphasising the role of explicit water molecules in the models**. Note that the later nomenclature *e.g. ttc* or *gtc* refers to glycine's dihedral angles. More details are given in the later sections (see Section 4.2.1).

The theoretical studies remain also inconclusive about the mechanism of N-Gly to Z-Gly tautomerization in water. An **intramolecular proton (H)-transfer (direct mechanism)** was mostly identified in these studies, which treated exclusively N-Gly conformer with cis (*c*) orientation between the amine and carboxyl group (*ccc* or *cct*),[59,61,65,146,283] a suitable glycine configuration for a facile direct H-transfer. Only few works report **a water-assisted tautomerization** that involves one or few molecules in the H-transfer mechanism.[68,282] These results are predominantly obtained if trans orientation between amine and carboxyl group were present among the studied conformers.

A long-term sampling, in the order of 0.5 to 1 ns, has been achieved with classical MD simulations, based on empirical force fields,[146,147] including a reactive force field.[58] Despite the longer sampling, feasible with classical MD simulations, the lack of polarization of the MM-atoms in the non-polarizable and non-reactive force fields was established insufficient to reproduce correctly enthalpy and entropy ratio in the tautomerization free energy.[58,146]

Nevertheless the theoretical works contributed valuable advances on the stability and tautomerization of glycine conformations in water medium, the accumulated knowledge is still incomplete. Moreover, it does not apply directly to the glycine conformational behaviour in the interstellar ice media. Hitherto, glycine tautomerization has been studied exclusively in view of its biological relevance at room temperature.[58,63,68,146,147,281]

In this chapter, we address the questions of glycine conformational space and the related transformations, including tautomerization mechanisms, in **water media and temperatures (from 20 to 250K) relevant with respect to the interstellar ice conditions**. A higher temperature of 450K was also included in order to model rapid heating of ices by cosmic rays or stellar winds matter fluxes. We considered **four glycine conformers**, each of them in four different orientations, embedded in water as supra-molecular models. **Water clusters** of higher sizes were studied: from n=4, 17, 27 H_2O molecules to a representative model (of about 350 H_2O molecules) of a **Low Density Amorphous (LDA) interstellar ice**. For these purpose, local optimizations and molecular dynamics studies were performed using **DFT and DFT(QM)/MM hybrid approaches**. Furthermore, a similar study was carried out using **DFTB and DFTB/MM class 1 FF methods**. Note that due to the DFT method's limitation, DFTB and DFTB/MM approaches appear as an alternative to enlarge the research field *e.g.* toward larger systems sizes, longer simulation times or improving the statistical averaging.

4.2 Initial Structures: N-Gly Isomers, LDA Ice and Water Cluster. General Strategy

In this section the strategy followed to build the different neutral glycine (N-Gly) initial configurations is presented. Note that, unless precised otherwise, this strategy is common for either DFT or DFTB investigation. The computational details respectively for DFT and DFTB explorations will be later report in Sections 4.3.1 and 4.4.1.

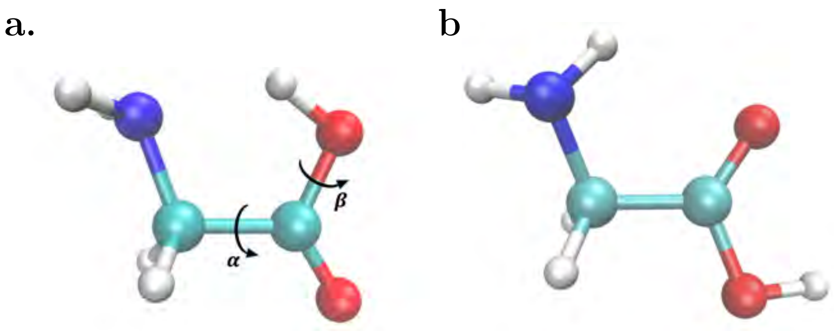
4.2.1 Neutral Glycine (N-Gly) Isomers

Eight most stable N-Gly isomers in gas phase were identified from CCSD(T) calculations.[55] In this set of glycine conformers the orientation of the carboxylic hydrogen with respect to N atom was considered. For the purpose of our work, only four out of the reported eight isomers (see Fig. 1.3) were chosen, disregarding the rotation around C-N bond. It represents various rotations of the -COOH group around the C-C bond and rotation of the carboxylic hydrogen as shown in Figure 4.1.

In this particular case, only DFT local optimization were performed for each N-Gly isomer (see Section 4.3.1 for computational details). Figure 4.1 shows the four N-Gly DFT optimized structures. In addition, values for the two characteristic dihedral angles are depicted, denoted $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (between -180 and $+180°$), corresponding to **N-C-C-O and to C-C-O-H dihedral angles** respectively. Finally, the DFT relative energies for each isomer are detailed.

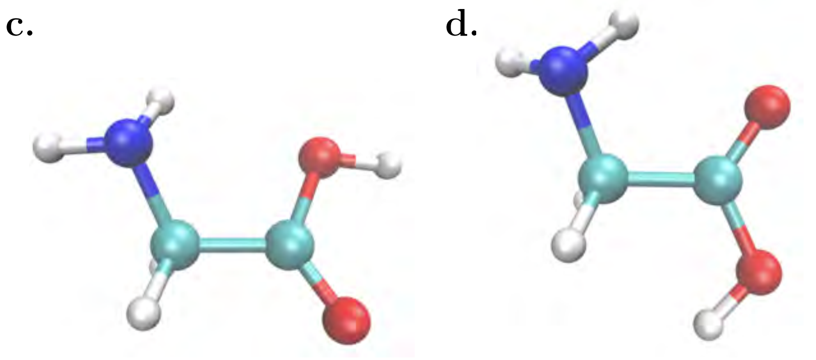
The α and β angles allow to depict **cis (abbreviated as *c*)** or **trans (abbreviated as *t*)** isomer's character. An angle with absolute value smaller than 30° allows to define a cis-conformation character, while an angle with absolute value larger than 90° corresponds to a trans-conformation character.

In the following, the N-Gly isomer will be denoted as ***xy*-N-Gly**, where x and y report the nature (c or t) of α and β angles, respectively. ***cc*-N-Gly** (Figure 4.1-a) refers to the neutral glycine isomers with $|\alpha|$ and $|\beta|$ angles close to 0° and ***tt*-N-Gly** (Figure 4.1-b) corresponds to a neutral glycine isomer with $|\alpha|$ and $|\beta|$ values close to 180°. In the two remaining isomers, α and β have different cis or trans character leading to ***ct*-N-Gly** (Figure 4.1-c) and ***tc*-N-Gly** (Figure 4.1-d) isomers.



cis-cis(*cc*)-N-Gly
$\alpha \sim [6°]$ $\beta \sim [-1.5°]$
$E_R = 0.0$ kcal.mol^{-1}

trans-trans(*tt*)-N-Gly
$\alpha \sim [162°]$ $\beta \sim [177°]$
$E_R = 2.6$ kcal.mol^{-1}



cis-trans(*ct*)-N-Gly
$\alpha \sim [-30°]$ $\beta \sim [-178°]$
$E_R = 4.0$ kcal.mol^{-1}

trans-cis(*tc*)-N-Gly
$\alpha \sim [166°]$ $\beta \sim [-2.5°]$
$E_R = 7.5$ kcal.mol^{-1}

Figure 4.1: DFT optimized neutral glycine (N-Gly) isomers. Cis and trans conformational character described by the α and β dihedral angles[284] .

Few preliminary studies were done on zwitterionic glycine (Z-Gly) in gas phase (more details about Z-Gly can be found in Section 4.4.3). Two DFT local optimizations were performed using two different initial Z-Gly geometries. As expected, the Z-Gly conformers transformed to N-Gly minimum energy structures during the DFT local geometry optimizations.

4.2.2 N-Gly isomers trapped in LDA ice

In order to model the interstellar ices, we have used the Low Density Amorphous (LDA) ice model from Ghesquière *et al.*[123] It consists of 352 water molecules, equilibrated previously from MD simulations.[285] Water molecules were described by TIP4P potential with an NPT ensemble to control the temperature and pressure of the interstellar ice conditions and the density of water was monitored to be approximately 0.958 g/cm^3. More details can be found in the reference.[123] Figure 4.2 shows this LDA ice model used in the present study with 20 Å cubic box dimensions.

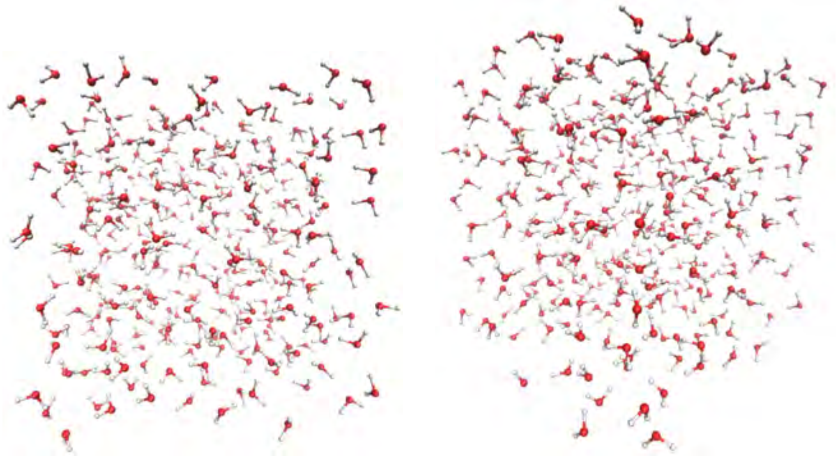


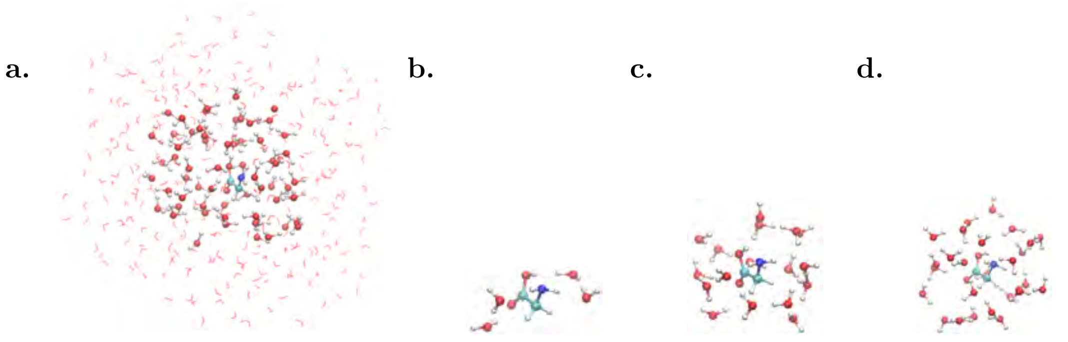
Figure 4.2: Benchmark LDA ice model from Ghesquière *et al.*[123] research (Two sites). This ice model was used (with no modifications) as the main bulk structure for all *xy*-N-Gly embedded system investigated in this research.

Each of the four minimum energy N-Gly isomers, described in Section 4.2.1 (Figure 4.1) was introduced into the LDA ice model. The N-Gly molecule was first positioned at the center of mass of the LDA ice model. Second, the closest 8 or 9 water molecules surrounding the N-Gly isomer were deleted to maintain a minimal distance of 2 Å between the N-Gly and water atoms. This criteria was established in order to avoid any convergence issue in the subsequent dynamical simulations. To further increase the sampling, we displaced the center of mass of *xy*-N-Gly with respect to the center of mass of LDA ice model and reoriented each *xy*-N-Gly isomer randomly in four different positions. We therefore generated **four different water neighbourhoods around every glycine isomer, resulting in sixteen *xy*-N-Gly-LDA ice systems**.

The QM-MM border was defined by a sphere with a radius $R_{cut-off} = 8.5$ Å and an origin at the glycine's center of mass. Doing so, the *xy*-N-Gly and its several first solvation shells, **containing 61-62 water molecules**, were treated at the DFT (or DFTB) level, while the remaining 282-283 water molecules were described at the MM level.

The structures were first optimized leading to a total of **sixteen stable *xy*-N-Gly-LDA ice models** (four per neutral glycine isomer), used as starting structures for the MD simulations. Note that the sixteen *xy*-N-Gly-LDA ice models were optimized by both DFT/MM and DFTB/MM class 1 FF methods.

As a representative example, Figure S.4-a reports a snapshot of LDA-embedded *cc*-N-Gly. Note that at temperatures higher than 170K, the LDA ice organization is not maintained.[123]



_cc_-Gly - (61/62) H_2O (QM) - 282 H_2O (MM) **_cc_-Gly- $4H_2O$** **_cc_-Gly- $17H_2O$** **_cc_-Gly- $27H_2O$**

Figure 4.3: Representative snapshots of the four different N-Gly embedded molecular systems extracted from the BOMD simulations. The atoms, treated at DFT level are shown as balls and those at MM level as lines. a) *cc*-Gly- 61 H_2O(QM)-282 H_2O(MM) model; b) *cc*-Gly- $4H_2O$ cluster; c) *cc*-Gly- $17H_2O$ cluster. d)*cc*-Gly- $27H_2O$ cluster[284] .

4.2.3 Water Cluster

The chemical evolution of glycine mostly depends on the interactions with the closest water molecules (first solvation shell). The dynamical evolution also depends on the water molecules further from the glycine. In order to characterize this effect, we have performed simulations in smaller water environments. They do not correspond to the most stable cluster structures (as in other works[58, 59, 286, 287]).

This finite size effect investigation was held, introducing each glycine isomer inside the ice model as described in the previous section with four various random displacements/orientations. The clusters were built by selecting only the glycine molecule and its n (4, 17 or 27) closest surrounding water molecules. The largest distance between glycine and water atoms are within 4.5, 6 and 7 Å for the clusters with n=4, 17 or 27 water molecules, respectively.

For these clusters, all atoms were treated at a DFT (or DFTB) level. The geometry optimizations were not carried out at variance in the case of *xy*-N-Gly -LDA ice models in order to maintain the original random reorientations of *xy*-N-Gly isomers inside the water clusters. **These xy-N-Gly-nH$_2$O initial structures allowed to generate 144 BOMD trajectories (3 water cluster sizes; 4 random positions/orientations, 4 conformational N-Gly isomers and 3 temperatures) of 10ps**. Some snapshots taken from these MD are exemplified in Figure S.4 b, c, and d.

4.3 DFT and DFT/MM Exploration

4.3.1 Computational Details

All DFT calculations were performed using the deMon2k code.[148] The DFT calculations were carried out with DZVP basis set,[276] as well as the automatically generated auxiliary basis set GEN-A2*[288] in combination with the PBE96 exchange-correlation functional.[275] The choice of the exchange-correlation functional was determined from previous theoretical studies where the PBE functional was shown to provide a fair description of glycine molecular systems.[53,289,290] The Restricted Kohn-Sham (RKS) SCF tolerance was set tighten between 10^{-6} and 10^{-9} a.u. In addition, dispersion corrections were added to the electronic energy through empirical C_6 coefficients.[291] Note that exclusively for the N-Gly gas phase local optimizations, the aug-cc-PVTZ basis set was used.

DFT/MM hybrid calculations have also been performed, within the electronic embedding incode additive QM/MM scheme in deMon2k.[169] The DFT region was treated as mentioned above and the MM region was described by the AMBER-FF99SB force field potential (TIP3P for water molecules)[292–295] which includes the Coulomb effects inside the MM part. QM-MM and MM interactions are described by the Lennard-Jonnes potential and Coulomb interactions between MM point charges atoms and QM electronic densities are considered, thus capturing the polarization of the QM region by the MM environment.

Local geometry optimization calculations were performed with the Broyden, Fletcher, Goldfarb, Shanno (BFGS)[212–215] Hessian update scheme, with a maximum root mean square (RMSQ) forces tolerance of 10^{-4} a.u.

Classical Born-Oppenheimer Molecular Dynamics (BOMD) were performed in the canonical ensemble. Each trajectory's length equals to 10 ps, with few exceptions noted in the following section. The integration time step was set to 1 fs. Linear and angular momentum conservation were forced with a tolerance value of 10^{-8} a.u. The temperatures were maintained by a chain of 5 Nosé-Hoover thermostats[272–274] with energy exchange frequency of 800 cm^{-1}. The BOMD simulations were carried out at three temperatures of 20, 250 and 450K.

4.3.2 Glycine Conformers and Reaction Path Analysis

In the course of the performed dynamics trajectories, a number of reshaping of the glycine molecule were observed. It involves conformerization between the different N-Gly forms, tautomerization toward a zwitterionic form, as well as the partial or total loss of an hydrogen atom leading to ionic compounds. In the following, we present and discuss the observed glycine conformational isomers and their ionized forms. A detailed description of reaction pathways is also provided.

4.3.2.A Neutral Conformers

The dynamical evolution of neutral glycine is tracked by the evolution of the two dihedral α and β angles (see Section 4.2.1). The changes of α and β along selected trajectories are exemplified in Figure 4.4.

For instance on the left panels, it is shown how at 20K, α and β dihedral angles remained almost invariant with respect to their initial values. The obtained small values of their standard deviations, reported in Table 4.1, confirm that the nature of the initial *xy*-N-Gly isomers are preserved along these dynamics. On the other hand,

***xy*-N-Gly**	$\alpha(°)$	$\beta(°)$
cc	5.3 ± 3.5	5.5 ± 3.0
ct	28.7 ± 11.1	167.1 ± 4.0
tt	157.8 ± 4.7	162.4 ± 6.6
tc	173.5 ± 3.7	8.9 ± 4.7
***tt*-Rotamer**	113.8 ± 6.5	176.3 ± 2.4

Table 4.1: Average absolute values for α and β dihedral angles and standard deviations for five selected trajectories (*cc*- *ct*- *tc*-and *tt*-N-Gly-27H_2O at T=20K).

Figure 4.4-right panels illustrate the transformation of N-Gly molecule from one isomer to another one. Such transformations were only observed for initial geometries

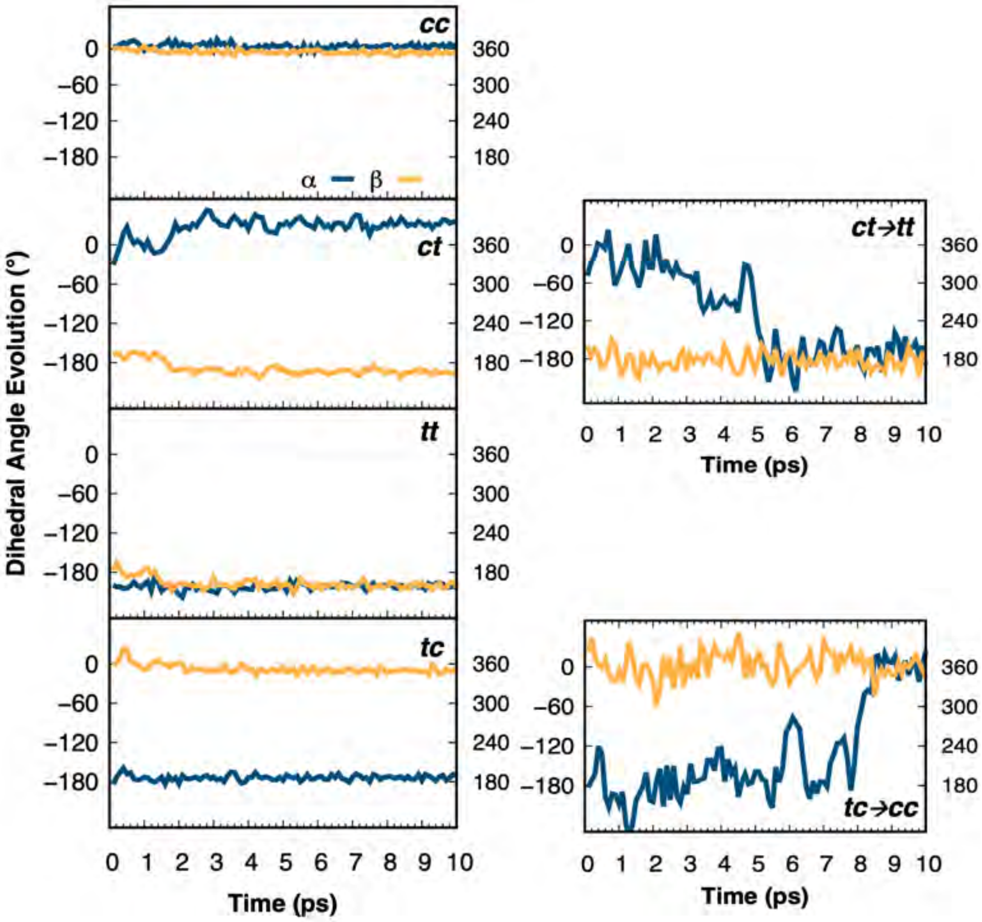
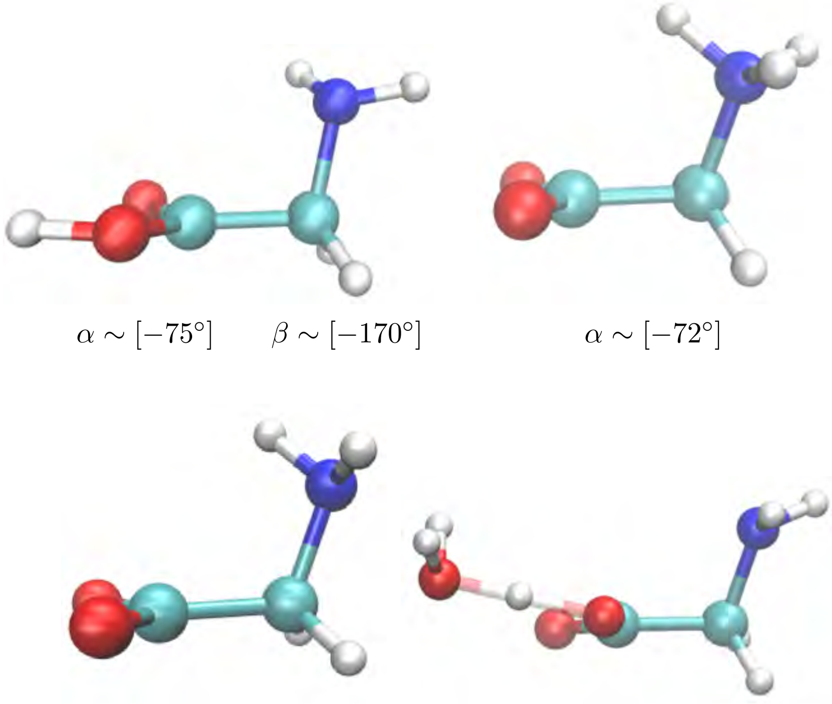


Figure 4.4: Left panels illustrate the monitored α and β dihedral angles for four *cc*-*ct*- *tc*-and *tt*-N-Gly-27H_2O molecular systems (dynamics at T=20K). On right panels, the isomer conformerization for both *ct*- and *tc*-N-Gly-17H_2O molecular systems are captured (dynamics at T=450K). A modulus of 2π was applied for angle values larger than 70°.

involving either *ct*-N-Gly or *tc*-N-Gly isomers. All conformerizations observed for the *ct*-N-Gly isomer, turned toward the *tt*-N-Gly one. Those observed for *tc*-N-Gly isomer resulted into *cc*-N-Gly one. These trends in isomers conformerization are consistent with the fact that *cc*-N-Gly and *tt*-N-Gly are the most energetically stable forms of neutral glycine in gas phase and in water.[68] In addition, rotations around the C-C bond (*ct*-N-Gly → *tt*-N-Gly and *tc*-N-Gly → *cc*-N-Gly) appear to be much more favourable than those around the C-O bond (*ct*-N-Gly → *cc*-N-Gly and *tc*-N-Gly → *tt*-N-Gly). This is consistent with the higher rotational barriers around the C-O bond than those around the C-C bond, computed at MP2[68] and CPCM-CCSDT[296] levels of theory.

Interestingly, a new geometry of neutral glycine (hereafter called **rotamer form**), unstable in gas phase, was observed during some trajectories. It corresponds to a rotation of the -OOH group around the C-C bond as shown in Figure

4.5-upper-left panel where the plane defined by the C-C bond and the nitrogen atom and the one defined by the same C-C bond and the carboxylic group of glycine are getting closer to orthogonality (more precisely $|\alpha| \sim 75°$). We note that this rotamer configuration appeared in some trajectories initiated from *cc*- *ct*- *tc*- or *tt*-N-Gly initial glycine isomers. In Figure 4.6, an example for rotamer formation from the *tt*-N-Gly isomer form is tracked from the evolution of α and β angles. It demonstrates that β dihedral angle oscillates around ±180°, thus indicating trans conformer, whereas the α dihedral angle shifted by ~60°. Similar conformational rotamers were attributed to transition state structures in the isomerization reactions of neutral glycine in water, studied with QM/MM umbrella sampling simulations at T = 300K.[68] Without fixing the degrees of freedom, our simulations revealed that these rotamers occur not only for the neutral but also, as will be seen below, for the zwitterionic and anionic forms of the glycine-water systems (see below and Figure 4.5), even in small water clusters with n=17 and 27 molecules.



$\alpha \sim [-75°]$ $\beta \sim [-170°]$ $\alpha \sim [-72°]$

$\alpha \sim [-85°]$ $\alpha \sim [-66°]$

Figure 4.5: Rotamer conformers, and corresponding α and β angles, for neutral (upper-left panel), zwitterion (upper-right panel), anion (lower-left panel) and proton delocalized (lower-right panel) glycine forms.

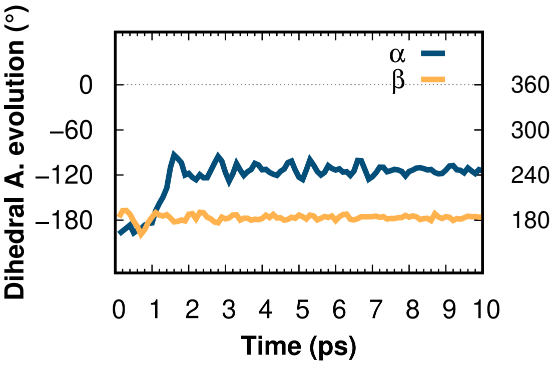


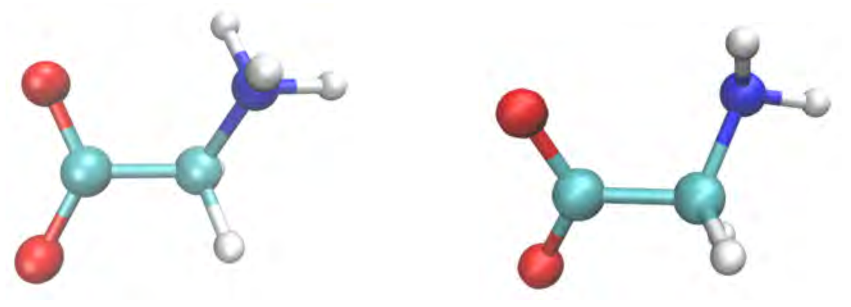
Figure 4.6: α and β dihedral angles dynamics evolution for the *tt*-N-Gly-27H_2O rotamer (dynamics at T=20K). A modulus of 2π was applied for angle values larger than 70°.

4.3.2.B Zwitterion Isomers

The zwitterionic glycine (Z-Gly) is another form of glycine molecule with a net neutral charge that has two charged functional groups (one positive and one negative). With respect to N-Gly, the zwitterionic form of glycine is characterized by the abstraction of proton from its carboxylic group and the hydrogenation of its amino group (see Figure 4.7-top,left). The **observation of Z-Gly during dynamics is a direct result of its stabilization by the surrounding water molecules**, since this form is unstable in gas phase as reported in Section 4.2. This form is particularly interesting as it is the most natural form of glycine reported in the literature in aqueous solution at room temperature or under the form of crystals.[297,298]

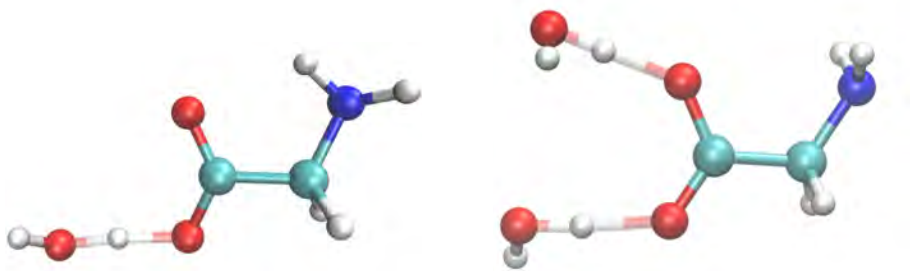
The dynamic evolution from N-Gly to Z-Gly molecular systems, appeared to be possible for any initial N-Gly isomer, excluding the *tt*-N-Gly. This can be related to the fact that *tt*-N-Gly isomer registers the largest distance between the proton (associated to the carboxylic group) and the nitrogen atom in glycine. In addition, from the energetic point of view, *tt*-N-Gly is particularly stable in gas phase.

As will be detailed in Section 4.3.3, the formation of Z-Gly is found to be more favourable when the initial N-Gly geometry corresponds to the *cc* isomer. This can be explained by the proximity between the hydrogen atom attached to the carboxilic group and the nitrogen atom in the amino group. Indeed, almost all the observed Z-Gly forms resulted from a **direct proton transfer**, where the carboxylic proton shifts toward the amino group, which is in agreement with previous studies on the tautomerization mechanism of glycine.[59,61,65,70,146] In Figure 4.8-top, **the direct proton transfer reaction path (N-Gly → Z_D-Gly)** is illustrated by the evolution of distances N-$H_{1,2,3}$ and O-$H_{3,X}$. H_1 and H_2 are the hydrogen atoms initially bonded to nitrogen, H_3 is the hydrogen atom initially involved in the carboxylic group and H_X is, at a given time, the closest hydrogen to any of the glycine oxygen atoms (O_g-H_X) or to the glycine nitrogen atom (N-H_X). We note the rapid transformation from N-Gly to Z_D-Gly that is observed only after 0.5 ps.



Zwitterion-Gly
($H_3\ N^+CH_2COO^-$)

Anion-Gly
($NH_2CH_2COO^-$)



H delocalized-Gly
$[OH \cdots H \cdots NH_2CH_2COO]^-$

Double H delocalized-Gly
$[OH \cdots H \cdots NH_2CH_2COO \cdots H \cdots OH]^-$

Figure 4.7: Illustration of four different glycine forms observed along dynamics: zwitterion glycine (top-left panel), anion (top-right panel) and one and two proton delocalization forms of glycine (bottom-left and right panels).

Figure 4.8-bottom, captures another type of proton transfer detected in the dynamics where the carboxilic proton migrates throughout the water molecules surrounding the glycine molecule. This **concerted proton transfer** was mostly observed with initial *ct*- and *tc*-N-Gly isomers and allowed to overpass the large distances between the two functional groups in these isomers. Only two concerted proton transfer mechanism were observed for initial *cc*-N-Gly isomers, out of the total of 29 Z-Gly forms registered. Figure 4.8-bottom shows the longer transfer's life time for the concerted proton transfer mechanisms with respect to the direct one. Indeed, the first proton transfer from the carboxylic group to one water molecule occurs at around 0.8 ps, followed by 2 ps of dynamics where the proton migrates

through different surrounding water molecules. Finally, a new bond between a proton from another H_2O molecule is formed with the nitrogen atom, terminating the **transformation from N-Gly → $\mathbf{Z_C}$-Gly** (after ~2.6 ps dynamics).

Let us finally mention that a **rotamer Z-Gly** form was also identified (see Figure 4.5-top right) with perpendicular carboxylic / nitrogen plans, similarly to the previously reported N-Gly rotamer cases.

4.3.2.C Anionic Forms

In some trajectories, a negatively charged glycine form was observed, namely a dehydrogenated glycine molecule resulting from the loss of the proton of the carboxylic group (see Figure 4.7-top right).

These forms were observed in trajectories with *tc*-, *ct*- and *tt*-N-Gly initial isomers, but not in those starting from the *cc*-N-Gly isomer. This can be explained from the proximity between the amino and carboxilic groups in *cc*-N-Gly isomer, leading to a direct proton transfer between these functional groups rather than the proton loss (see Section 4.3.2.B). On the other hand, for *tc*- *ct*- and *tt*-N-Gly isomers, other mechanisms involving water molecules are favoured. The already presented concerted proton transfer mechanism requires a specific molecular organization of the water molecules.

Figure 4.9-top monitors the N-H_1, N-H_2 and O_g-$H_{3,X}$ bond distances for a trajectory leading to glycine anion in a cluster with 27 water molecules. At 1.9 ps, the carboxilic proton is transfered to a water molecule but, contrary to the Z-Gly form, this proton remains attached on a water molecule and no N-H bond is formed after 10 ps.

Similarly to the dynamics of the neutral and zwitterion glycine forms, a rotamer configuration was also observed for the glycine anion (see Figure 4.5-bottom left).

4.3.2.D Hydrogen Delocalization Patterns

The last glycine form observed in the dynamics consists in **a proton delocalization in between the caboxylic functional group of glycine and one (or two) water molecules**. The carboxilic hydrogen is shared between two oxygen atoms, namely that of the closest water molecule and the one to which it was ini-

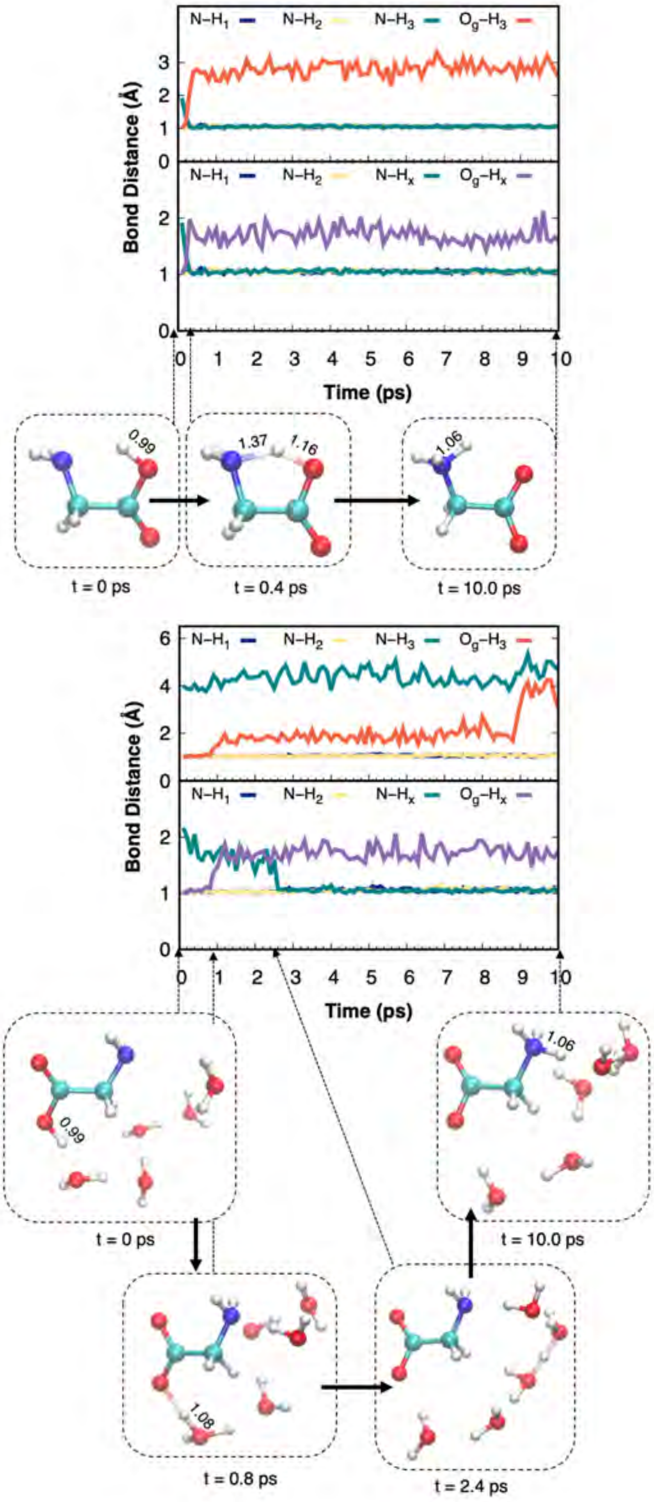


Figure 4.8: Snapshots and time evolution of characteristic bond distances for two direct (top) or concerted (bottom) zwiterionic tauromerization mechanisms, with mechanisms forms extracted from clusters simulations. Top: *cc*-Gly-27H$_2$O (T=250K) and bottom: *tc*-Gly-27H$_2$O (T=250K). O$_g$ refer to glycine oxygen atoms. H$_1$ and H$_2$ (resp. H$_3$) identify the protons which, at $t = 0$ ps, belong to the amino (resp. carboxylic) group. At each time, O$_g$-H$_x$ (resp. O$_w$-H$_x$) is the closest distance between any proton and any of the carboxylic oxygen (resp. water oxygen) atom whereas N-H$_x$ is the third closest distance between the nitrogen atom and any proton.

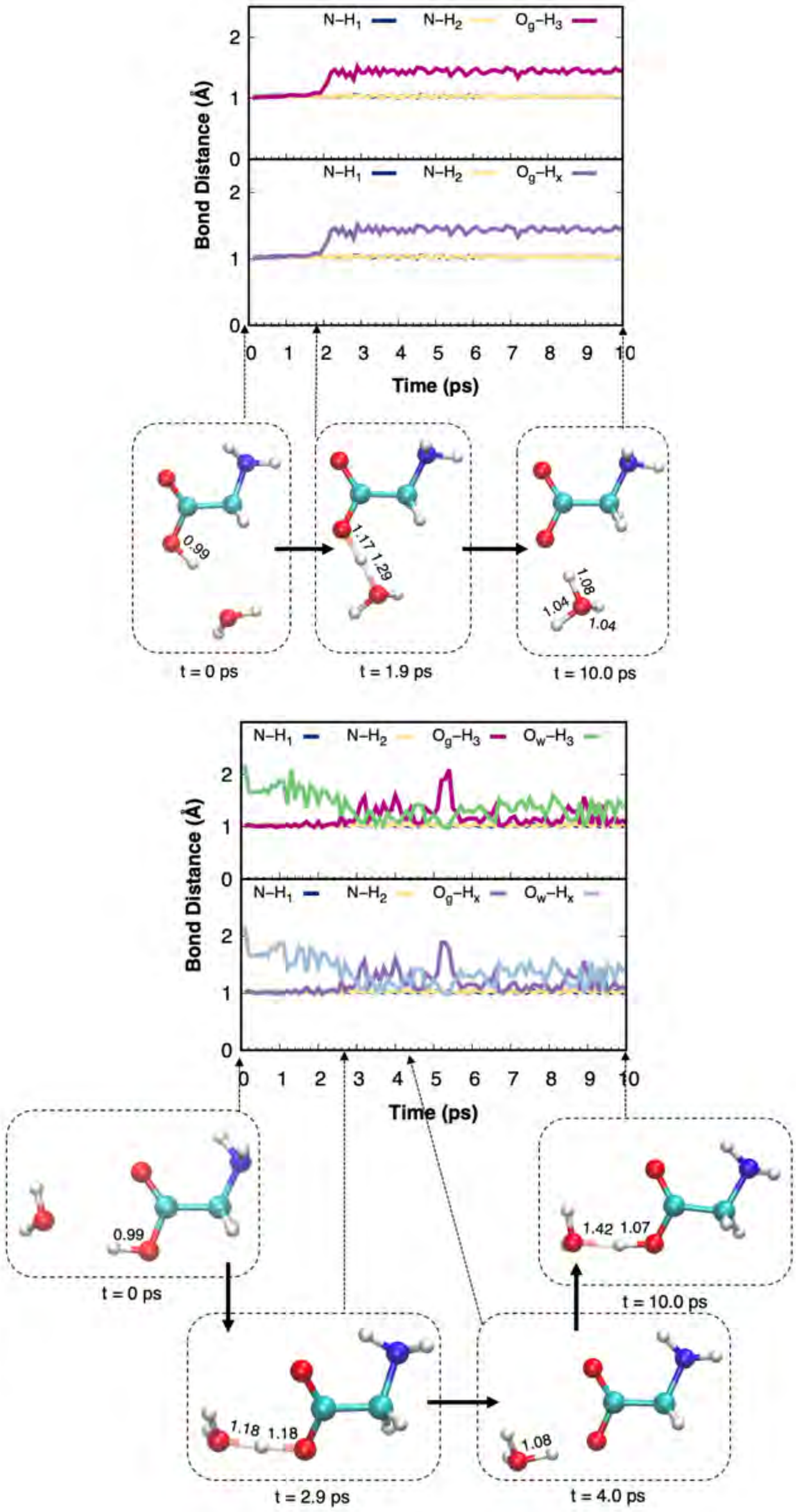


Figure 4.9: Snapshots and time evolution of characteristic bond distances for two different transformation mechanisms leading to anion (top) or H-delocalised (bottom) forms extracted from clusters simulations. Top: *tc*-Gly-27H_2O (T=20K) and bottom: *tt*-Gly-17H_2O (T=250K). O_g and O_w refer to glycine or water oxygen atoms. H_1 and H_2 (resp. H_3) and O_g-H_x (resp. O_w-H_x) have the same representation as in Figure 4.8.

tially bonded (see Figure 4.7-bottom left). Sometimes, the hydrogen atom remains at equal distance between the two oxygen atoms (see Figure 4.9-bottom capture at t=2.9 ps), whereas, in some others, it oscillates between two oxygen atoms (see Figure 4.9-bottom snapshots at t=4.0 ps and t=10.0 ps).

Proton delocalized glycine can be defined as an **oscillation between an unstable anion and a neutral glycine forms**. The interchanging of H proton and bond formation at different stages of the dynamics is illustrated in Figure 4.9 (bottom) which shows the crossing curves corresponding to O_g-$H_{3,X}$ and O_w-$H_{3,X}$ distances where O_g and O_w are the involved oxygen atoms on the glycine and water molecules, respectively, and O_w-H_X is the distance between O_w and its third closest hydrogen atom.

Very rarely, a **mechanism with two delocalised protons** could be identified during the dynamics (see Figure 4.7-bottom right). The dehydrogenation of one of the carboxylic oxygen atom is rapidly followed by the hydrogenation of the other one, in a concerted hydrogen transfer mechanism through the water environment. This double dehydrogenation / hydrogenation mechanisms occurs several times along the trajectories.

Similarly to the anionic glycine form, the proton delocalization patterns were observed only in dynamics with initial *ct*-, *tc*- and *tt*-N-Gly isomers. Due to the proximity of the two functional groups in *cc*-N-Gly isomer, N-Gly → Z-Gly transformation is favourable and prevents the hydrogen delocalization.

Finally, identical to previous glycine forms, a proton delocalised rotamer was observed in some dynamics as illustrated in Figure 4.5 (bottom right).

Anionic or proton-delocalized intermediates or transition state structures were not considered in former works, to the best of our knowledge, in mechanistic studies of glycine tautomerization reactions.

Let us finally mention, that addressing the proton delocalization character[299–302] would in principle require a quantum treatment of hydrogen atoms, for instance making use of Path Integral MD schemes.[303,304] These latter studies are beyond the scope of the present work.

4.3.3 Effect of water environement complexity vs system temperature

In this section, a deeper quantitative analysis is performed to discuss the **role of the water environment size and of the temperature on the glycine evolution**. We address the dependence on the temperature and the size of the water environment of the various *xy*-N-Gly-nH_2O and *xy*-N-Gly-LDA structural transformations, reported in the previous sections.

The results of the molecular dynamics simulations at T=20, 250 and 450K for 3 ps and their continuations until 10 ps are gathered in Figure 4.10-A and -B. We mention that, in the case of the simulations with the LDA ice model, only one out of four trajectories achieved at 3 ps could be continued up to 10 ps due to computational cost. The latter were taken among those which did not show any transformation in the simulated first 3 ps.

For convenience, we will call in the following the water clusters of different sizes as: **(i) small clusters** to refer to clusters with n=4 H_2O molecules; **(ii) intermediate clusters** to refer to clusters with n=17 and 27 water molecules and **(iii) LDA ice** to refer to the largest model (343-344 water molecules). Similarly, we will call T=20, 250K the **low temperatures** and 450K the **high temperature**.

We first discuss the dynamical evolution of the ***cc*-N-Gly isomer**. In Figure 4.10-A and B upper panels, the initial neutral *cc*-N-Gly geometry was maintained in some trajectories while the tautomerization toward a Z-Gly form was observed in the trajectories of other replicas (see Section 4.3.2.B). The latter occurs most frequently through a **direct proton transfer**. The formation of Z-Gly conformer was often observed (of about 75 and 100 %) in the intermediate cluster sizes, at the three temperatures (light blue histogram columns in Figure 4.10). In the case of the LDA ice model, the zwitterionic form appears in 25 % at T=20 and 250K, whereas the high temperatures (T=450K) favoured the Z-Gly tautomerization.

The following common features were captured for trajectories with initial ***tt*-, *ct*- and *tc*-N-Gly-n$\mathbf{H_2O}$(LDA) structures** (Figure 4.10-A and -B middle and lower panel). For all the small clusters, glycine molecules remained practically in their neutral forms whatever the temperature, the initial isomer structure and the simulation length (3 or 10 ps). Indeed, in these small systems, the water network is insufficient to completely surround the glycine molecules, which limits the transformation possibilities. This is in line with recent studies on the free-energy surfaces at

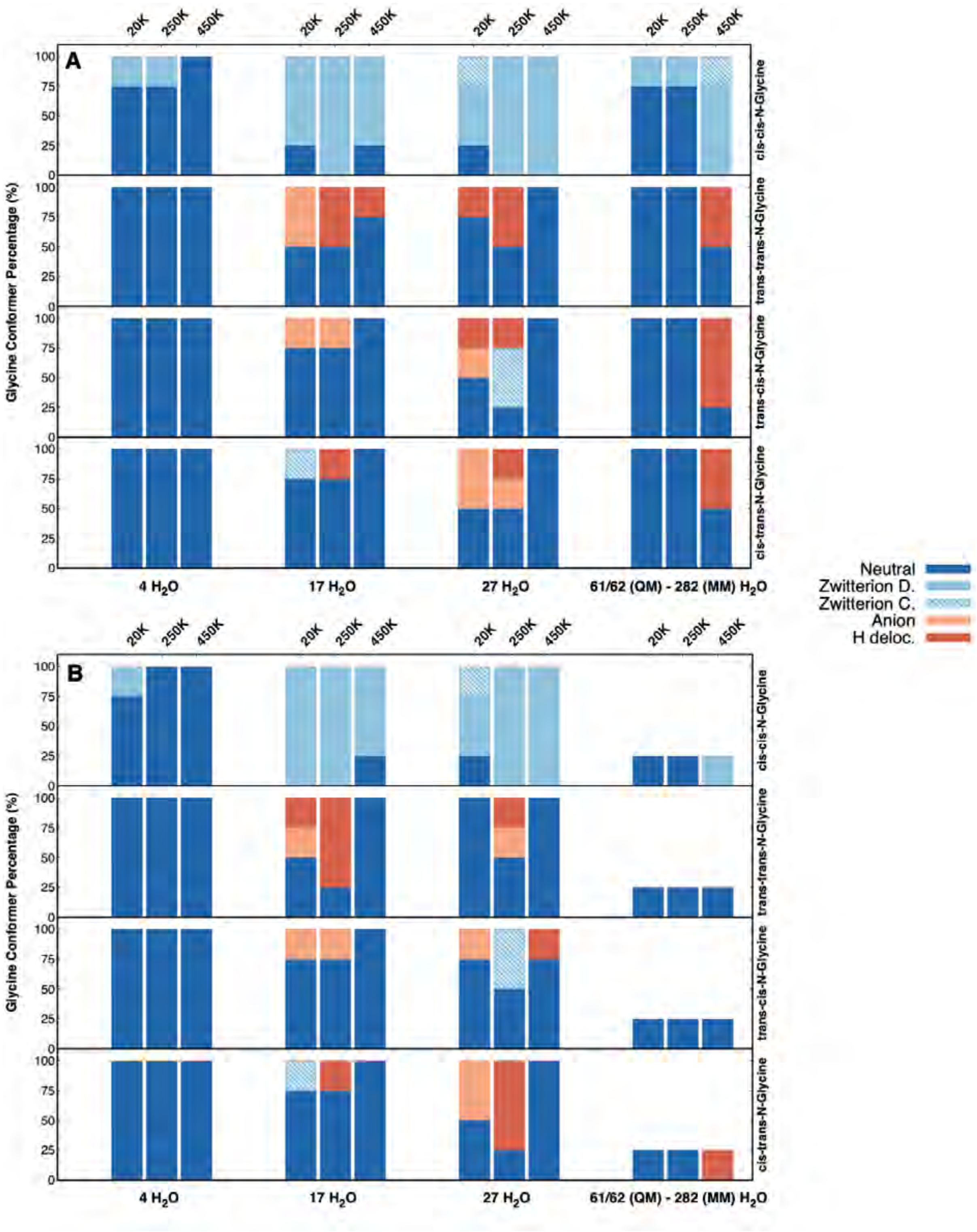


Figure 4.10: Dominant glycine transformations observed during the first part (3 ps, top A) and full part of trajectories (10 ps, bottom B) differing by the initial isomer, its initial localization, the number of water molecules and the temperature. For the largest systeme (LDA ice), a single simulation was performed up to 10 ps per isomer/size/temperature due to the great computational time requirements.

300K with biased meta-dynamics simulations, demonstrating that the zwitterionic form in Gly(H_2O)$_4$ cluster is unstable.[305]

Furthermore, in the intermediate cluster sizes all the transformations reported in Section 4.4.3 can be observed, namely anion (Fig. 4.10 orange columns), proton delocalization (red) and zwitterion glycine. When N-Gly tautomerization toward Z-Gly form was observed, it was in every cases due to a **concerted proton mechanism**. This is in line with the frequent observations of anion and proton delocalization glycine forms, as these patterns can be seen as preliminary forms for such concerted mechanism. It can be extrapolated that in longer dynamics ($>$ 10 ps) Z-Gly tautomerization might take place. Surprisingly, N-Gly form is seen to be always maintained at the highest 450K temperature. This can be explained by the fact that glycine molecule moves from the cluster's center toward its borders. Indeed, the water molecules preferred to build a hydrogen-bond network between them by displacing the glycine molecule toward clusters borders, where the water coordination effect is smaller and therefore less favourable to chemical transformations of glycine. Note that for the cases of *ct*- and *tc*-N-Gly isomers, the migrations of glycine toward the clusters' borders was often associated with a conformer transformation ($ct \rightarrow tt$ or $tc \rightarrow cc$).

On the other hand, in **the ice representative model**, such a glycine migration toward the surface was not observed due to the larger number of water shells around the glycine system. Rearrangements of water molecules in this system did not occur because **the hydrogen-bond network between the water molecules in the outer solvation shells of glycine was already formed**, at variance to the small and intermediate sizes models. Therefore, in *tt*-, *ct*- and *tc*-N-Gly-LDA ice molecular systems, **the glycine molecule remains neutral in all simulations at low temperatures**. At 450K, proton delocalization patterns are observed. Note that these proton delocalization patterns were also sometime observed on very short and limited periods in some trajectories denoted as "Neutral", because such transfers were not representative of the overall trajectory. This points out to the unstable character of proton delocalization form in these cases.

When comparing series of simulations at both, 3 ps (Fig.4.10-A) and 10 ps (Figure 4.10-B), it appears that trajectories characterized as proton delocalization up to 3 ps were later on identified as anion forms until the end of the 10 ps dynamics.

The opposite transformations (from anion to proton delocalised) were also observed. The oscillations between these two unstable forms, even for the anion conformer, can be explained by the proton proximity between carboxilic group of glycine and water molecule involved in this transfer. The proton originally attached to the carboxilic functional group, remains unbounded but close to the carboxylic group. Similarly, considering simulations only up to 3 ps, some proton delocalization forms were identified and latter stabilized toward the neutral form. This points out the importance of the length of the trajectories to discuss the tautomerization of glycine in water.

4.3.4 Conclusions for the DFT and DFT/MM Exploration and Astrochemical Implications

In this study consisting on the exploration of the tautomerization mechanisms and the effects of water environment and temperature in different glycine molecular systems, the following general trends were identified:

- In the ice representative models, the neutral glycine is stabilized at low temperatures (20 and 250K), but not at the high temperature (450K). In the latter case, the tautomerization transformation towards Z-Gly form were mostly observed for the *cc*-N-Gly-LDA, whereas for the other isomers proton delocalization forms were observed.
- The trends reported in the ice representative model, and in particular the tautomerization, can only be observed when explicitly considering the large water network around glycine, the simulations with a smaller number of water molecules leading to different conclusions. In the case of n = 4, 17 and 27 H_2O clusters, the glycine solvation shell tends to disappear because of the water hydrophilic character, which favours the aggregation of water molecules between them, thus displacing the glycine toward cluster borders. Consequently, some of the glycine functional groups (carboxilic or amino) remain unbounded to water, which stabilizes the N-Gly conformer. This is well evidenced for the case of small (-4H_2O) cluster. Even at high temperatures (T=450K), the lack of water embedding the glycine molecule prevents the transformation to Z-Gly. This is consistent with previous free energy studies that evidenced the stabilization of neutral glycine in gas phase or surrounded by few water molecules.[59, 68, 305]
- With sufficient number of molecules, the efficiency of the tautomerization process (N-Gly → Z-Gly) depends on the nature of the glycine isomers. This is

also the case of the the associated mechanism. For instance the direct mechanism occurred only from the *cc*-N-Gly-nH2O isomers, whereas for *ct*- and *tc*-N-Gly molecular systems, the proton transfer was assisted by the water molecules in a concerted mechanism.

This work enforces the idea that considering the various glycine isomers is mandatory to provide a complete picture of the tautomerization mechanisms. In addition to the initial isomers structures, the other factor driving the tautomerization is the representation of the water environment. Both water and glycine dynamics are also temperature dependent. For intermediate size clusters, water - water aggregation occurs faster at T = 450K, which stabilises N-Gly. At this temperature and in the large size water model (starting from LDA-ice model), *cc*-N-Gly is fully converted to Z-Gly, whereas for the other three isomers, the delocalization between H_2O and -COO- group dominates.

In addition, we demonstrated that **the glycine tautomerization process is strongly influenced by the nature of its conformational isomer**. The Z-Gly isomer is relatively rapidly formed from *cc*-N-Gly predominantly via a direct H-transfer, whereas the *ct*-, *tc*- and *tt*-N-Gly isomers undergo different transformations such as anion and proton delocalised forms.

From the stand point of Astrochemistry, our study suggests that glycine in the cold interstellar ices should remain neutral up to at least 250K consistently with the experimental results in Iopollo *et al.*[26]. These authors reported that glycine is formed mainly in its neutral form on interstellar ice and that heating the system resulted in the loss of neutral glycine isomer's spectral signature, and signature of zwitterionic glycine started to appear. This is in line with our simulations of the ice model, where transformations toward zwitterionic forms or H-delocalized forms are favored when increasing the temperature.

From the stand point of chemical simulations, our study strongly suggests that modeling interstellar bulk chemistry requires few water layers (our LDA ice model) surrounding the reactants to avoid self-aggregation of the water molecules.

4.4 DFTB and DFTB/MM Exploration

4.4.1 Computational Details

All calculations were performed using the deMonNano code[149] employing the second-order DFTB level of theory with BIO parameters. A Fermi distribution temperature of 100K was used along with the class IV/charge model 3 (CM3) charges with a parametrization of $D_{\text{O-H}} = 0.129$ and $D_{\text{C-H}} = 0.1$[306,307] (recall Eq. 2.62 in Chapter 2). The London dispersion function with a damping function correction was also included in the calculations (see Eq. 2.60 in Chapter 2)[192] .

DFTB/MM hybrid calculations were performed within the new electronic embedding incode additive QM/MM scheme in deMonNano (see Chapter 3). The DFTB region was treated as mentioned above and the MM region was described similarly than in the DFT/MM calculation, making use of the AMBER-FF99SB force field potential (TIP3P for water molecules)[292–295] which includes the Coulomb interactions inside the MM part.

Local geometry optimization calculations were performed with the Steepest Descent Method (SDC). Classical Born-Oppenheimer Molecular Dynamics (BOMD) were performed in the canonical ensemble. The total simulation time of 10ps was chosen and the time step was set to 0.5fs. Linear and angular momentum conservation were forced and a spherical potential wall (1Ha) of 20Å diameter around the glycine molecular systems was used. The temperatures were maintained by a chain of 5 Nosé-Hoover thermostats[272–274] with energy exchange frequency of 800 cm^{-1}. The BOMD simulations were carried out at the same three temperatures of 20, 250 and 450 K.

4.4.2 DFTB potential Reparametrization

Preliminary calculations for both local optimization and MD simulations were carried out using BIO parameters. Nevertheless, the preliminary local optimizations of Z-Gly molecules in gas phase showed a poor description of the sp^3 hybridization in the amino group. This lack of description was translated in a **C-C glycine's bond breaking** during the local optimizations. In addition, from a preliminary study of the effect of water environment complexity in glycine systems performed using DFTB2 approach with BIO parameters (computational details in Section 4.4.1) showed that **interactions between glycine and water molecules were poorly described**. This can be seen in in Figure 4.11 which collects all trajectories for *cc*, *ct*, *tc*- and *tt*-N-Gly-nH_2O glycine cluster systems. Contrary to the DFT study (see Section 4.3.3) where glycine conformers, tautomerization processes and partial or total deprotonation processes were observed (see Figure 4.10), very few transformations were captured in the DFTB simulations with BIO parameters. Only 3.5% of transformations were observed, all consisting of an unstable proton delocalization (red histogram columns).

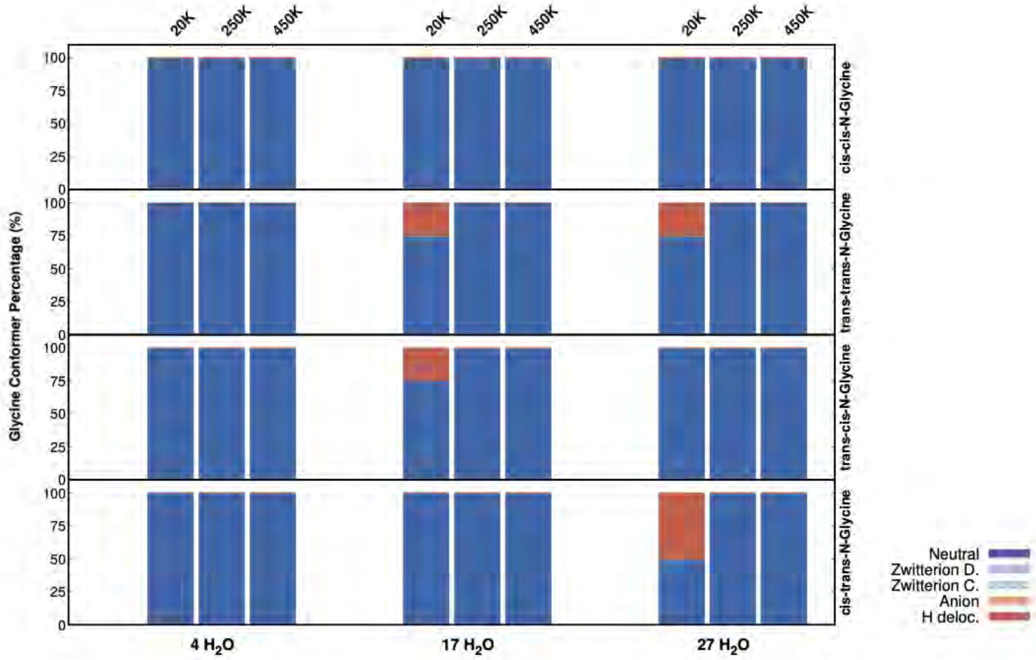


Figure 4.11: Trajectories for DFTB2 calculations with BIO parameters (10ps dynamics). For the largest systeme (LDA ice), four simulations were performed up to 10ps per isomer/size/temperature.

It is for these reasons, that a need of reparametrization appeared as was also suggested in M. P. Kroonblawd *et al.*[205] work. This reparamentrization aims to decrease the repulsive character between the C-C, N-H and O-H pairs of atoms.

Based on a set of DFT forces and making use of the **Force Match (FM) extension** implemented in deMonNano, a reparametrization of the $V_{rep}^{ab}(\mathbf{R}_a - \mathbf{R}_b)$ repulsive term in DFTB forces was done (see Section 2.3.3). This reparametrization was done for the **C-C (V_{rep}^{CC}), N-H (V_{rep}^{NH}) and O-H (V_{rep}^{OH}) atomic pairs**.

A set of DFT calculations were performed, consisting of: **(i)** two local optimizations for the neutral and zwitterionic glycine gas phase molecules (*cc* and *tt* isomers). **(ii)** Molecular dynamics simulations of 3ps for each *cc*- *ct*- *tc*- and *tt*-N-Gly-17H_2O system at both 20 and 250K. Note that computational details reported in Section

4.3.1 were maintained. The previously listed simulations were combined in different manners when **training the Force Matched DFTB potentials**, leading to different reparametrized V_{rep}. Figure 4.12 compares the original V_{rep} from BIO parameters (black line) with the new reparametrized ones (V_{rep}^{CC}, V_{rep}^{NH} and V_{rep}^{OH}) from different simulation trainings. The distinctions between the training simulations, came from the use of different R$_{cut-off}$ radius, the inclusion of dispersion or from the chosen MD snapshot simulations, different isomers and temperatures.

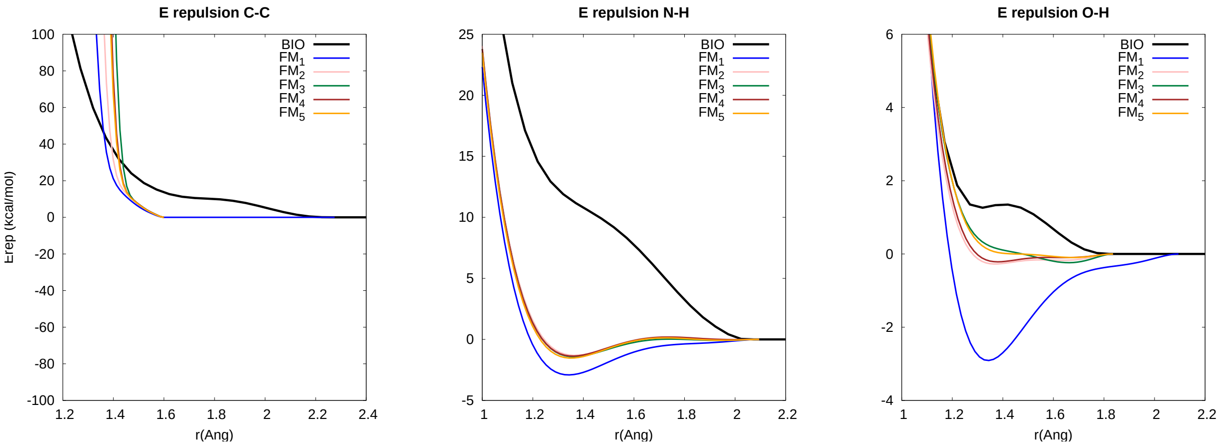


Figure 4.12: Curves for DFTB potential reparametrizations carried out by the Force Match extension in deMonNano package. The DFTB potential computed by BIO parameters (black line) is compared with the repulsive reparametrized potential from different DFT trainings (color lines).

From the different set of reparametrization curves in Figure 4.12, the orange curves were chosen as the preferred V_{rep}^{CC}, V_{rep}^{NH} and V_{rep}^{OH} new DFTB repulsive potentials. This latter, used the best trainning set, *i.e* includes London dispersion correction and the most complete set of DFT MD snapshots: the *cc*- and *tt*-N-Gly gas phase DFT local optimization forces and the *cc*- and *tt*-N-Gly-17H$_2$O DFT MD simulation forces at both 20 and 250K. Figure 4.13 shows how this new reparametrization **enhances the attractive character of the C-C, N-H and O-H atomic pairs potential**.

In practice, for the particular case of glycine molecular systems treated in these work, the repulsive energy terms in the Slater-Koster tables (see Section 2.3.3) for the BIO parameters are replaced by the new reparametrized DFTB repulsive potentials. Note that these modifications are only done for the C-C, O-H and N-H atom pairs. It is important to stress that this reparametrized DFTB potential are specific for glycine molecular systems. Making use of this reparametrization in any other molecular systems or performing new reparametrization on the DFTB repulsive potentials would require new benchmark studies.

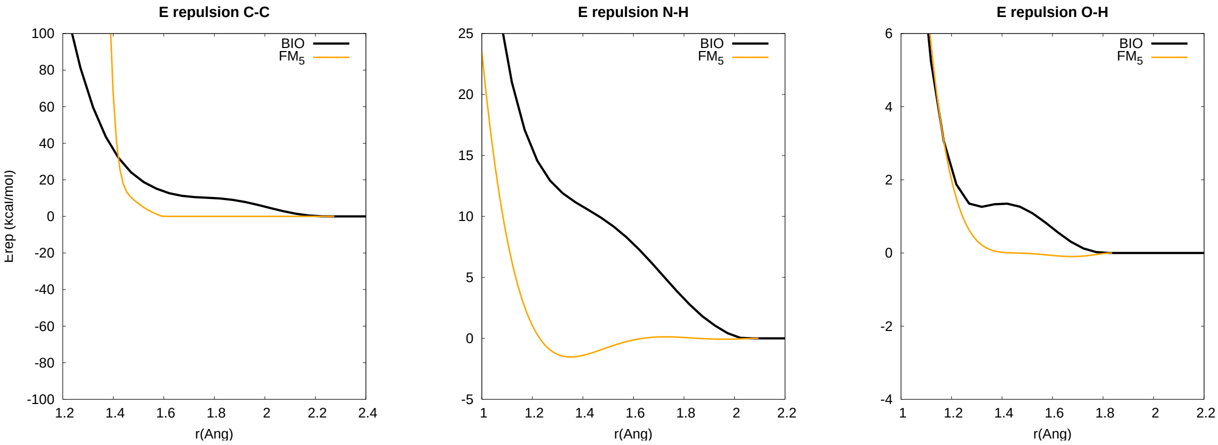


Figure 4.13: Comparison of DFTB repulsion potential from BIO parameters (black line) with the chosen repulsive reparametrized potential.

4.4.3 Reparametrized DFTB Potential for Glycine Conformers and Reaction Path

The first goal was to ensure that the same conformers and tautomerization mechanisms observed in the DFT study (see Section 4.3) are also captured in the DFTB calculations. From the several local optimizations and the MD dynamics performed in xy-N-Gly-nH$_2$O and xy-N-Gly-LDA systems using the DFTB and DFTB/MM class 1 FF approaches with the new reparametrized BIO parameters, great improvements were achieved:

1. From the Z-Gly local optimizations, thanks to the new reparametrized potential, **the C-C bond breaking was avoided**.
2. In the local optimizations, the neutral glycine minimum energy structures were now accessible. This is in line with DFT results (see Section 4.2.1).
3. It is observed that this new reparamentrization favours the O-H and the N-H interactions between glycine and water molecules. Therewith, in the MD DFTB calculations, the same glycine conformers reported in Section 4.4.3 were observed. The zwitterionic glycine, the anionic and the proton delocalized forms were captured (see Figure 4.7). In addition, the rotamer forms for all four glycine conformers (see Figure 4.5) were also detected.
4. In the particular cases where Z-Gly conformers were observed during dynamics, **both direct and concerted proton transfer** reaction paths were observed. This is consistent with the DFT results (see Section 4.3.2.B)

In the following, the study of water environment and temperature effects on glycine system's is reported for trajectories obtained by the DFTB approach using the new reparametrized V_{rep}^{CC}, V_{rep}^{NH} and V_{rep}^{OH} DFTB repulsive potentials.

4.4.4 Effect of water environement complexity vs system temperature and DFTB comparison

In this section, we address the dependence on the temperature and the size of the water environment of the various *xy*-N-Gly-nH_2O and *xy*-N-Gly-LDA structural transformations.

The results of all trajectories at T=20, 250 and 450K for 3ps and their continuations until 10ps are gathered in Figure 4.14-A and -B. Without the computational resources limitations in DFT calculations, all four trajectories achieved at 3 ps could be continued up to 10 ps for the LDA ice model.

The same nomenclature employed in the DFT analysis is used in this section. Therewith, clusters will be divided into small (4 H_2O molecules) and intermediate (17 and 27 H_2O clusters), followed by the LDA ice model (343-344 water molecules). Similarly, we will call T=20, 250K the low temperatures and 450K the high temperature.

In Figure 4.14-A and B upper panels, the dynamical evolution of the ***cc*-N-Gly isomer** is captured. The initial neutral *cc*-N-Gly geometry was maintained in some trajectories while **the main trend observed was the tautomerization toward the Z-Gly conformer which is consistent with DFT simulations**. Besides, few transformations to anion and proton delocalized forms were also observed. As it was found in the DFT calculations, the Z-Gly transformations were mainly due to a direct proton transfer. The formation of Z-Gly conformer was often observed in the intermediate cluster sizes, at the three temperatures. In the case of the LDA ice model, the zwitterionic form appears about 75% at T=20 and about 100% at 250 and 450 K. While the DFT calculations confirmed that the high temperature favoured the Z-Gly tautomerization, this temperature dependence cannot be extracted from the DFTB simulations.

For trajectories with initial ***tt*-, *ct*- and *tc*-N-Gly-nH_2O (LDA) structures**, for all the small clusters glycine molecules remained neutral whatever the temperature and the initial isomer structure (Figure 4.14-A and -B middle and lower panel).

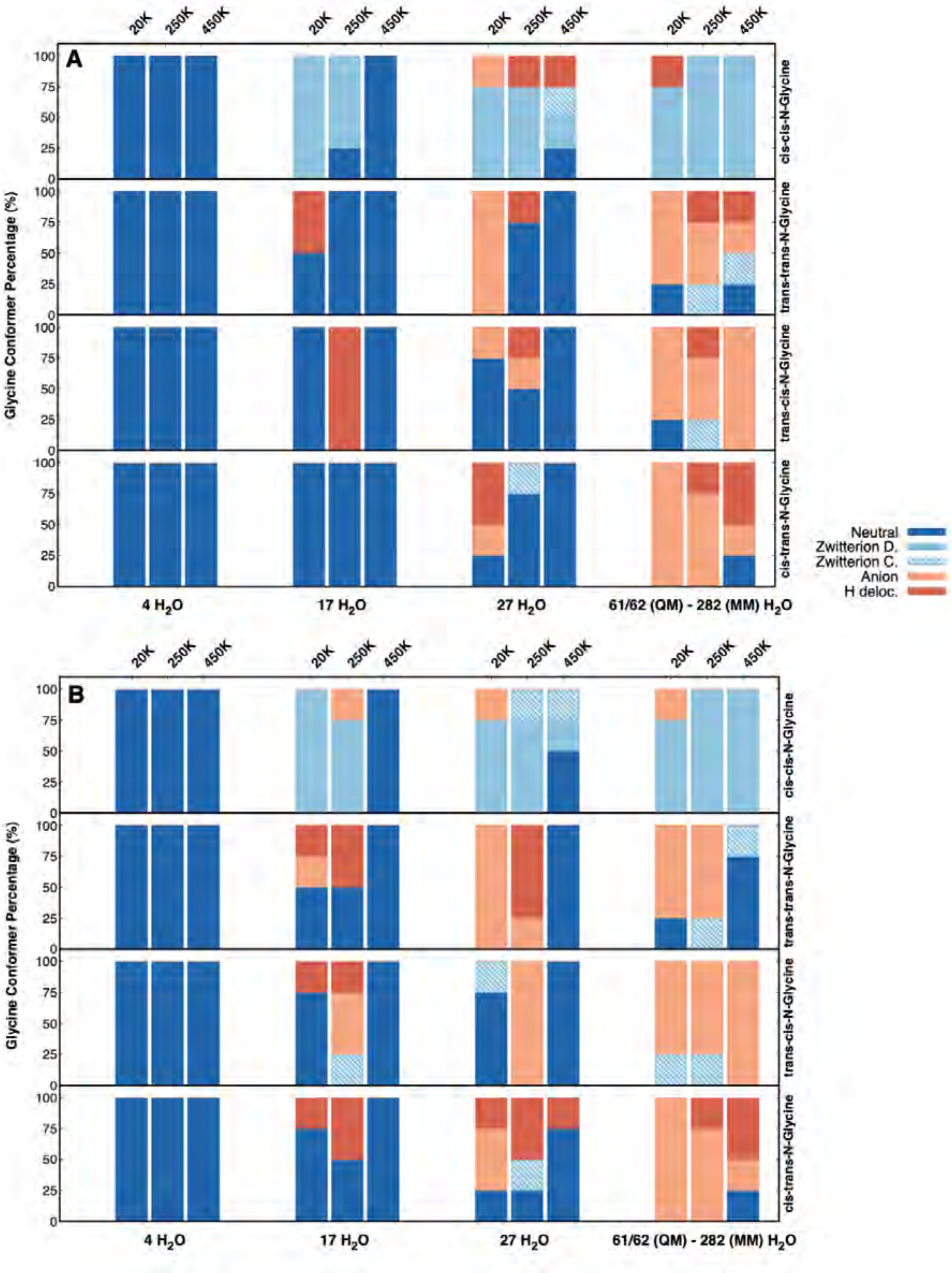


Figure 4.14: Dominant glycine transformations observed during the first part (3ps, top A) and full part of trajectories (10ps, bottom B) differing by the initial isomer, its initial localization, the number of water molecules and the temperature.

Similarly to the DFT trajectories, in these small systems, the water network is insufficient to completely surround the glycine molecules, which limits the transformation possibilities. Furthermore, in the intermediate cluster sizes all zwitterionic (concerted proton mechanism), anion or H-delocalized forms can be observed at a higher percentage than in DFT calculations. N-Gly forms are always maintained at the highest 450K temperature in line with the DFT results. In some cases, the same mechanism observed in DFT simulation was also captured by the DFTB trajectories at high temperature. The glycine molecule moves from the cluster's center toward its borders. Again, the water molecules preferred to build a hydrogen-bond network and the glycine molecule is displaced toward clusters borders. On the other hand, at high temperature, also a partial or complete evaporation of water molecules was observed, thus the water network being insufficient to favour transformations.

On the other hand, in the ice representative model, neutral glycine molecule was barely present. At all three temperatures, transformation patterns were observed, the anionic form being the most favourable one. Note that at high temperature, similarly to the cluster systems, the evaporation of water molecules was observed, compromising the well delimitation of the QM-MM border cut off. Results at 450K for the ice representative model and for simulation times larger than $\sim$4 ps must be carefully analysed as some mixing atoms between QM and MM region occurred.

When comparing series of simulations at both 3 ps (Fig.4.10-A) and 10 ps (Figure 4.10-B), it appears that trajectories characterized as proton delocalization up to 3 ps, were later on identified as anion forms until the end of the 10 ps dynamics.

4.4.5 Conclusions for the DFTB and DFTB/MM Exploration

Thanks to the new reparametrization of V_{rep}^{CC}, V_{rep}^{NH} and V_{rep}^{OH} DFTB repulsive potential, great improvements have been obtained :

- Z-Gly local optimizations, showed that neutral glycine minima structures can be explored. The C-C bond breaking appearing when performing local optimizations using standard BIO parameters is now avoided with the new reparametrization.
- While in DFTB calculations with BIO parameters very few tautomerization mechanisms were observed (Figure 4.11), the four glycine conformers (neutral, zwitterionic, anion and proton delocalized) and their respective transformation mechanisms are detected with the new reparametrization.
- All trajectories for the larger ice model systems were completed up to the

10 ps thanks to the less computational cost of DFTB calculation with respect to DFT ones.

However, it must be mentioned that, in the DFTB/MM MD trajectories, a higher percentage of tautomerization transformations appeared compared to the DFT ones. In the ice representation model (see Figure 4.14), it is noticed that only in few trajectories the neutral glycine structures were maintained. This may be due to an overestimation of the attractive character of V_{rep}^{OH} and V_{rep}^{NH} DFTB potentials.

Furthermore, in the LDA ice model case at high temperature (450 K), almost a complete evaporation of water molecules was observed, compromising the accuracy in the results.

To conclude, although an improvement in the glycine study is observed with the new reparametrized BIO parameters, some few tunnings in the V_{rep}^{OH} and V_{rep}^{NH} DFTB potentials still must be done before broadening the glycine study. In particular, as a preliminary idea, an intermediate tunning between BIO standard parameters and the V_{rep}^{OH} and V_{rep}^{NH} reparametrization used in this chapter could be done. In future works, from the optimal reparametrized DFTB potentials, an extension to larger systems sizes as well as an improvement on the statistics sampling will be considered.

CHAPTER 5

Real Time - Time Dependent - DFTB (RT-TD-DFTB) Implementation in deMonNano and First Application

As introduced in Chapter 1, high energy collision processes can exist in the ISM or circumstellar medium, *e.g* cosmic rays (CRs) or solar winds and they may induce electronic excitation. So far in this thesis, only the electronic ground state stationary states were considered, and therefore the time independent Schrödinger equation was solved. Nevertheless, in the study of these ultrafast dynamic processes, the description of the excited electronic states and the dynamical time evolution of electronic systems requires to go beyond the electronic stationary ground state case and therefore to solve the time dependent Schrödinger equation.

The time dependent (TD)-DFT method appears as great tool to solve the time dependent Schrödinger problem, using either the Linear Response (LR)-TD-DFT or the Real Time (RT)-TD-DFT approaches.

In this chapter, a brief introduction on the tight binding extension, TD-DFTB, is presented. In particular, a new incode RT-TD-DFTB implementation in deMonNano is reported in combination with the DFTB/MM class 1 method already presented in Chapter 3. A benchmark study based on the comparison of absorption spectra is given. Finally, a preliminary study of the behaviour of N-Glycine molecule trapped in the LDA model ice under solar wind radiation is reported along with a similar study on free flying PAHs systems.

5.1 General Introduction

The resolution of the time dependent Schrödinger equation solved for **fixed nuclei** was introduced in Chapter 2 (recall Eq. 2.40 in Section 2.3.2). From the reported TD-DFT method based on the Runge and Gross theorems,[174] two main approaches are used in practice to solve the TD-Schrödinger problem: **(i)** The Real Time-TD-DFT (RT-TD-DFT) modelling explicitly the time evolution of an electronic system. **(ii)** The Linear Response-TD-DFT (LR-TD-DFT) mostly solved using Casida's formulation.[173] This approach models the particular case of a response under an electromagnetic oscillating field perturbation (weak perturbation). More details can be found in Section 2.3.2.

Several software packages have reported both LR- or/and RT-TD-DFT method implementation such as: Quantum Dissipative Dynamics (QDD),[308] VASP,[309,310] Quantum ESPRESSO[311] and deMon2k[148] code.

Time Dependent-DFTB method

As an extension to the TD-DFT method, an analogous methodology exists in the tight binding framework, known as the **time dependent-DFTB** method with both similar Real Time (**RT-TD-DFTB**) and Linear Response (**LR-TD-DFTB**) approaches. Similar to Casidas' work,[173] Niehaus *et al.*[202] provided a solution for the LR-TD-DFTB problem.

Several code implementations including LR-TD-DFTB and RT-TD-DFTB approaches can be found such as: DFTB+[185] , DFTBaby[190] and Amsterdam Molecular Suite (AMS)[312] for the Linear Response approximation and DFTB+[185] for the Real Time-TD-DFTB approach. In the case of the deMonNano code, at present only LR-TD-DFTB approach has been implemented.[313]

With the aim of properly modelling the electron dynamics between different excitation levels, a **new RT-TD-DFTB implementation in deMonNano in combination with the DFTB/MM class 1 method** was carried out and is reported in the following (see Section 5.2).

5.2 Real Time-Time Dependent-DFTB Method

Based on Runge and Gross[174] theorem and on the Khon-Sham TD-DFT formulation method, a first attempt for the explicit propagation of the electronic MOs is given by:

$$i\frac{\partial \phi_i(\mathbf{r},t)}{\partial t} = H[\rho(\mathbf{r},t)]\phi_i(\mathbf{r},t) \tag{5.1}$$

In practice, a reformulation of Eq. 5.1 can be done under the form of the **Liouville-Von-Neumann (LvN) electron equation of motion** for the electron density $\rho(\mathbf{r},t)$ propagation[314] :

$$i\frac{\partial \rho(\mathbf{r},t)}{\partial t} = [H[\rho(\mathbf{r},t)],\rho(\mathbf{r},t)] \tag{5.2}$$

Extending Eq. 5.2 to the TD-DFTB framework leads to replace the $H[\rho(\mathbf{r},t)]$ by the DFTB Kohn-Sham operator, with $\phi_i(\mathbf{r},t)$ being the molecular orbitals expanded as a **linear combination of the DFTB minimal atomic orbitals** (see Sec. 2.3.3):

$$\phi_i(\mathbf{r},t) = \sum_{\mu}^{M_{\text{basis}}} c_{\mu i}(t)\chi_\mu(\mathbf{r}) \tag{5.3}$$

From Eq. 5.3, it is observed that the time dependency is held in the $c_{\mu i}(t)$ molecular orbitals coefficients. In addition, due to the imaginary unit i present in Eq. 5.2, the complex character of the equation is transfered to the $c_{\mu i}(t)$ coefficients. Therefore, two time dependent terms, for the real and imaginary parts respectively constitute the total coefficients $c_{\mu i}(t) = c^R_{\mu i}(t) + ic^I_{\mu i}(t)$.

In the following, only developments regarding the TD-DFTB approach are reported although it must be noted that TD-DFT equations are closely related. They can be extrapolated straightforwardly only by replacing the DFTB Kohn-Sham operator by the DFT one.

Hand in hand with Eq. 5.3, another simplification in Eq. 5.2 is done by expressing the density $\rho(\mathbf{r},t)$ in terms of the **density matrix $\mathbf{P}_{\mu\nu}$** in the DFTB minimal atomic orbital basis:

$$\rho(\mathbf{r},t) = \sum_{\mu}^{M_{\text{basis}}} P_{\mu\nu}(t)\chi_{\mu}(\mathbf{r})\chi_{\nu}(\mathbf{r}) \tag{5.4}$$

with

$$P_{\mu\nu}(t) = n_i \sum_{i}^{occ} c^*_{\mu i}(t)c_{\nu i}(t) + c.c \tag{5.5}$$

Note from Eq. 5.5 that the complex character in $c_{\mu i}(t)$ coefficients is also found in $P_{\mu\nu}(t)$ matrix containing both the real and imaginary components.

Therewith, **the Liouville-Von-Neumann equation (LvN)**[314] (Eq. 5.2) is reformulated as:

$$iS\frac{\partial P(t)}{\partial t}S = H(t)P(t)S - SP(t)H(t) \tag{5.6}$$

with S being the atomic orbitals overlap matrix $S_{\mu\nu} = \langle\chi_{\mu}(\mathbf{r})|\chi_{\nu}(\mathbf{r})\rangle$ and $H(t)$ the DFTB Kohn-Sham operator in the atomic orbitals basis.

In deMonNano, we decided to work within an orthogonalized basis given by the **Löwdin symmetric orthogonalization (LSO)**[314] of the DFTB non-orthogonal atomic basis. Under the LSO approach, a new orthonormal basis ϕ' is build from a non-orthogonal basis set ϕ by $\phi' = S^{-\frac{1}{2}}\phi$, where $S^{-\frac{1}{2}}$ is the inverse root-square overlap matrix. In the following, *prime* will refer to matrices expressed in the othonormalized basis. The **orthonormalized $\mathbf{P'(t)}$** and $\mathbf{H'(t)}$ Hamiltonian matrices are given by:

$$\begin{aligned} P'(t) &= S^{1/2}P(t)S^{1/2} \\ H'(t) &= S^{-1/2}H(t)S^{-1/2} \\ S^{-1/2}SS^{-1/2} &= Id \end{aligned} \tag{5.7}$$

Note the changing in the exponentials' sign in $S^{1/2}$ and $S^{-1/2}$ respectively for $P'(t)$ and $H'(t)$ derived from $S^{-1/2}SS^{-1/2} = Id$. More details of the algebra derivation can be found in reference.[315]

From this orthonormalization, Eq.5.6 is simplified as:

$$i\frac{\partial P'(t)}{\partial t} = [H'(t), P'(t)] \tag{5.8}$$

Figure 5.1 illustrates the main steps in the mathematical derivation of the RT-TD-DFTB formulation. The explicit electron dynamic simulation consists on the propagation of the density matrix $P'(t)$. In the following section, the RT-TD-DFTB algorithm is detailed along with the specifications relevant to the deMonNano implementation.

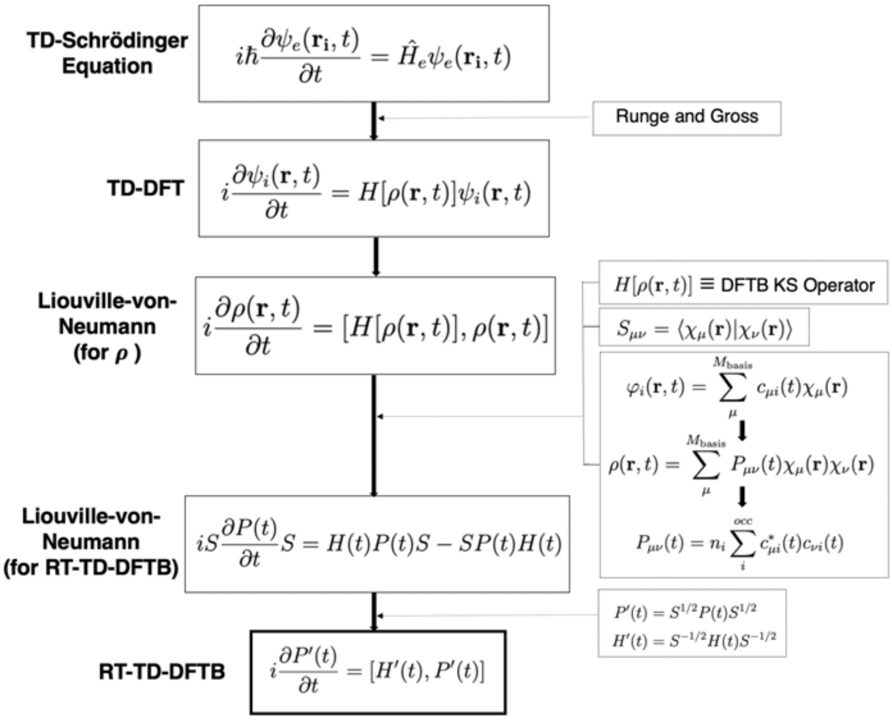


Figure 5.1: Scheme for the RT-TD-DFTB formulation from the starting TD-Schrödinger equation.

5.2.1 RT-TD-DFTB in deMonNano

In this section we present the development work performed in this thesis consisting in the implementation of the RT-TD-DFTB framework in addition to several time propagators. Firstly, we break down Eq. 5.8 in different steps for **the $\boldsymbol{P(t)}$ density matrix propagation** under the RT-TD-DFTB approach. Before detailing the main stages in the propagation procedure, few preliminary steps are carried out in order to build $P'(t)$ and $H'(t)$ matrices.

Note that in the RT-TD-DFTB approach, $\mathbf{H(t)}$ is the DFTB Hamiltonian operator corresponding either to $\mathbf{H^0_{\mu\nu}}$ (**for DFTB0**) or to $\mathbf{H^0_{\mu\nu}} + \mathbf{H^1_{\mu\nu}(q)}$ (**for DFTB2**) (recall Eq.2.47 and 2.55 in Chapter 2). In the following, we explain only the development of RT-TD-DFTB for the DFTB2 case, where $H(t)$ or $H'(t)$ are named respectively as $H^{\text{DFTB}}(t)$ and $H'^{\text{ DFTB}}(t)$.

In Figure 5.2 first right panel, the preliminary steps for the $P(t)$, H^{DFTB} and

$H'^{\,\mathrm{DFTB}}$ calculation are illustrated. From a $P(t)$ matrix, the q atomic charges are obtained. Note that although $P(t)$ is a complex matrix, only the real part of it is involved in the computation of q atomic charges. From the computation of q atomic charges, $\mathrm{H}^1_{\mu\nu}(q,t)$ matrix is calculated and added to $\mathrm{H}^0_{\mu\nu}(t)$ as in standard DFTB (recall Eq.2.56).

Straightforwardly, $H^{\mathrm{DFTB}}(t) = H^0(t)+H^1(q,t)$ is transformed into its orthonormalized basis using the root-square overlap matrix $S^{1/2}$ (see Eq. 5.7). The same is done with $P(t)$. Therewith, H'^{DFTB} and $P'(t)$ matrices are introduced in the LvN equation (Eq. 5.8) in order to propagate $P'(t)$ (see Figure 5.2).

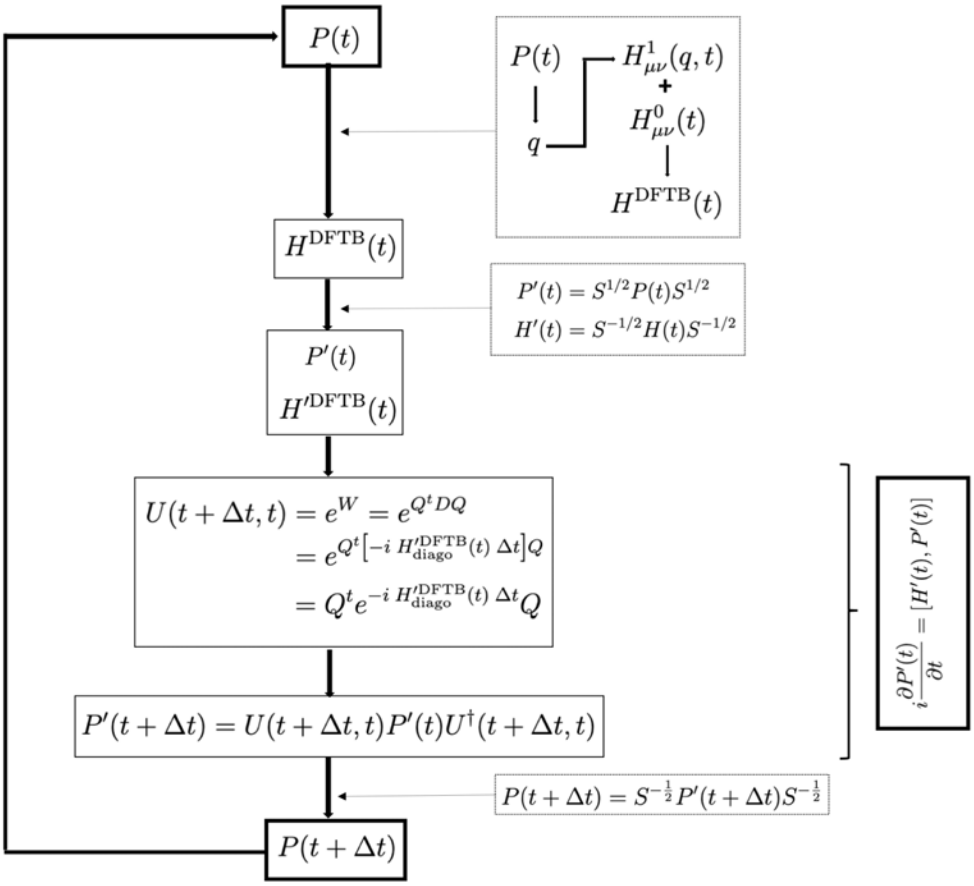


Figure 5.2: Illustration of RT-TD-DFTB algorithm development in deMonNano.

The numerical propagation of $P'(t)$ within the LvN equation can be done by the time discretization into small time steps. An **evolution operator $\mathbf{U(t+\Delta t,t)}$** is introduced to express the density matrix $P'(t+\Delta t)$ at time $t+\Delta t$ from the density matrix $P'(t)$ at time t :

$$P'(t+\Delta t,t) = U(t+\Delta t,t)P'(t)U^{\dagger}(t+\Delta t,t) \tag{5.9}$$

where the evolution operator $U(t+\Delta t, t)$, with t being the simulation time and Δt the simulation time step, is given by:

$$U(t+\Delta t, t) = T \exp\left\{-i \int_{t_i}^{t_i+\Delta t} H'(\tau) d\tau\right\} = e^W \tag{5.10}$$

In Eq. 5.10, T constitutes the time-ordering operator ensuring that the operators correlated to later times will always appear to the left of the ones correlated with earlier times. This is important when building the algebra of the different propagator forms (see Section 5.2.2):

$$T\left\{A(x)B(y)\right\} = \begin{cases} A(x)B(y) \text{ if } \tau_x > \tau_y, \\ \pm B(y)A(x) \text{ if } \tau_x < \tau_y. \end{cases} \tag{5.11}$$

As a first approach on the evaluation of $U(t+\Delta t, t)$ matrix, a general exponential matrix e^W is computed:

$$U(t+\Delta t, t) = e^W \tag{5.12}$$

In the following paragraphs, we focus on the computation of $U(t+\Delta t, t)$ and the propagation of $P'(t+\Delta t, t)$. The computation of W matrix is detailed in Section 5.2.2.

The calculation of $U(t+\Delta t, t)$ matrix can be a complicated task. In order to evaluate the exponential matrix e^W, the W matrix is first diagonalized by a real unitary matrix Q: $W = Q^t D Q$ with D the diagonal matrix. In this manner, e^W is build from the exponential of the diagonal elements of matrix D:

$$U(t+\Delta t, t) = e^W = e^{Q^t D Q} = Q^t e^D Q \tag{5.13}$$

The last two stages in the propagation of $P(t)$ are given by the application of $U(t+\Delta t, t)$ to $P'(t)$ (Eq. 5.9). Note that in 5.9, $P'(t)$ and its propagated matrix $P'(t+\Delta t, t)$ are complex matrices and therefore the complex transpose $U^\dagger(t+\Delta t, t)$ is applied during propagation. On the other hand, Q is constituted only by real part matrix elements, thus the diagonalization in Eq.5.13 involves Q^t is instead of $Q^\dagger$.

With the new $P'(t+\Delta t, t)$ propagated density matrix, a transformation into atomic basis is done applying the inverse root-square overlap matrix: $P(t+\Delta t, t) = S^{-1/2} P'(t+\Delta t, t) S^{-1/2}$. Therewith, $P(t)$ is updated by $P(t) \leftarrow P(t+\Delta t, t)$ and the similar procedure is followed until the end of the simulation time.

In a simulation performed with this new RT-TD-DFTB implementation, the first step in the calculation consists of choosing an initial density matrix $P(t)$ to start the loop in the RT-TD-DFTB algorithm (Figure 5.2).

Note that the presented RT-TD-DFTB algorithm is common for all DFTB approaches, yet for the DFTB0 case all steps concerning atomic charges are avoided. Finally, this new implementation allows the use of DFTB (DFTB0, DFTB2, DFTB3) approaches as well as **the DFTB/MM class 1 implementation** reported in Chapter 3. All modifications in the DFTB0/DFTB2 Hamiltonian operator for the additive coupling and electrostatic embedding are therefore included in H^{DFTB} and $H'^{\,\mathrm{DFTB}}$ (recall Chapter 3 Sec. 3.2).

5.2.2 Time Propagators

From the interest to improve the RT-TD-DFTB algorithm efficiency, sophisticated extensions exist while computing e^{W}. Several reviews of the different propagators have been done in ref.[316,317]

In this work, firstly **Magnus Propagator** and later the **Predictor-Corrector scheme** were implemented in deMonNano. Note that the RT-TD-DFTB algorithm introduced in previous section (Figure 5.2) will be maintained, however few changes in the computation of $U(t+\Delta t, t)$ matrix (Figure 5.2-forth left panel) are done.

Magnus Propagator (MP)

As suggested in Álvarez-Ibarra *et al.*,[318] the **Magnus operator**[319] is considered to be one of the most robust propagators. The Magnus propagator expresses the time-unordered exponential by a series of commutator integrals (Ω_i):[252]

$$\begin{aligned}
&T\exp\left\{-i\int_{t_i}^{t_i+\Delta t} H'(\tau)d\tau\right\} = e^{W} = e^{\sum_{i=1}^{\infty}\Omega_i} \\
&\Omega_1(t+\Delta t, t) = -i\int_{t}^{t+\Delta t} H'(\tau)d\tau \\
&\Omega_2(t+\Delta t, t) = -i\int_{t}^{t+\Delta t} d\tau_1 \int_{t}^{\tau_1} d\tau_2\,[H'(\tau_1), H'(\tau_2)] \\
&\ldots
\end{aligned} \tag{5.14}$$

In this work, the exponential series was restricted to the second-order ($W = \Omega_1$) also known as the **Second-Order Magnus propagator (SOMP)**. In this approximation only one H'^{DFTB} at a time $t + \Delta t/2$ is computed:

$$W = \Omega_1(t+\Delta t, t) \simeq -iH'^{\text{DFTB}}\left(t+\frac{\Delta t}{2}\right)\cdot \Delta t \tag{5.15}$$

More sophisticated extensions to the Magnus operators exist, *e.g* Furth-order Magnus Propagator (FOMP). In this approximation, two commutator integrals Ω_1 and Ω_2 are evaluated at different time intervals. The FOMP extension is out of the scope of this thesis, more details can be found in reference.[318]

With the new formulation of W matrix, the evolution operator $U(t + \Delta t, t)$ is given by:

$$\begin{aligned} U(t+\Delta t, t) =& e^{W} = e^{-iH'^{\text{DFTB}}\left(t+\frac{\Delta t}{2}\right)\cdot\Delta t} \\ =& e^{Q^t D Q} = Q^t e^{-iH'^{\text{DFTB}}_{digonal}\left(t+\frac{\Delta t}{2}\right)\cdot\Delta t} Q \end{aligned} \tag{5.16}$$

Note that $W = -iH'^{\text{DFTB}}\left(t+\frac{\Delta t}{2}\right)\cdot\Delta t$ matrix is diagonalized as explained in Eq. 5.13), where now D is the diagonal matrix with $H'^{\text{DFTB}}_{digonal}\left(t+\frac{\Delta t}{2}\right)$ matrix elements. In the RT-TD-DFTB algorithm, the step for the computation of $U(t + \Delta t, t)$ (see Figure 5.2-forth left panel) is now replaced by Eq. 5.16.

The propagation with the Second-order Magnus Propagator scheme needs the computation of $H'^{\text{DFTB}}\left(t+\frac{\Delta t}{2}\right)$ which is unknown. There exists two different methods to compute the latter matrix:[252] 1) the iterative algorithm[252] and 2) two-step Predictor-Corrector scheme.[318,320] Due to the time-consuming procedure in the first case, in this thesis, only the latter **Predictor-Corrector scheme** was implemented in deMonNano.

Predictor-Corrector Scheme

In Van Voorhis work,[320] several steps were detailed in order to properly compute the W matrix from the Second-order Magnus Propagator formula. A first **predictor** stage is performed by estimating the *future* H'^{DFTB} matrices solely from the past ones. However, at a very long simulation time this approach will drag certain errors. For this reason, a second stage **corrector** is carried out finding a balance during the *predictor-corrector* calculations using the same Magnus expansion.

Figure 5.3, illustrates the steps followed in the W matrix computation and in the $P(t)$ propagation by means of the Predictor-Corrector scheme[320] :

1. **Predictor**: from the two stored H^{DFTB} matrices from previous $t_{n-2} + \Delta t/2$ and $t_{n-1} + \Delta t/2$ time steps, a new H^{DFTB} at $t + \Delta t/4$ matrix is extrapolated

by means of:

$$H^{\mathrm{DFTB}}\left(t+\frac{\Delta t}{4}\right)=-\frac{3}{4}H^{\mathrm{DFTB}}\left(t_{n-2}+\frac{\Delta t}{2}\right)+\frac{7}{4}H^{\mathrm{DFTB}}\left(t_{n-1}+\frac{\Delta t}{2}\right) \tag{5.17}$$

2. **Predictor**: with $H^{\mathrm{DFTB}}\left(t+\frac{\Delta t}{4}\right)$, the diagonalized matrix to evaluate e^{W} is build and thus the evolution operator $U\left(t+\frac{\Delta t}{2},t\right)$, which is used to predict $P\left(t+\frac{\Delta t}{2}\right)$.

3. **Corrector**: $P\left(t+\frac{\Delta t}{2}\right)$ is used to calculate $H^{\mathrm{DFTB}}\left(t+\frac{\Delta t}{2}\right)$

4. **Propagator**: the propagation is done by calculating the new e^{W} with $H^{\mathrm{DFTB}}\left(t+\frac{\Delta t}{2}\right)$ and thus the $U\left(t+\Delta t,t\right)$ (Eq. 5.15 and 5.16 in Magnus Propagator). With Eq. 5.9 the propagated matrix $P\left(t+\Delta t,t\right)$ is calculated.

5. **Update** There is an extra step of matrices updating where:

$$\begin{aligned} P(t) &\leftarrow P\left(t+\Delta t,t\right)\\ H^{\mathrm{DFTB}}\left(t_{n-2}+\frac{\Delta t}{2}\right) &\leftarrow H^{\mathrm{DFTB}}\left(t_{n-1}+\frac{\Delta t}{2}\right)\\ H^{\mathrm{DFTB}}\left(t_{n-1}+\frac{\Delta t}{2}\right) &\leftarrow H^{\mathrm{DFTB}}\left(t+\frac{\Delta t}{2}\right) \end{aligned} \tag{5.18}$$

The Predictor-Corrector scheme implementation in deMonNano was chosen as the main propagation scheme in the following. A great advantage of this propagator scheme is that it optimizes the Δt time step (reduced computational time), while preserving the results accuracy.

5.3 Benchmark Implementation and Absorption Spectra

In order to benchmark the implementation of RT-TD-DFTB method in deMonNano, a comparison between several absorption spectra was performed. On one hand, a set of absorption spectra for different molecules where calculated by the Linear Response-TD-DFTB method in deMonNano. On the other hand, the absorption spectra for the same molecules were obtained from RT-TD-DFTB calculations with a **δ-function pulse** external perturbation. The latter function also known as **δ-kick function**, is characterized by its weak and linear perturbation mimicking the one modelled in LR-TD-DTFB calculations.

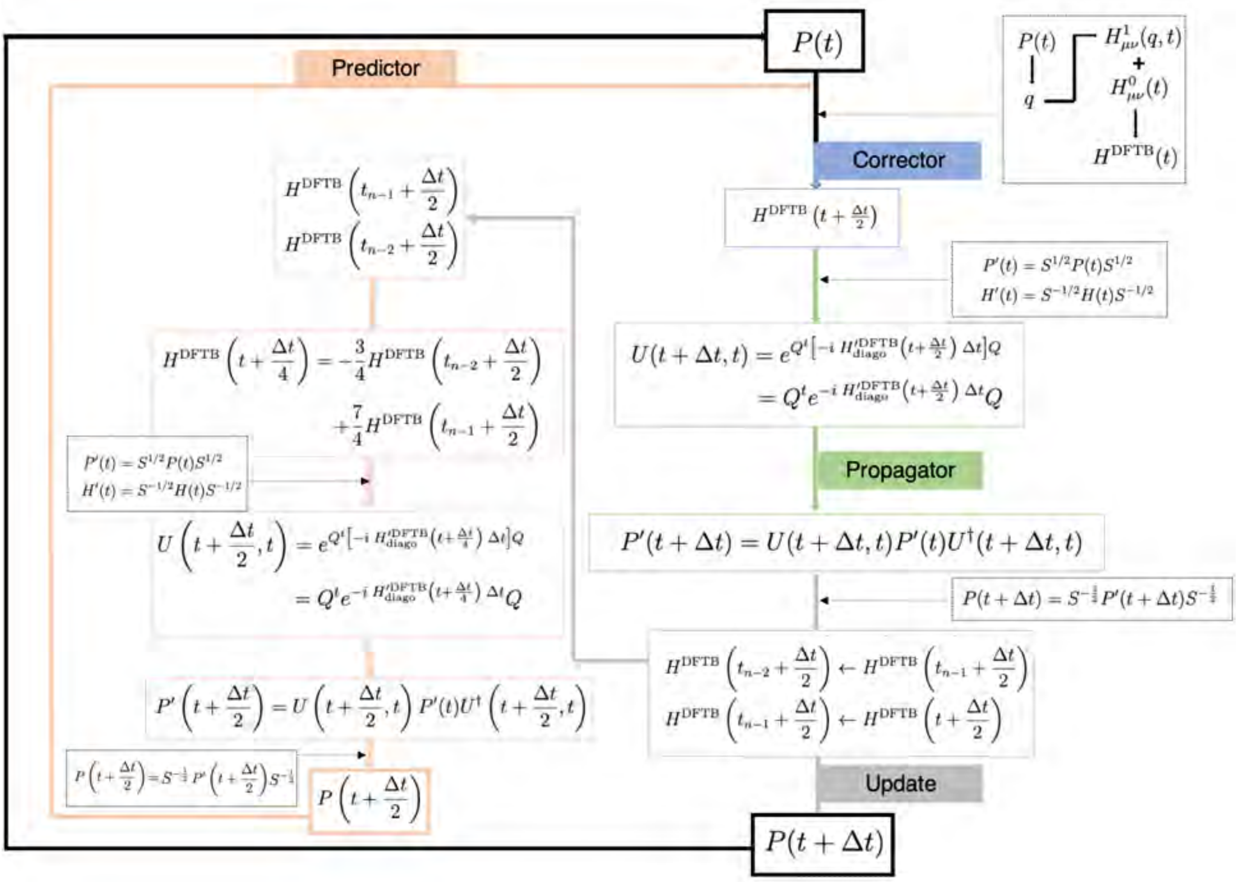


Figure 5.3: Illustration of the RT-TD-DFTB algorithm development combined with the Predictor-Corrector propagator scheme.

The calculation of absorption spectra are of great interest in Astrophysics. Many works have been reported on the computation of absorption spectra in the near-IR, visible and near-UV spectral range.[78,88,103,321] For these reasons, the calculation of absorption spectra obtained by RT-TD-DFTB methods with a delta-function pulse, are not only a tool for the benchmark implementation but also an instrument for future studies.

In this section a brief description of the δ-kick pulse and its new implementation in deMonNano as part of this thesis work is presented. In addition, the absorption spectra comparison is reported.

5.3.1 δ-Function Pulse of a Dipole Field

In this approach, the electronic spectrum is obtained from the response of a system submitted to excitations covering all possible frequencies. Such multi-wavelenght excitations is represented by:[322]

$$E(t) = \frac{1}{\pi}\int_0^{+\infty} \varepsilon_0(\cos\omega t)\mathbf{k^0} d\omega \tag{5.19}$$

where $\mathbf{k^0}$ is the unit vector giving the polarization direction ($\mathbf{k^0} \in x, y, z$) of the field and ε_0 is the unitless field strength. A mathematical equivalence exists applying a delta Dirac function pulse perturbation at a time zero, also know as **δ-kick pulse**:[322,323]

$$\mathbf{E_{kick}}(t) = \varepsilon_0\mathbf{k^0}\delta(t) = \frac{1}{\pi}\int_0^{+\infty} \varepsilon_0(\cos\omega t)\mathbf{k^0} d\omega \tag{5.20}$$

$\delta(t)$ is a Dirac type function characterized by its unique impulse for a very short period of time. From the mathematical side, the one-dimension δ-Dirac function $\delta(x)$ can be obtained from a δ_c function:[324,325]

$$\delta_c(x) = \begin{cases} 0 \text{ for } |x| > c \\ \frac{1}{2c} \text{ for } |x| < c. \end{cases} \tag{5.21}$$

with $\delta_c(x)$

$$\int_{-\infty}^{\infty} \delta_c(x)dx = 1 \tag{5.22}$$

Defining $\delta(x)$ as:

$$\delta(x) = \lim_{c\to 0} \delta_c(x) \tag{5.23}$$

Among all the properties of $\delta(x)$ one can mention:

$$\delta(x) = 0 \quad \text{if } x \neq 0$$

$$\int_{-\infty}^{\infty} \delta(x)dx = \int_{0^-}^{0^+} \delta(x)dx = 1 \tag{5.24}$$

The computation of the absorption spectra are obtained by a single real time propagation after a δ-kick perturbation. Based on Yabana's work,[322] the electronic wave function (Eq. 5.1) **just after the δ-kick perturbation** of the MOs is given by:[326]

$$\psi_i(t = 0^+) = \exp\left(i\, \varepsilon_0\, \mathbf{k^0} \cdot \mathbf{r}\right) \psi_i(t = 0^-) \tag{5.25}$$

In the following, Eq. 5.25 is going to be extended to the RT-TD-DFTB method. In addition, specific details on the δ-kick implementation in deMonNano are reported.

5.3.2 δ-Kick function Pulse in deMonNano

In this section, the above presentation of the δ-kick pulse is extended to RT-TD-DFTB approach replacing the evolution of electronic wave function by the evolution of the density matrix $P(t)$. Similar simplifications used in previous sections (recall equations 5.3 and 5.5) yield to a reformulation of Eq. 5.25 in terms of the density matrix $\mathbf{P}_{\mu\nu}$ in the DFTB minimal atomic orbital basis. More details on the equation's derivation can be found in references.[326, 327]

In practice, the implementation of Eq. 5.25 within the DFTB framework, is done by the computation of a matrix $\tilde{\mathbf{K}}$:

$$\tilde{K} = -i\frac{1}{2}\varepsilon_0 \mathbf{k_0} S_{\mu\nu} \left(\boldsymbol{R_a} + \boldsymbol{R_b}\right) \tag{5.26}$$

$$P'(t = 0^+) = e^{\tilde{K}'} P'(t = 0) \left(e^{\tilde{K}'}\right)^\dagger \tag{5.27}$$

$\mathbf{R}_a$ and $\mathbf{R}_b$ are the coordinates of a and b atoms. $\tilde{\mathbf{K}}'$ expression in orthonormal basis computed making use of the previously introduced algebra expressions in Eq. 5.7. The new exponential matrix is calculated with the diagonalized $\tilde{K}$ matrix:

$$e^{\tilde{K}} = e^{Q^T \tilde{K}_{diago} Q} = Q^T e^{\tilde{K}_{diago}} Q \tag{5.28}$$

Figure 5.4 illustrates the preliminary steps following in the perturbation of the molecular system. From this initial perturbation of the density matrix $P(t = 0^+)$, the evolution of the molecular system after the δ-kick perturbation will be done by means of RT-TD-DFTB methods monitoring the density matrix $P(t + \Delta t)$ as reported in Section 5.2.1. Note that the RT-TD-DFTB algorithm in Figure 5.2 is maintained, however the initialization of $P(t)$ is given by the replacement of $P(t) \leftarrow P(t = 0^+)$. Finally, the absorption spectra is obtained from the Fourier Transformation of the dipole moment, where $\boldsymbol{\mu(t)}$ is obtained from the propagated $P(t)$ density matrix using the formula in reference:[326]

$$\alpha(\omega)\mathbf{k^0} = \frac{2m_e}{e\hbar\pi\varepsilon_0}\omega \int_0^T sin(\omega t) g(t) \left[\boldsymbol{\mu}(0) - \boldsymbol{\mu}(t)\right] dt \tag{5.29}$$

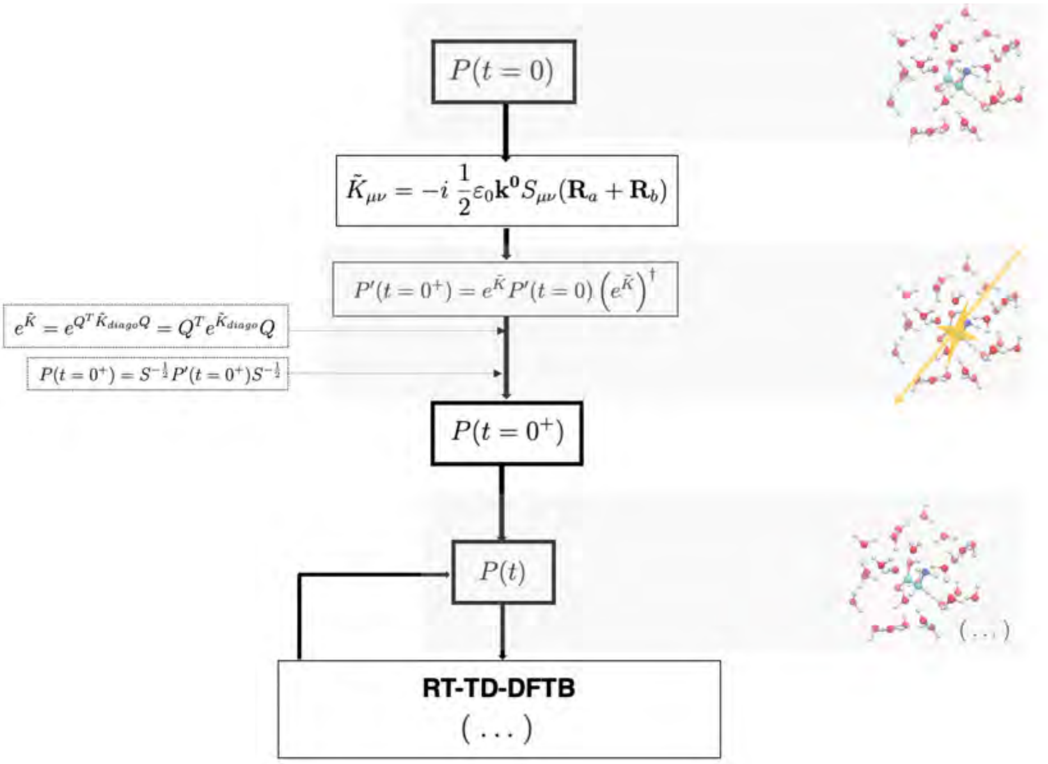


Figure 5.4: Schema for the δ-kick pulse external perturbation within the RT-TD-DFTB algorithm implemented in deMonNano. Right panels illustrate the perturbation steps in an example molecular system.

The dipole strength tensor $\alpha(\omega)$ is connected to the folded oscillator strength $f_{I\kappa}$ via:

$$\mathbf{k^0}_\kappa \alpha(\omega) \mathbf{k^0}_\kappa = \sum_I f_{I\kappa} \tilde{g}(\omega - \omega_I) \tag{5.30}$$

Here κ refers to any Cartesian coordinates x, y, z. In Eq. 5.29, T is the total simulation time and $g(t)$ is an envelope function removing the effects of finite simulations time. In Eq. 5.30, $\tilde{g}(\omega)$ is the normalized Fourier Transformation.[326] In order to compare both LR- and RT-TD-DFTB absorption spectra, the same envelope function $\tilde{g}(\omega)$ has been applied to the LR- stick spectra.

5.3.3 Benchmark Absorption Spectra Comparison

Three simple molecules in gas phase, acetylene (C_2H_2), water molecule (H_2O) and *trans-trans*-glycine (NH_2CH_2COOH), were chosen as benchmark. Each molecular system was perturbed by a δ-kick pulse function in the three k_x^0, k_y^0 and k_z^0 polarization directions with $\varepsilon_0 = 0.003$. The total RT-TD-DFTB simulation were performed using the Second-Order Magnus Propagator with a time step of 0.5 as and a total simulation time of 50 fs. MAT DFTB parameters were chosen in these calculations.

LR-TD-DFTB simulations were performed for the same three molecular systems. Figure 5.5, illustrates the **three absorption spectra** computed in both LR-TD-DFTB (red line) and RT-TD-DFTB simulations (blue line). It must be recalled that the geometrical configurations were chosen arbitrarily. They do not exemplify any real system. Any analysis of the properties and characteristic of this geometries is out of the scope of this work.

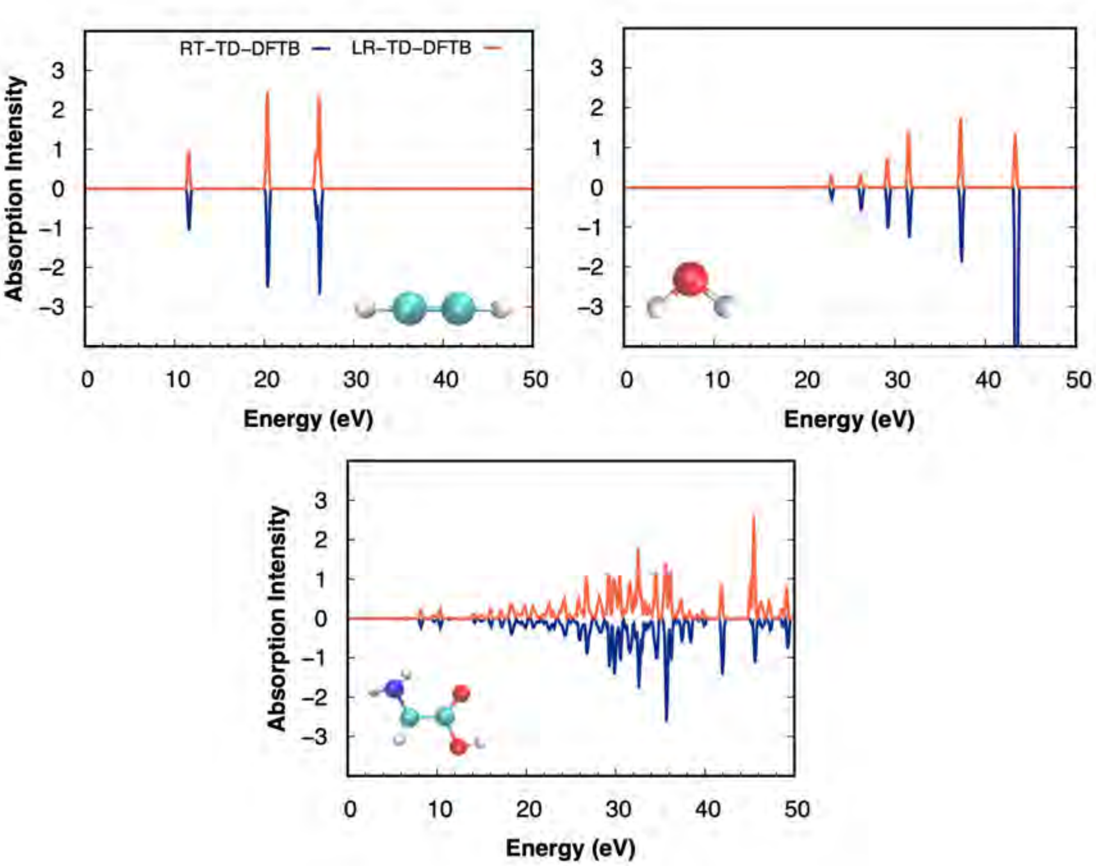


Figure 5.5: Absorption spectra obtained by both LR-TD-DFTB (red line) and RT-TD-DFTB (blue line) approaches.

In Figure 5.5-top left panel, the **absorption spectra for the acetylene molecule** is shown. Both LR- and RT spectra appear to be mostly similar with the three main absorption peaks at 11.6, 20.4 and 26.2 eV and with similar intensities at each three peaks.

In Figure 5.5-top right panel, the **absorption spectra for a single water molecule** at both LR- and RT-TD-DFTB methods are depicted. Similarly to previous results, both LR- and RT- methods give similar absorption spectra with six main peaks at 22.9, 26.2, 29.2, 31.6, 37.2 and 43.3 eV. Note that the latter peak at 43.3 eV was found to be slightly blue shifted in the RT- spectra (0.2 % with respect to LR- results). Furthermore, for the first five peaks, similar intensities were found between LR- and RT- spectra.

Finally, in Figure 5.5-bottom central panel, the **absorption spectra for the *trans-trans*-neutral glycine molecule** is illustrated. In this latter, many absorption peaks appear. Similar patterns were found in both spectra calculated by LR- and RT- approaches. The more predominant peaks were found at 26.8, 30.0, 32.5, 34.5, 37.4, 42.0 and 45.5 eV. Only peaks between 42.0 and 50.0 eV energy values were slightly blue shifted (∼0.4% with respect the LR- spectra). It is also observed that for peaks lower than 35 eV, the intensities were found similar in both LR- and RT- spectra.

Note that only results for a range between 0 to 50 eV energy values are reported. This correspond to the range were the accuracy of LR-TD-DFTB results are believed to be assured. In the case of water and glycine absorption spectra, some variation in the RT- absorption intensities were found with respect to the LR- ones. These differences, along with the small shifts on the main absorption peaks were noted only for high absorption energies (above 35 eV). The difficulties in computing high energy absorption modes using LR-, may be explained by the large number of high single particle energy excitations that are needed to be included to properly compute Casida's equations. On the other hand, also RT-TD-DFTB spectra might present some errors when extracting the high energy modes due to the chosen time step. It is possible that a smaller time step would be needed in order to have a proper description of the high frequency excitations.

As a last remark, thanks to the RT-TD-DFTB implementation, the decomposition of the RT- absorption spectra in its three x, y, z components can be extracted. Figure 5.6-upper left panel illustrates the total RT-TD-DFTB absorption spectra for acetylene molecule compared to its three component lines. It is shown that only the spectra for the x-direction component appears during the simulation (green line). This resonance to the perturbation only for the x-component of the acetylene molecule can be explained due to the linearity of the molecule. Furthermore, for the case of the water molecule (see Figure 5.6-upper right panel), it is observed that only y- and z-components are the active ones (yellow and purple line). This is explained due to the planarity (yz) of water molecule. Finally, in the case of glycine absorption spectra (see Figure 5.6-bottom panel), all three x-, y- (predominantly) and z-components are activated.

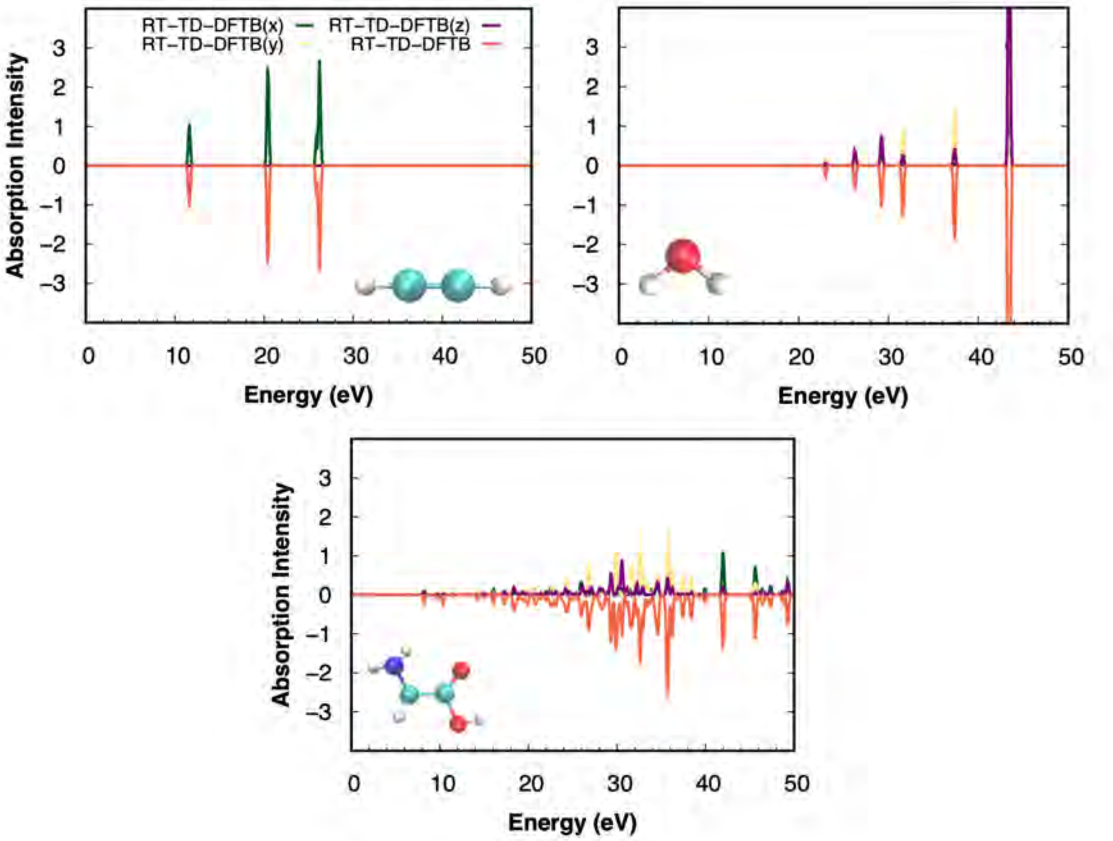


Figure 5.6: Absorption spectra obtained by RT-TD-DFTB (blue line) approach. RT-TD-DFTB spectra curve is splitted into its three $k^0(x)$ (gree line), $k^0(y)$ (yellow line) and $k^0(z)$ (purple line) components.

From the previously discussed results, it can be concluded that the RT-TD-DFTB method implemented in deMonNano under a linear and weak perturbation (δ-kick pulse) reproduced successfully the absorption spectra obtained by LR-TD-DFTB approach. In addition, the difficulties found in long time scale simulations for LR-TD-DFTB approaches are now reduced by the RT-TD-DFTB method. This latter, is found to be accurate at both short and long simulation's time scales. In the following, two first case studies are reported were the δ-kick pulse is replaced by a non-linear and high energy collision perturbations.

5.4 Ion Collision Induced Electron Dynamics in Molecular Systems

In this section, we report first simulations of ion collision induced electron dynamics with the new RT-TD-DFTB development. Firstly, the **ion-N-Glycine col-**

lision has been studied mimicking the solar wind collision processes. Subsequently, **ion-PAH collision** has been performed mimicking the low energy cosmic ray collision appearing in the protostellar regions.

In each study, the colliding ion corresponds to a charged proton particle (H^+) with a constant kinetic energy. The initial kinetic energies were chosen depending on the targeted astrophysical regions, respectively 10 keV for the ion-N-Gly collision or 100 keV for the ion-PAH collision. Note that, the chemical reactivity (*e.g* chemisorption, fragmentation, etc) may take place in a second step is not investigated here.

5.4.1 Ion-N-Glycine Collision

Computational Details

From the reported results in Chapter 4, four different N-Glycine molecular systems were chosen for this study. On one hand, the two most stable *cc*- and *tt*-N-Gly molecules in gas phase were chosen and treated by the DFTB2 approach using BIO parameters. On the other hand, the two stable *cc*-N-Gly-LDA and *tt*-N-Gly-LDA molecular systems both in their neutral forms were chosen. *cc*- and *tt*-N-Gly systems and several molecules from the first solvation shells (containing 61-62 water molecules), were treated at the DFTB2 level using BIO DFTB parameters, while the remaining 282-283 water molecules were treated at the MM level described by the AMBER-FF99SB force field potential (TIP3P for water molecules) (see Chapter 4 Section 4.4.1).

The electron propagation was performed by the RT-TD-DFTB/MM class 1 FF method using the Second-Order Magnus Propagator coupled to a Predictor-Corrector scheme with a time step of 1 as. The total simulation time consisted of 2 fs or 6 fs respectively for the gas phase and the glycine embedded in ice model. These systems were submitted to a proton collision at 10keV (energy deposition). Three different proton trajectories and impact parameters were set: **1)** one perpendicular to the N-C-C plane of glycine colliding on the middle of the C-C bond **(CC)**, **2)** one perpendicular to the N-C-C plane of glycine impacting at the nitrogen atom **(N)** and **3)** the latter one perpendicular to the C-C-O plane of glycine colliding at the oxygen atom **(OH)** (where the oxygen atom is the one attached to H in the carboxilic group). The proton at its initial position was placed either 7Å away for the isolated glycine or 20Å away from the N-Gly-LDA sytems.

Figure 5.7-upper panels illustrate the set of impact parameters respectively for the *cc*-N-Gly and *cc*-N-Gly-LDA. Similarly, Figure 5.8-upper panels depict the *tt*-N-Gly and *tt*-N-Gly-LDA impact parameters. These figures capture the real time **charge evolution (CE)** consisting the fluctuation of the charge with respect to the initial charge values. Note that, for the ice modelled systems (Figure 5.8 and 5.7 right panels) the CE for the complete water environment was represented (blue line) in addition to the glycine molecule charge evolution.

Ultrafast-charge evolution in Ion-*cc*-N-Glycine collision

From this preliminary simulations, Figure 5.7-left panel depicts the charges' time evolution for ***cc*-N-Gly isolated molecule**. It can be observed that in trajectories passing through the central C-C bond (CC), the positively charged proton H^+ induces electron displacement towards the C_O carbon (attached to the carboxilic group) and to the C_N carbon atom (attached to nitrogen). The latter being slightly higher. In addition, most of the proton-attracted electron density is taken from the nitrogen atom (N), from the oxygen atoms constituting the carboxilic group (O_H and O) and from the hydrogen atoms attached to C_N atom (H_{CH}).

In the cases were the impact occurred at the nitrogen atom (N), first electron displacement is induced to the C_N atom, where the electron densities are taken from the bonded carbon (C_O) and H_{CH} hydrogen atoms attached to C_N. Thereafter, a secondary displacement (∼0.05 fs after) occurred, where the previously located electron density in C_N atom migrates to the nitrogen atom. The electron density at ∼0.5 fs is highly located at N atom.

Finally, for trajectories passing through the O_H oxygen attached to the hydrogen atom (OH), the induced electron displacement was mostly located at this same atoms. The electron redistribution is taken mainly from the H_{OH} hydrogen atom and from the C_O carbon atom. Charge motion from the other oxygen atom (O) also appeared.

Note that in all three dynamics, there is a delay of ∼0.2 fs where any electron displacement is observed. This indicates that collision have not yet occurred, and the atomic charges are similar to the initial ones.

Similar electron displacement effects are observed in the case of the ***cc*-N-Gly-LDA molecular system**, however the impact appears latter in time (∼1.3 fs) as the proton passes through the ice shells before reaching the glycine molecule (Figure 5.7-middle and right panels). In Figure 5.7-middle panel, the induced electron

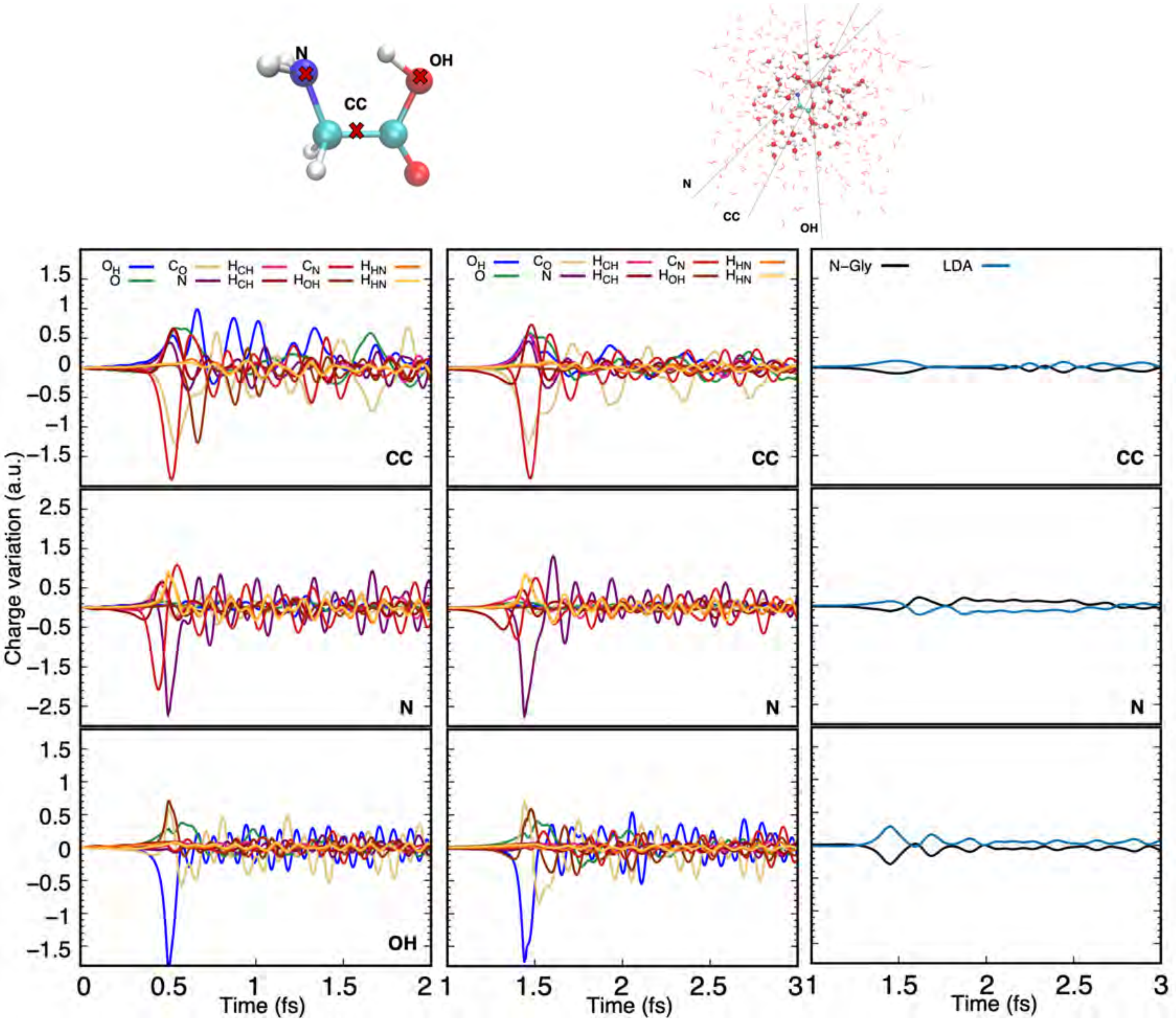


Figure 5.7: Ultrafast charge evolution for *cc*-N-Gly (left panel) and *cc*-N-Gly-LDA (middle and right panels) after collision with a 10 keV H^+ impacting the center of a central C-C glycine bond (CC) or the nitrogen atom perpendicular to the N-C-C plane (N) or the oxygen atoms (O-H) perpendicular to the C-C-O plane (OH). On the left panel, full 2 fs simulations are represented, however 3 fs out of the 6 fs simulation time are represented on the middle and right panels.

displacement is represented only for each individual glycine's atoms. On the other hand, Figure 5.7-right panel monitors the electron displacement for the total molecular charge of glycine (black line) in addition to the charge evolution of the LDA environement (blue line). It is interesting to observe that, in the collision process, electron displacement mostly occurs inside the glycine molecule (intramolecular displacement). However we note that, small intermolecular electron displacement also exist between the glycine molecule and the water environment.

Ultrafast-charge evolution in Ion-*tt*-N-Glycine collision

The charge time evolution for simulation runs in *tt*-N-Gly and *tt*-N-Gly-LDA molecular systems are reported in Figure 5.8. In all trajectories of the isolated system, the electron displacements are found very similar to the ones reported for *cc*-N-Gly molecule. However, there are two major differences: **(i)** when the proton impacts the middle C-C bond (CC), it is observed that the electron density taken from the O_H atoms is significantly lower that in the case of the *cc* isomer; **(ii)** for trajectories where H^+ impacts the nitrogen atom (N), contrary to the *cc*-N-Gly trajectories, almost all electron density is located at the N atom. The induced charge evolution observed at the carbon C_N atom for the *cc* isomer, is almost absent in the case of the *tt*-N-Gly. A charge oscillation between the glycine molecule and the ice environment is observed from Figure 5.8-right panel, however, as already noticed in the case of *cc*-N-Gly trajectories, this intermolecular charge motion is smaller than the intramolecular one.

It must be mentioned that in all trajectories, once the projectile is gone, the electron accumulation becomes energetically unfavorable and, due to electron-electron repulsion, the electronic density flows away from the impact point (or from the highly located electron density in an atom) and the corresponding atoms suffers and undergo electron loss. This is the starting point of a period of charge oscillation in the glycine systems. Similar attraction of the electrons through the impact point followed by repulsion has been reported in several different contexts, *e.g* irradiated biomolecules, and is known as **ebb-and-flow effect**.[328, 329]

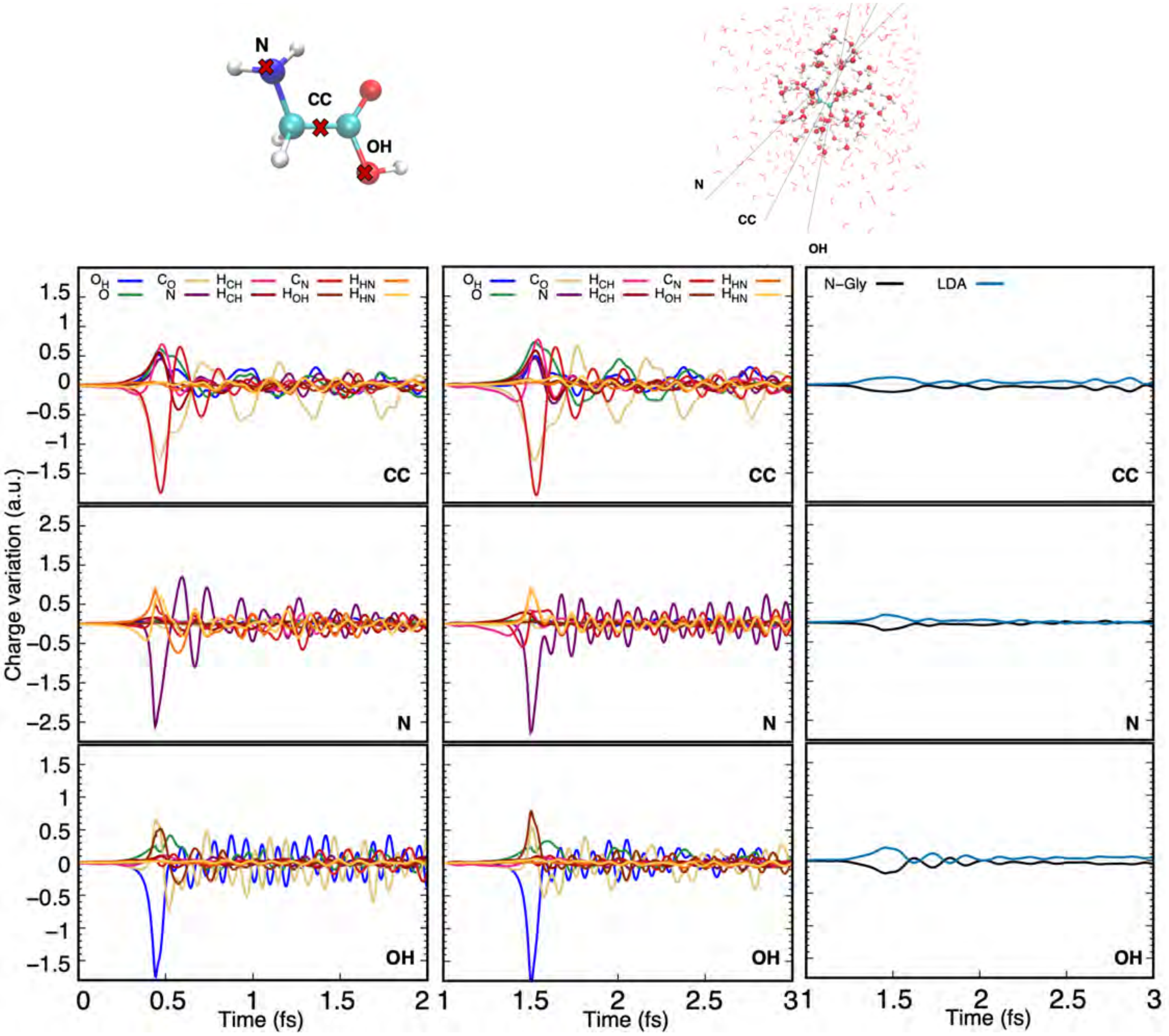


Figure 5.8: Ultrafast charge evolution for *tt*-N-Gly (left panel) and *tt*-N-Gly-LDA (middle and right panels) after collision with a 10 keV H^+ impacting the center of a central C-C glycine bond (CC) or the nitrogen atom perpendicular to the N-C-C plane (N) or the oxygen atoms (OH-) perpendicular to the C-C-O plane (OH). On the left panel, full 2 fs simulations are represented, however 3 fs out of the 6 fs simulation time are represented on the middle and right panels.

5.4.2 Ion-PAH Collision

Computational Details

Anthracene and octacene molecules in gas phase were previously optimized using the DFTB2 approach with MAT parameters and a Fermi distribution temperature of 100K. The class IV/charge model 3 (CM3) charges were used with a parametrization of $D_{\text{C-H}} = 0.1$.[306]

The electron propagation was done by the RT-TD-DFTB method using the Second-order Magnus Propagator coupled to a Predictor-Corrector scheme with a time step of 1as and a total simulation time of 1fs. Both molecular systems where submitted to a proton (H^+) collision at 100 keV. The proton trajectory is perpendicular to the PAH plane and the impact point is located either: **1)** in the middle of a central aromatic ring **(C)**, **2)** on the center of a C-C bond in the middle of the PAHs **(CB)** or **3)** on the center of a C-C bond at the border of the PAHs **(B)**. The initial position of proton was 5Å away from the PAHs impact point. Figure 5.9-upper panels illustrates the set of impact parameters respectively for the antracene and octacene molecules.

Ultrafast-charge evolution in Ion-PAH collision

A set of subfragments were defined for the study of the time charges evolution each of them constituted by a pair of C-C and the respective attached hydrogen atoms (see Figure 5.9-upper panels). As preliminary runs, Figure 5.9-lower panels report the time evolution charges for the subfragments. It can be observed that the positively charged proton H^+ induces electron displacements inside the PAH toward the impact point. Larger charge accumulations are obtained for trajectories colliding a bond than for trajectories passing through an aromatic cycle center (C). Most of the proton-attracted electron density is taken from the nearby fragments which acquire simultaneously positive charges.

For the central bond collision (CB), electrons can be attracted from both sides of the bond leading to a larger charge accumulation than for a border bond collision.

Similarly than in previous ion-N-glycine collisions, the ebb-and-flow effect and the period of charge oscillation in the PAH are observed. Interestingly, the size of the PAH seems to have a minor effect on the electron dynamics regarding the intensity of electronic density accumulation, as well as the initial charge dynamics. This can be explained because, at the short timescale, the charge dynamics is very local. This is less true at longer timescales and it can be noticed that high frequency charge oscillations remain clearly visible in the anthracene case, whereas in octacene the charge oscillation becomes more spread over the full PAH, reducing the oscillations intensities.

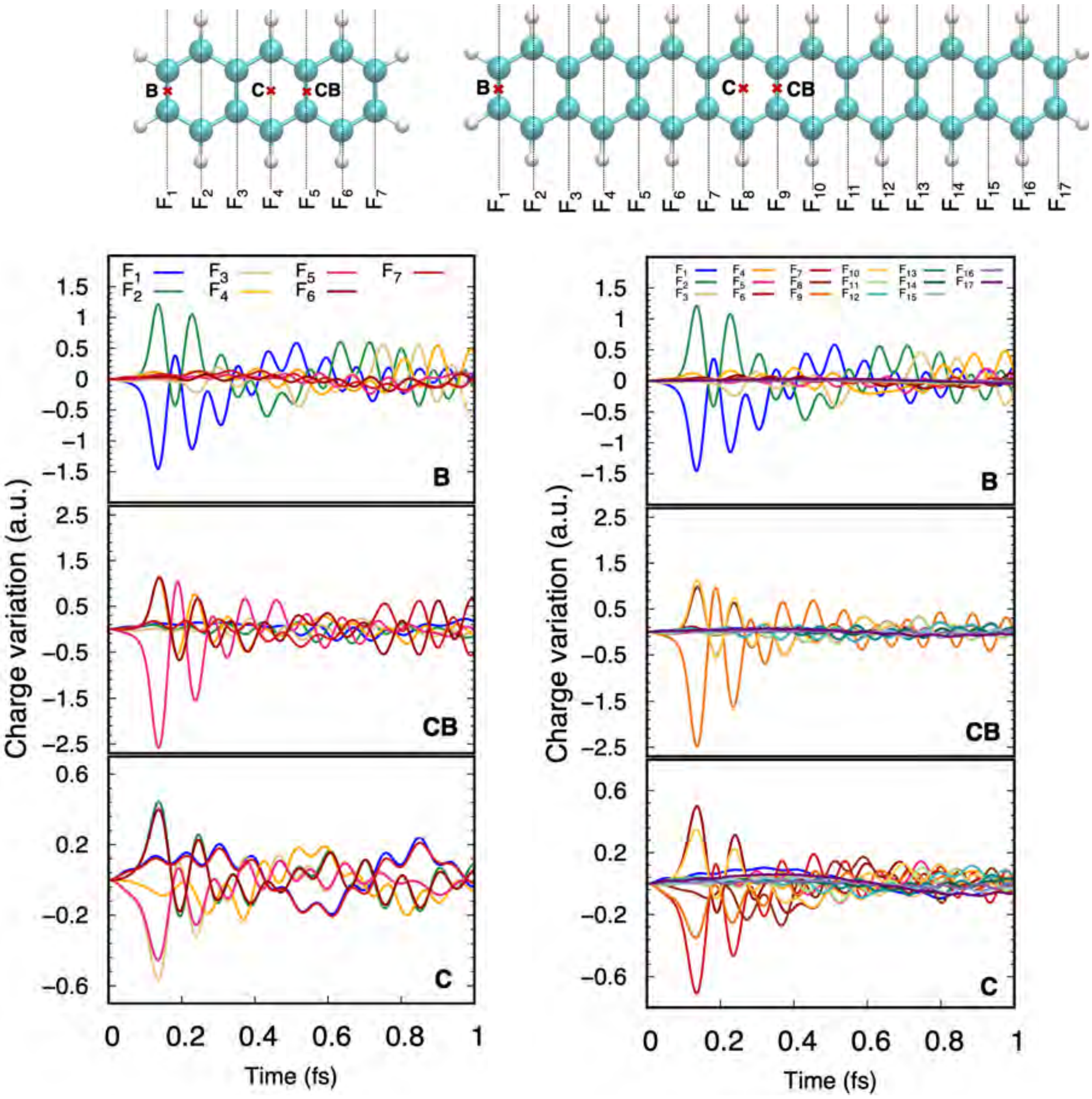


Figure 5.9: Ultrafast charge evolution for anthracene and octacene after collision with a 100 keV proton impacting the center of a central aromatic ring (C) or an central (CB) or border (B) bond.

5.5 Conclusions and Perspectives

In this chapter, an implementation of the **Real Time-Time Dependent-DFTB (RT-TD-DFTB)** method in deMonNano has been reported. Several DFTB (DFTB0, DFTB2, DFTB3) levels of theory can be employed for the explicit electron dynamic simulation, as well as the **hybrid DFTB(QM)/MM class 1 FF** approach detailed in Chapter 3.

Within the RT-TD-DFTB approach, two new implementations for the Second-Order Magnus Propagator[319] (SOMP) coupled with the Predictor-Corrector scheme[320] are detailed. With this latter, a notably reduction in the computational time was observed ($\sim$80% faster for the PC) while the accuracy in calculations was preserved.

A benchmark study has been performed, comparing several absorption spectra obtained either from LR-TD-DFTB[313] or from RT-TD-DFTB methods in deMonNano. In the latter case, a last implementation consisting of a **δ-kick pulse** initialization was needed in order to mimic the linear and weak perturbation modelled by LR-TD-DFTB approach. A set of three small molecules were selected, NH_2CH_2COOH, C_2H_2 and H_2O. In the absorption spectra comparison, it was observed how the main absorption peaks were found similar in the spectra computed by the LR-TD-DFTB and RT-TD-DFTB methods.

In addition, a set of preliminary simulations of **ion collision induced electron dynamics** have been carried out for two molecular systems. Six different N-Glycine and PAH molecular systems were submitted to a proton collision, while the **ultrafast charge evolution** has been monitored. Several induced electron displacements were observed under the effect of the proton collision.

As a perspective, we note that, the electronic stopping power (ESP), *i.e.* the energy deposition during collision can be studied using RT-TD-DFTB approach. We have performed preliminary calculations to compare the ESP obtained by RT-TD-DFTB and RT-TD-DFT methods. Two simple test case systems were modelled: a simple H^+-water molecule collision (1MeV) and a H^+-coronene collision (from 0.01MeV to 1MeV). In Figure 5.10, the ESP distribution (upper panels) and the average energy deposition (lower panels) are illustrated. An underestimation of both energies obtained from the RT-TD-DFTB methods is observed. From a first drawn hypothesis, it is thought that this understimation might be explained by the lack of description of the core electron excitation in DFTB method. Another perspective of this work consists of including the proper description of the ionization processes

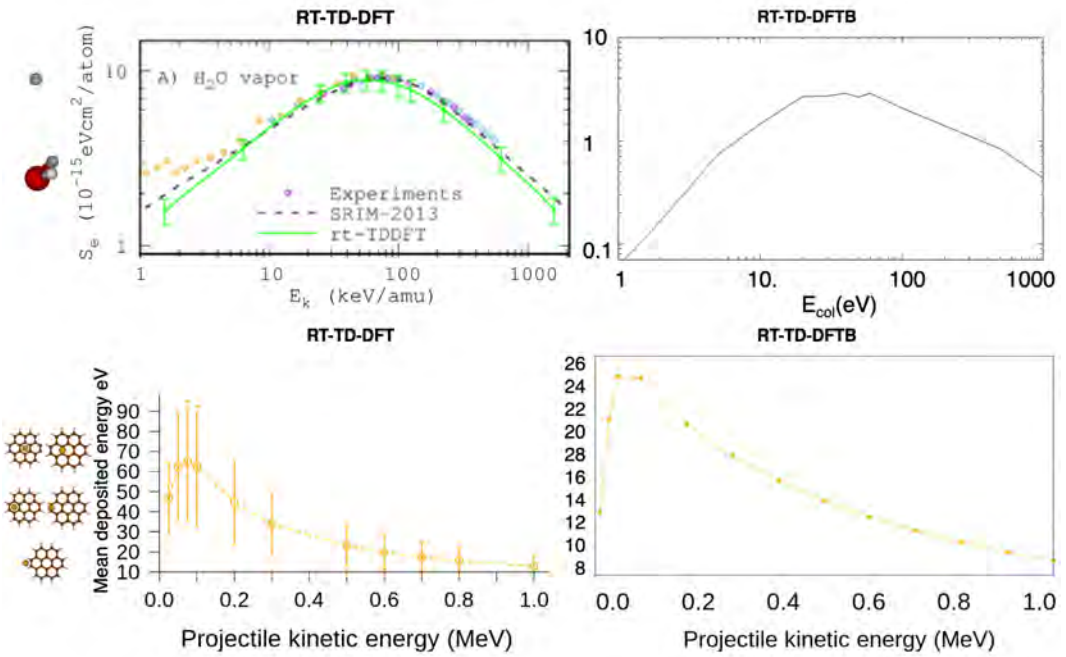


Figure 5.10: Preliminary ESP and energy deposition benchmark calculations for H$^+$-water collision (top panels) and H$^+$-coronene collision (bottom panels). First RT-TD-DFT results from Gu. B *et al.*[330] (top panel). Coronene study performed by A de La Lande (private comunication).

after collision (more details can be found in General Conclusions and Perspectives).

Moreover, better improvements in the orbital description could be done by adding some diffuse or polarization orbitals in the DFTB basis. On the other hand, this orbital's tunning should be done carefully in order to preserve the main advantage of the DFTB method, *i.e.* the calculations computational cost. More details can be found in General Conclusions and Perspectives. In the future, an extended benchmark study should be performed.

In this chapter, explicit electron dynamics has been simulated. In order to capture longer time scale processes, *e.g* chemisorption, fragmentation, etc, a coupling between the electronic and nuclear dynamics is needed. In the future, this coupling will be done by introducing the Non Adiabatic Molecular Dynamics (NAMD) Ehrenfest[331] scheme.

General Conclusions and Perspectives

General Conclusions

As stated in the general introduction, this thesis work aims to understand the behaviour of complex astrochemical molecules such as glycine prebiotic molecule, immersed in a interstellar ice model under different conditions relevant for the Interstellar Medium. In addition, the behaviour of this astrochemical system under the effect of high energy irradiation processes was also investigated. This thesis has contributed to broadening the knowledge of glycine conformational space and the related transformations including tautomerization mechanisms, in ice-water media at different temperatures relevant with respect to the interstellar ice conditions. In addition, preliminary results have been provided on the ultrafast electron charge evolution of ion-N-Glycine collision. Similar results were provided for the ion-PAH collision.

In order to achieve these objectives, two new implementations were done in the deMonNano[149] software package. Firstly, in order to deal with the large system sizes, the combination of Quantum Mechanics (QM) and Molecular Mechanics (MM) approaches have been implemented where the DFTB level of theory has been used for the energy and forces computation in the QM part, and class 1 force fields potentials have been included in the molecular mechanics description. Furthermore, for the study of ultrafast evolution phenomena, the Real Time-Time Dependent-DFTB (RT-TD-DFTB) method has been implemented and applied to the explicit simulation of electron dynamics. Both DFTB(QM)/MM class 1 FF and the RT-TD-DFTB implementations were combined and applied to simulating explicitly the ultrafast electron dynamics of large sized systems (N-Glycine embedded in a LDA model ice) under ion collision.

Hybrid QM/MM methods have been widely used in computational chemistry. In Chapter 3, a complete introduction of the methods and their different schemes has been presented. In this thesis, the **hybrid DFTB(QM/MM) implementation was performed successfully using the additive coupling approach, in both, mechanical and electrostatic embedding schemes**. Note that three DFTB0, DFTB2 and DFTB3 levels of theory have been employed. This new implementation allows to use any class 1 force field potential, as long as the user provides the corresponding parameters. To the date, AMBERF99[258] and OPLS-AA[260] parameters have been used. It is important to mention that, in this implementation, **the polarization of the QM region by the charges of the MM part is achieved.**

RT-TD-DFTB methods have been extensively used for simulating the explicit electron dynamics after an external perturbation. A general introduction of the approach was reported in Chapter 5. The **successful implementation of RT-TD-DFTB method has been done including the Second-order Magnus Propagator[319] (SOMP) coupled with the Predictor-Corrector[320] scheme**. This latter ensures a significant reduction of the computational resources while the accuracy of calculations is maintained. The main external perturbations that has been targeted in this work are the ion-collision type. In addition, a δ-kick pulse perturbation has also been developed, drawn by the interest of simulating linear and weak perturbations. From this latter implementation, several absorption spectra of a set of different molecules after a δ-kick pulse perturbation were successfully calculated with the RT-TD-DFTB approach, providing a benchmark study.

Regarding the applications, many efforts have been dedicated to the conformational study of glycine molecule and the related transformation in a water ice environment and are reported in Chapter 4. QM/MM 10 ps molecular dynamic simulations have been carried out at three temperatures (20, 250 and 450K) for four different neutral glycine (N-Gly) isomers (*cc*, *ct*, *tc*, *tt*) immersed in three water clusters sizes (n=4, 17 and 27 H_2O molecules) and in a large size LDA ice model (343-344 H_2O molecules). All simulations have been done by both DFT/MM and

DFTB/MM class 1 methods. From the **DFT/MM dynamics**, five main conclusion can be drawn:

- For the N-Gly in the LDA ice model, **all conformers have been found to be stable at temperatures of 20 and 250K**. Nevertheless, several transformations have been observed for a high temperature (450K).
- **The large water network around glycine must be explicitly considered in order to capture the tautomerization mechanisms** or any intermediate transformation (proton delocalization or anionic form).
- With sufficient number of molecules, the efficiency of the tautomerization process N-Gly → Z-Gly (zwitterion glycine) depends on the nature of the glycine isomers. For example, in the case of *cc*-N-Gly-nH_2O, mostly direct proton transfer mechanism occurred while, in the case of *ct*- and *tc*, the concerted proton transfer mechanism has been captured. **This confirms that considering the various glycine isomers is mandatory to provide a complete picture of the tautomerization mechanism.**
- It has been demonstrated that **the tautomerization process is strongly influenced by the nature of its conformational isomers. Z-Gly isomer is predominantly observed for *cc*-N-Gly, whereas *ct*-, *tc*- and *tt*-N-Gly undergo other kinds of transformations *e.g* anion and proton delocalized forms**.
- **From the Astrochemistry point of view, this research suggests that glycine in the cold interstellar ices should remain in its neutral form up to at least 250K consistently with the experimental results in Iopollo *et al.*[26]**

From the latter statement, the first question outlined in the introduction chapter in this thesis has been answered: *Does glycine molecule maintains its original structure under the effect of water-ice (LDA) environment and the effect of temperature?*. Yes at low temperatures. For a range of temperatures between 20 to 250K, glycine molecules maintain its neutral structure for all the four different *cc, ct, tc, tt* isomers.

From the **DFTB/MM** molecular dynamic calculations at 10 ps, a need of **reparametrization of BIO DFTB standard parameters** appeared. Thanks to reparametrizations of C-C, N-H and O-H atoms pairs repulsion potentials, tautomerization process in glycine molecules have been observed along with anionic and proton transfer molecular transformations in agreement with DFT calculations. Nevertheless, the new reparametrizations are thought to underestimate the repulsive character of O-H and N-H repulsive energies.

Finally in Chapter 5, the preliminary studies of the ultrafast electronic dynamic induced by ion collision have been carried out using the RT-TD-DFTB method. In first place, the N-Glycine molecule in gas phase and immersed in a LDA model ice has been bombarded by a proton particle mimicking a solar winds collision with a kinetic energy of 10keV. Three different impact parameters were chosen and the ultrafast electron charge evolution has been monitored. **Several electron displacements were observed induced by the proton impact.**

From this statement, the second question highlighted in the introduction chapter of this work has been answered: *What is the electronic evolution of glycine molecule induced by a collision?* It can be observed that electron charge motions mostly occur within glycine molecule (intramolecular displacement). A charge oscillation between the glycine molecule and the ice environment is observed however, this intermolecular displacement is smaller than the intramolecular one. In all cases, an ebb-and-flow effect was captured, where the electron attraction through the impact point was followed by its immediate repulsion. A period of charge oscillation was then started.

Finally, induced electron dynamics was also captured during the collision processes in the ion-PAH collision study.

General Perspectives

In previous paragraphs, the progresses achieved in this thesis have been presented. Nevertheless, still a range of improvements exists with the aim to enhance the method implementations and to enhance the understanding of glycine molecular system in conditions relevant for the Interstellar Medium.

In first place, an upgraded scheme for the DFTB/MM class 1 FF approach will be performed. This latter will consist of the coupling of polarizable atoms description at the MM level with the DFTB quantum part. This new approach will allow to capture not only the polarization of QM atoms by the MM charges, but also **to take into account the polarization of MM atoms by the QM charges**. The work of Iftner *et al.*[263] for argon atoms and similar works performed at the DFT level[169, 252] will be taken as a reference.

In the case of **RT-TD-DFTB method**, the preliminary ion-collision calculations performed for a water or a coronene molecule evaluating the energy stopping power (ESP) and the energy deposition, point difficulties in their proper description. Several reasons can be involved:

- The lack of core-electron description in the DFTB method ban the access to the possible core-electron excitations. Nevertheless, it is believed that depending on the kind of collision and deposited energy, the core-electron excitations will appear or not. In the future, a complete benchmark study comparing with DFT calculations will be performed in order to properly understand this phenomena.
- As reported in Spiegelman *et al.*,[313] the description of the Rydberg states is not reached in DFTB, as only valence orbitals are described, thus, the access to the Rydberg excitation processes is banned. This is one of the reason why **ionization or electron attachment are not properly described in DFTB.** In order to overcome these deficiencies, in the future, it can be interesting to include diffuse basis functions that would give access to Rydberg states, or to include a discrete set of virtual orbitals.[313]
- The problem of the electron-hole (or electron lost in the continuum) when modeling ionization processes has been widely reported in literature. Due to this lack of description of the electron lost in the continium in DFTB, the ionization process cannot be modelled. Different tools have been suggested to its resolution such as the use of **Complex Absorbing Potential (CAP)**.[332] This latter consists of adding a supplementary imaginary potential to the Kohn-Sham potential, where the excited electrons in the non-bonding (virtual) orbitals will be removed when they reach a certain threshold distance or energy criteria. In addition, an expanded effect of the non-bonding levels is achieved yielding to a continuum representation.[329, 333] Several studies[178, 329]are going to be taken as a reference to implement the CAP potential or its variances in deMonNano.

At last, thanks to the RT-TD-DFTB implementation, the explicit electron dynamics can be simulated however only for fixed ions. In order to capture long scale processes, e.g chemisorption, fragmentation, dissociation, etc, a **coupling between the electronic and nuclear dynamics is needed**. These dynamics,also known as **Non-Adiabatic Molecular Dynamics (NAMD)**, are beyond the Born-Oppenheimer approximation. To the date, Tully's FSSH approached coupled with

LR-TD-DFTB has been implemented in deMonNano.[313] However, this later uses Linear Response approach with TD-DFTB to compute the excited states. At each simulation step, the LR-excited states, have to be computed turning into long and expensive computational simulations. In addition, although the LR-TD-DFTB approach gives all different LR-exited states, it deals with an inappropriate description of the S_0/S_1 (ground state/first excited state) non-adiabatic coupling description.[190] Therefore, another approach must be employed when looking for a good description of the electron dynamics, *e.g* electron relaxation in a molecular system excited by CRs collision. This is the case of Ehrenfest scheme[331] combining the Real Time-TD-DFTB approach with the nuclear dynamics classical description. In the future, this new implementation will be performed in the deMonNano software.

To conclude, different perspectives are also considered, regarding on the better understanding of the chemical behaviour of glycine interstellar ice systems before and after high energy collision processes:

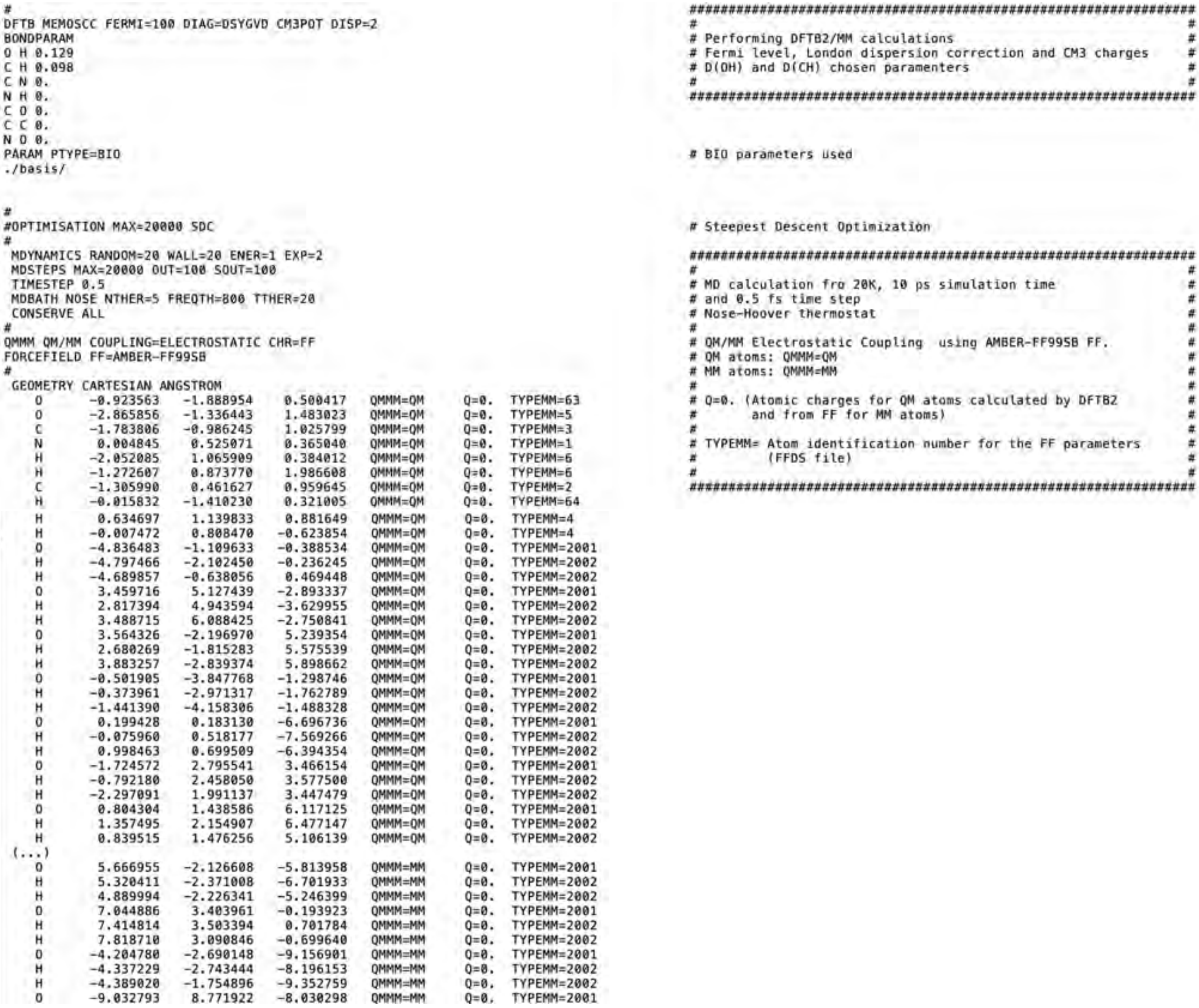
- Firstly, several improvements of the BIO reparametrized DFTB potential will be performed, in particular on the N-H and O-H atomic pairs repulsion energies. With this refinement, a better description of glycine systems by the DFTB method is expected. Therewith, a complete study will be carried out using DFTB/MM class 1 approach, where similar MD simulations as reported in Chapter 4 will be performed increasing the statistics.
- With the new implementation of RT-TD-DFTB with the Ehrenfest approach combining electron and nuclei dynamics, a comprehension of the chemical evolution of glycine interstellar system after collision is expected to be achieved. Incorporating the nuclear dynamics will allow to model fragmentation, dissociation or other chemical processes that are expected to appear after collision phenomena. Therewith, in the future, Non-Adiabatic Molecular Dynamics will be performed for the glycine molecular systems under collision processes. Note that similar dynamics are intended for PAHs compounds.
- In a recent work of Kroonblawd *et al.*,[334] the formation of nitrogen-containing polycyclic aromatic hydrocarbons (NPAHs) from a possible shock synthesis from impacting material like glycine-water mixtures was reported. This work sparks new ideas of upgraded molecular systems that can be studied in the future with a clear connection between both prebiotic glycine and PAH compounds studied in this thesis.

Examples of deMonNano Input

List of deMonNano inputs

- Example of DFTB/MM FF class 1
- Example of FFDS external file (AMBER-FF99SB force)
- Example of Ion-Collision input

Example of DFTB/MM FF class 1



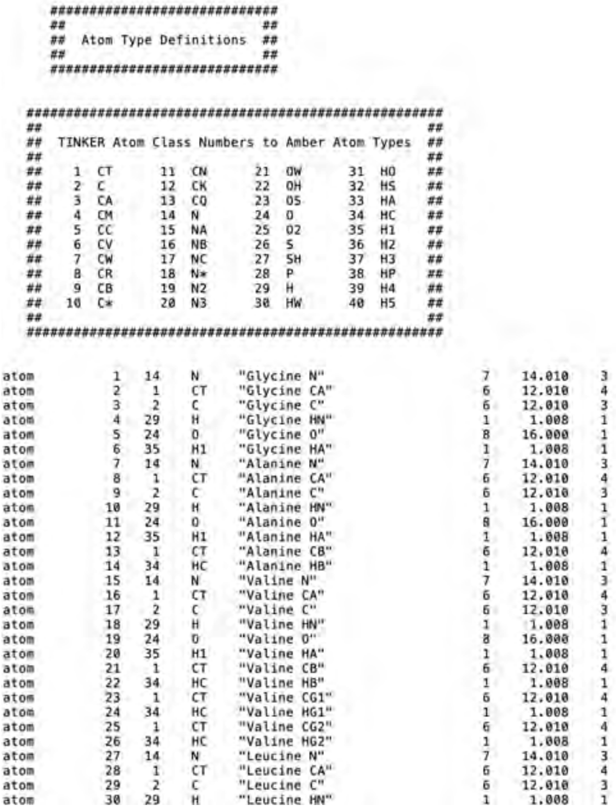
```
#
DFTB MEMOSCC FERMI=100 DIAG=DSYGVD CM3POT DISP=2
BONDPARAM
O H 0.129
C H 0.098
C N 0.
N H 0.
C O 0.
C C 0.
N O 0.
PARAM PTYPE=BIO
./basis/


#
#OPTIMISATION MAX=20000 SDC
#
 MDYNAMICS RANDOM=20 WALL=20 ENER=1 EXP=2
 MDSTEPS MAX=20000 OUT=100 SOUT=100
 TIMESTEP 0.5
 MDBATH NOSE NTHER=5 FREQTH=800 TTHER=20
 CONSERVE ALL
#
QMMM QM/MM COUPLING=ELECTROSTATIC CHR=FF
FORCEFIELD FF=AMBER-FF99SB
#
 GEOMETRY CARTESIAN ANGSTROM
    O      -0.923563   -1.888954    0.500417   QMMM=QM      Q=0.   TYPEMM=63
    O      -2.865856   -1.336443    1.483023   QMMM=QM      Q=0.   TYPEMM=5
    C      -1.783806   -0.986245    1.025799   QMMM=QM      Q=0.   TYPEMM=3
    N       0.004845    0.525071    0.365040   QMMM=QM      Q=0.   TYPEMM=1
    H      -2.052085    1.065909    0.384012   QMMM=QM      Q=0.   TYPEMM=6
    H      -1.272607    0.873770    1.986608   QMMM=QM      Q=0.   TYPEMM=6
    C      -1.305990    0.461627    0.959645   QMMM=QM      Q=0.   TYPEMM=2
    H      -0.015832   -1.410230    0.321005   QMMM=QM      Q=0.   TYPEMM=64
    H       0.634697    1.139833    0.881649   QMMM=QM      Q=0.   TYPEMM=4
    H      -0.007472    0.808470   -0.623854   QMMM=QM      Q=0.   TYPEMM=4
    O      -4.836483   -1.109633   -0.388534   QMMM=QM      Q=0.   TYPEMM=2001
    H      -4.797466   -2.102450   -0.236245   QMMM=QM      Q=0.   TYPEMM=2002
    H      -4.689857   -0.638056    0.469448   QMMM=QM      Q=0.   TYPEMM=2002
    O       3.459716    5.127439   -2.893337   QMMM=QM      Q=0.   TYPEMM=2001
    H       2.817394    4.943594   -3.629955   QMMM=QM      Q=0.   TYPEMM=2002
    H       3.488715    6.088425   -2.750841   QMMM=QM      Q=0.   TYPEMM=2002
    O       3.564326   -2.196970    5.239354   QMMM=QM      Q=0.   TYPEMM=2001
    H       2.680269   -1.815283    5.575539   QMMM=QM      Q=0.   TYPEMM=2002
    H       3.883257   -2.839374    5.898662   QMMM=QM      Q=0.   TYPEMM=2002
    O      -0.501905   -3.847768   -1.298746   QMMM=QM      Q=0.   TYPEMM=2001
    H      -0.373961   -2.971317   -1.762789   QMMM=QM      Q=0.   TYPEMM=2002
    H      -1.441390   -4.158306   -1.488328   QMMM=QM      Q=0.   TYPEMM=2002
    O       0.199428    0.183130   -6.696736   QMMM=QM      Q=0.   TYPEMM=2001
    H      -0.075960    0.518177   -7.569266   QMMM=QM      Q=0.   TYPEMM=2002
    H       0.998463    0.699509   -6.394354   QMMM=QM      Q=0.   TYPEMM=2002
    O      -1.724572    2.795541    3.466154   QMMM=QM      Q=0.   TYPEMM=2001
    H      -0.792180    2.458050    3.577500   QMMM=QM      Q=0.   TYPEMM=2002
    H      -2.297091    1.991137    3.447479   QMMM=QM      Q=0.   TYPEMM=2002
    O       0.804304    1.438586    6.117125   QMMM=QM      Q=0.   TYPEMM=2001
    H       1.357495    2.154907    6.477147   QMMM=QM      Q=0.   TYPEMM=2002
    H       0.839515    1.476256    5.106139   QMMM=QM      Q=0.   TYPEMM=2002
 (...)
    O       5.666955   -2.126608   -5.813958   QMMM=MM      Q=0.   TYPEMM=2001
    H       5.320411   -2.371008   -6.701933   QMMM=MM      Q=0.   TYPEMM=2002
    H       4.889994   -2.226341   -5.246399   QMMM=MM      Q=0.   TYPEMM=2002
    O       7.044886    3.403961   -0.193923   QMMM=MM      Q=0.   TYPEMM=2001
    H       7.414814    3.503394    0.701784   QMMM=MM      Q=0.   TYPEMM=2002
    H       7.818710    3.090846   -0.699640   QMMM=MM      Q=0.   TYPEMM=2002
    O      -4.204780   -2.690148   -9.156901   QMMM=MM      Q=0.   TYPEMM=2001
    H      -4.337229   -2.743444   -8.196153   QMMM=MM      Q=0.   TYPEMM=2002
    H      -4.389020   -1.754896   -9.352759   QMMM=MM      Q=0.   TYPEMM=2002
    O      -9.032793    8.771922   -8.030298   QMMM=MM      Q=0.   TYPEMM=2001
```

```
####################################################################
#                                                                  #
# Performing DFTB2/MM calculations                                 #
# Fermi level, London dispersion correction and CM3 charges       #
# D(OH) and D(CH) chosen paramenters                               #
#                                                                  #
####################################################################


# BIO parameters used


# Steepest Descent Optimization

####################################################################
#                                                                  #
# MD calculation fro 20K, 10 ps simulation time                    #
# and 0.5 fs time step                                             #
# Nose-Hoover thermostat                                           #
#                                                                  #
# QM/MM Electrostatic Coupling  using AMBER-FF99SB FF.             #
# QM atoms: QMMM=QM                                                #
# MM atoms: QMMM=MM                                                #
#                                                                  #
# Q=0. (Atomic charges for QM atoms calculated by DFTB2           #
#       and from FF for MM atoms)                                  #
#                                                                  #
# TYPEMM= Atom identification number for the FF parameters        #
#         (FFDS file)                                              #
#                                                                  #
####################################################################
```

Example of FFDS external file (AMBER-FF99SB force)

```
      ##############################
      ##                          ##
      ##  Force Field Definition  ##
      ##                          ##
      ##############################

forcefield              AMBER-FF99SB
vdwtype                 LENNARD-JONES
radiusrule              ARITHMETIC
radiustype              R-MIN
radiussize              RADIUS
epsilonrule             GEOMETRIC
vdw-14-scale            2.0
chg-14-scale            1.2
electric                332.0522173
dielectric              1.0

      #############################
      ##                         ##
      ##  Literature References  ##
      ##                         ##
      #############################

V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling,
Comparison of Multiple Amber Force Fields and Development of Improved
Protein Backbone Parameters, PROTEINS, 65, 712-725 (2006)  [PARM99SB]

J. Wang, P. Cieplak and P. A. Kollman, "How Well Does a Restrained
Electrostatic Potential (RESP) Model Perform in Calcluating Conformational
Energies of Organic and Biological Molecules?", J. Comput. Chem., 21,
1049-1074 (2000)  [PARM99]

W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr.,
D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman,
"A Second Generation Force Field for the Simulation of Proteins, Nucleic
Acids, and Organic Molecules", J. Am. Chem. Soc., 117, 5179-5197 (1995)
[PARM94]

G. Moyna, H. J. Williams, R. J. Nachman and A. I. Scott, "Conformation in
Solution and Dynamics of a Structurally Constrained Linear Insect Kinin
Pentapeptide Analogue", Biopolymers, 49, 403-413 (1999)  [AIB charges]

W. S. Ross and C. C. Hardin, "Ion-Induced Stabilization of the G-DNA
Quadruplex: Free Energy Perturbation Studies", J. Am. Chem. Soc.,
116, 6070-6080 (1994)  [Alkali Metal Ions]

J. Aqvist, "Ion-Water Interaction Potentials Derived from Free Energy
Perturbation Simulations", J. Phys. Chem., 94, 8021-8024 (1990)
[Alkaline Earth Ions, radii adapted for Amber combining rule]

Current parameter values are available from the Amber site, located
at http://ambermd.org/
```

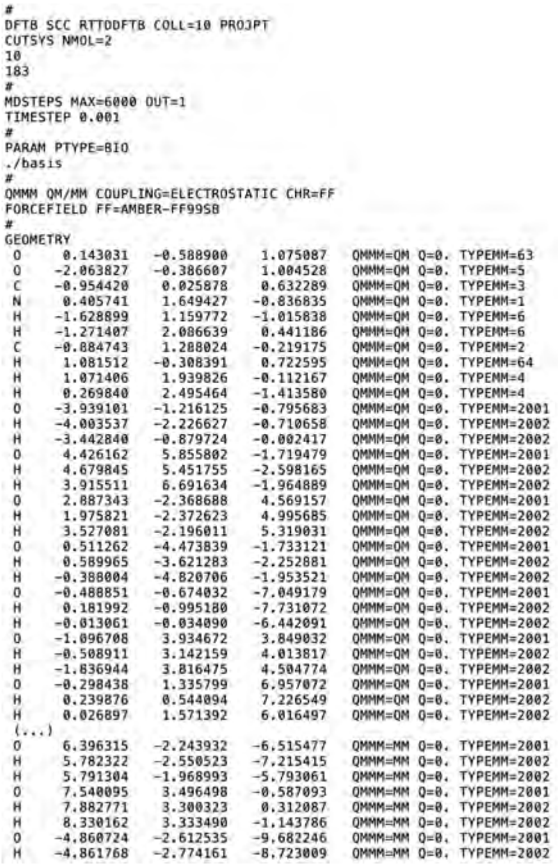
(...)



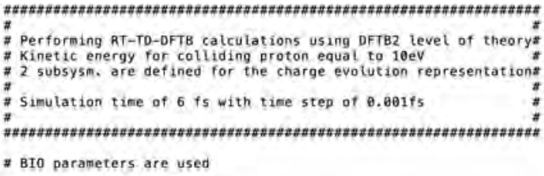
```
      #############################
      ##                         ##
      ##  Atom Type Definitions  ##
      ##                         ##
      #############################

   ######################################################
   ##                                                  ##
   ##  TINKER Atom Class Numbers to Amber Atom Types   ##
   ##                                                  ##
   ##    1  CT      11  CN      21  OW      31  HO     ##
   ##    2  C       12  CK      22  OH      32  HS     ##
   ##    3  CA      13  CQ      23  OS      33  HA     ##
   ##    4  CM      14  N       24  O       34  HC     ##
   ##    5  CC      15  NA      25  O2      35  H1     ##
   ##    6  CV      16  NB      26  S       36  H2     ##
   ##    7  CW      17  NC      27  SH      37  H3     ##
   ##    8  CR      18  N*      28  P       38  HP     ##
   ##    9  CB      19  N2      29  H       39  H4     ##
   ##   10  C*      20  N3      30  HW      40  H5     ##
   ##                                                  ##
   ######################################################

atom          1   14    N     "Glycine N"                   7    14.010    3
atom          2    1    CT    "Glycine CA"                  6    12.010    4
atom          3    2    C     "Glycine C"                   6    12.010    3
atom          4   29    H     "Glycine HN"                  1     1.008    1
atom          5   24    O     "Glycine O"                   8    16.000    1
atom          6   35    H1    "Glycine HA"                  1     1.008    1
atom          7   14    N     "Alanine N"                   7    14.010    3
atom          8    1    CT    "Alanine CA"                  6    12.010    4
atom          9    2    C     "Alanine C"                   6    12.010    3
atom         10   29    H     "Alanine HN"                  1     1.008    1
atom         11   24    O     "Alanine O"                   8    16.000    1
atom         12   35    H1    "Alanine HA"                  1     1.008    1
atom         13    1    CT    "Alanine CB"                  6    12.010    4
atom         14   34    HC    "Alanine HB"                  1     1.008    1
atom         15   14    N     "Valine N"                    7    14.010    3
atom         16    1    CT    "Valine CA"                   6    12.010    4
atom         17    2    C     "Valine C"                    6    12.010    3
atom         18   29    H     "Valine HN"                   1     1.008    1
atom         19   24    O     "Valine O"                    8    16.000    1
atom         20   35    H1    "Valine HA"                   1     1.008    1
atom         21    1    CT    "Valine CB"                   6    12.010    4
atom         22   34    HC    "Valine HB"                   1     1.008    1
atom         23    1    CT    "Valine CG1"                  6    12.010    4
atom         24   34    HC    "Valine HG1"                  1     1.008    1
atom         25    1    CT    "Valine CG2"                  6    12.010    4
atom         26   34    HC    "Valine HG2"                  1     1.008    1
atom         27   14    N     "Leucine N"                   7    14.010    3
atom         28    1    CT    "Leucine CA"                  6    12.010    4
atom         29    2    C     "Leucine C"                   6    12.010    3
atom         30   29    H     "Leucine HN"                  1     1.008    1
```

(...)

Example of Ion-Collision input



```
#
DFTB SCC RTTDDFTB COLL=10 PROJPT
CUTSYS NMOL=2
10
183
#
MDSTEPS MAX=6000 OUT=1
TIMESTEP 0.001
#
PARAM PTYPE=BIO
./basis
#
QMMM QM/MM COUPLING=ELECTROSTATIC CHR=FF
FORCEFIELD FF=AMBER-FF99SB
#
GEOMETRY
 O      0.143031   -0.588900    1.075087   QMMM=QM Q=0. TYPEMM=63
 O     -2.063827   -0.386607    1.004528   QMMM=QM Q=0. TYPEMM=5
 C     -0.954420    0.025878    0.632289   QMMM=QM Q=0. TYPEMM=3
 N      0.405741    1.649427   -0.836835   QMMM=QM Q=0. TYPEMM=1
 H     -1.628899    1.159772   -1.015838   QMMM=QM Q=0. TYPEMM=6
 H     -1.271407    2.086639    0.441186   QMMM=QM Q=0. TYPEMM=6
 C     -0.884743    1.288024   -0.219175   QMMM=QM Q=0. TYPEMM=2
 H      1.081512   -0.308391    0.722595   QMMM=QM Q=0. TYPEMM=64
 H      1.071406    1.939826   -0.112167   QMMM=QM Q=0. TYPEMM=4
 H      0.269840    2.495464   -1.413580   QMMM=QM Q=0. TYPEMM=4
 O     -3.939101   -1.216125   -0.795683   QMMM=QM Q=0. TYPEMM=2001
 H     -4.003537   -2.226627   -0.710658   QMMM=QM Q=0. TYPEMM=2002
 H     -3.442840   -0.879724   -0.002417   QMMM=QM Q=0. TYPEMM=2002
 O      4.426162    5.855802   -1.719479   QMMM=QM Q=0. TYPEMM=2001
 H      4.679845    5.451755   -2.598165   QMMM=QM Q=0. TYPEMM=2002
 H      3.915511    6.691634   -1.964889   QMMM=QM Q=0. TYPEMM=2002
 O      2.887343   -2.368688    4.569157   QMMM=QM Q=0. TYPEMM=2001
 H      1.975821   -2.372623    4.995685   QMMM=QM Q=0. TYPEMM=2002
 H      3.527081   -2.196011    5.319031   QMMM=QM Q=0. TYPEMM=2002
 O      0.511262   -4.473839   -1.733121   QMMM=QM Q=0. TYPEMM=2001
 H      0.589965   -3.621283   -2.252881   QMMM=QM Q=0. TYPEMM=2002
 H     -0.388004   -4.820706   -1.953521   QMMM=QM Q=0. TYPEMM=2002
 O     -0.488851   -0.674032   -7.049179   QMMM=QM Q=0. TYPEMM=2001
 H      0.181992   -0.995180   -7.731072   QMMM=QM Q=0. TYPEMM=2002
 H     -0.013061   -0.034090   -6.442091   QMMM=QM Q=0. TYPEMM=2002
 O     -1.096708    3.934672    3.849032   QMMM=QM Q=0. TYPEMM=2001
 H     -0.508911    3.142159    4.013817   QMMM=QM Q=0. TYPEMM=2002
 H     -1.836944    3.816475    4.504774   QMMM=QM Q=0. TYPEMM=2002
 O     -0.298438    1.335799    6.957072   QMMM=QM Q=0. TYPEMM=2001
 H      0.239876    0.544094    7.226549   QMMM=QM Q=0. TYPEMM=2002
 H      0.026897    1.571392    6.016497   QMMM=QM Q=0. TYPEMM=2002
 (...)
 O      6.396315   -2.243932   -6.515477   QMMM=MM Q=0. TYPEMM=2001
 H      5.782322   -2.550523   -7.215415   QMMM=MM Q=0. TYPEMM=2002
 H      5.791304   -1.968993   -5.793061   QMMM=MM Q=0. TYPEMM=2002
 O      7.540095    3.496498   -0.587093   QMMM=MM Q=0. TYPEMM=2001
 H      7.882771    3.300323    0.312087   QMMM=MM Q=0. TYPEMM=2002
 H      8.330162    3.333490   -1.143786   QMMM=MM Q=0. TYPEMM=2002
 O     -4.860724   -2.612535   -9.682246   QMMM=MM Q=0. TYPEMM=2001
 H     -4.861768   -2.774161   -8.723009   QMMM=MM Q=0. TYPEMM=2002
```



```
#################################################################
#                                                               #
# Performing RT-TD-DFTB calculations using DFTB2 level of theory#
# Kinetic energy for colliding proton equal to 10eV             #
# 2 subsysm. are defined for the charge evolution representation#
#                                                               #
# Simulation time of 6 fs with time step of 0.001fs             #
#                                                               #
#################################################################

# BIO parameters are used
```

List of Publications

- **Coupling density functional based tight binding with class 1 force fields in a hybrid QM/MM scheme.** Maysá Yusef Buey, Tzonka Mineva, Mathias Rapacioli *Theor Chem Acc* **141**, 16 (2022). DOI: 10.1007/s00214-022-02878-6
- **Temperature driven transformations of glycine molecules embedded in interstellar ice.** Maysá Yusef Buey, Tzonka Mineva, Dahbia Talbi, Mathias Rapacioli *Phys. Chem. Chem. Phys.*, (2024), Advance Article. DOI: 10.1039/d3cp03575a.
- **Addressing electronic and dynamical evolution of molecules and molecular clusters : DFTB simulations of energy relaxation in Polycyclic Aromatic Hydrocarbons.** Mathias Rapacioli, Fernand Spiegelman, Maysá Yusef Buey *Phys. Chem. Chem. Phys.*, (2023), Advance Article. DOI: 10.1039/D3CP02852F (Review Article).

Résumé étendu en français

Les molécules complexes présentes dans le milieu interstellaire, telles que les molécules organiques complexes (COM pour *Complex Organic Molecules*) et les hydrocarbures aromatiques polycycliques (PAH pour *Polycyclic aromatic hydrocarbon*), interessent la communauté scientifique depuis longtemps. La compréhension de leur formation, de leur comportement dans/sur les glaces interstellaires ou de leur évolution sous l'influence d'irradiations des rayonnement cosmique, a pris une grande importance car on pense qu'elles jouent un rôle clé dans l'apparition de la vie sur Terre.

Cette thèse est principalement consacrée à l'étude de la molécule prébiotique de glycine immergée dans une glace interstellaire amorphe de faible densité (LDA pour *Low Density Amorphous*) et des PAH isolés, en utilisant des méthodes de chimie théorique. Il s'agit d'un travail complexe, compte tenu de la variété des conformations des COM et de leurs évolutions possibles dans différentes conditions interstellaires. Cette complexité a nécessité le développement d'outils théoriques qui ont été implémentés dans le logiciel deMonNano. Tout d'abord, afin de traiter les systèmes de grande taille, la combinaison des approches de la mécanique quantique (QM pour *Quantum Mechanics*) et de la mécanique moléculaire (MM) a été implémentée où le niveau de théorie DFTB (pour *Density Functional Tight-Binding*) est utilisé pour le calcul de l'énergie et des forces dans la partie QM et le potentiel des champs de force de classe 1 est utilisé dans la description de la MM. Ensuite, pour l'étude des phénomènes ultrarapides, la méthode RT-TD-DFTB (pour *Real Time-Time Dependent-Density Functional Tight-Binding*) a été aussi implémentée. Les implémentations DFTB(QM)/MM et RT-TD-DFTB ont été combinées.

L'évolution dynamique de la molécule de glycine a d'abord été étudiée en intégrant la glycine dans la glace LDA et en utilisant l'approche QM/MM. Les questions de l'espace conformationnel de la glycine et des transformations associées ont été abordées pour des conditions astrophysiques pertinentes. Les résultats ont confirmé que le vaste réseau d'eau autour de la glycine et que les divers isomères de la glycine doivent être explicitement pris en compte afin de comprendre les mécanismes de tautomérisation. Du point de vue de l'astrochimie, cette recherche suggère que la glycine dans les glaces interstellaires froides devrait rester sous sa forme neutre jusqu'à au moins 250 K.

Enfin, les études préliminaires de la dynamique électronique ultrarapide induite par la collision ionique ont été réalisées à l'aide de la méthode RT-TD-DFTB, où la molécule de N-Glycine dans le LDA a été bombardée par un proton. Les résultats ont confirmé que les migrations de charges électroniques se produisent principalement à l'intérieur de la molécule de glycine (migration intramoléculaire), bien qu'une faible migration intermoléculaire soit observée entre la molécule de glycine et l'environnement de glace. La migration des électrons a également été observée lors de l'étude de la collision ion-PAH. Dans les deux cas, un effet de flux et de reflux a été observé.

En conclusion, cette thèse a contribué à l'élaboration de nouveaux outils de calcul, permettant d"etudier des systèmes de grande taille et la dynamique électronique ultrarapide. En ce qui concerne les applications, ce travail de thèse a permis d'élargir les connaissances relatives à l'espace conformationnel de la glycine et aux transformations qui y sont liées. De plus, la réponse électronique de ces systèmes soumis à des collisions avec des vents stellaires ou des particules de rayons cosmiques a été simulée. Grâce aux nouveaux outils théoriques qui ont été implémentés et de leurs extensions, de nouvelles perspectives sont ouvertes pour l'étude théorique de l'évolution chimique des COMs induite par l'irradiation et des phénomènes d'ionisation.

Dans ce qui suit, les différents chapitres de ce manuscrit sont introduits, en soulignant les idées et les conclusions principales.

Chapitre 1. Molécules Complexes dans le Milieu Interstellaire

La compréhension de l'univers observable et de sa composition est l'un des principaux objectifs de l'astrophysique. Les progrès des observations astronomiques et des études expérimentales ou théoriques en laboratoire ont permis de mieux comprendre les composés astronomiques, leur formation et leur évolution possible. Au cours des dernières décennies, les recherches sur la formation, la distribution et le comportement des molécules complexes dans l'espace ont pris de l'importance car on pense qu'elles jouent un rôle clé dans l'apparition de la vie sur Terre. L'hypothèse principale repose sur l'idée que les grains de glace emprisonnant des molécules complexes s'agglomèrent en glaces de grande taille. Ces glaces interstellaires se transforment finalement en comètes dans les systèmes planétaires en formation et arrivent sur la Terre primitive, constituant ainsi les précurseurs des origines de la vie sur Terre.

Dans ce travail de thèse, nous étudierons des molécules complexes immergées dans les glaces interstellaires ou dans leurs formes isolées. Cette thèse sera principalement consacrée à l'étude de la **molécule prébiotique de glycine immergée dans une glace interstellaire**. La glycine (voir Figure S.1) est l'un des acides aminés les plus simples présents dans de nombreuxplusieurs systèmes biologiques et a déjà été détectée à deux reprises dans la queue de deux comètes différentes.

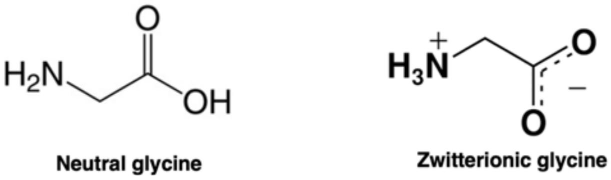


Figure S.1: Représentation de Lewis de la molécule de glycine prébiotique dans ses deux principaux conformère : neutre et zwitterionique.

Parailleurs, en raison de leur présence importante dans le milieu interstellaire, les hydrocarbures aromatiques polycycliques (PAH) (voir Figure S.2) sous leur forme isolée seront également étudiés.

Ce chapitre d'introduction présente le contexte astrochimique. Des détails sur les glaces interstellaires, la molécule prébiotique glycine et les composés PAH sont présentés. De plus, une brève introduction sur les processus de collision à haute énergie se produisant dans l'ISM est rapporté.

Les **objectifs de cette thèse** sont les suivants :

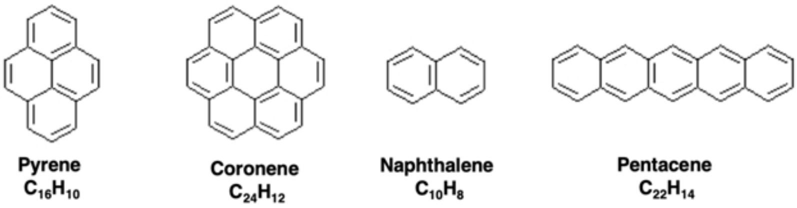


Figure S.2: Illustration de la taille et des conformations de plusieurs PAH.

- Nous abordons les questions de l'espace conformationnel de la glycine et des transformations associées, y compris les mécanismes de tautomérisation, dans les milieux glace-eau et à différentes températures pertinentes par rapport aux conditions de la glace interstellaire.

  La molécule de glycine conserve-t-elle sa structure originale sous l'effet de l'environnement eau-glace LDA et sous l'effet de la température ?

- Nous abordons les questions de la dynamique électronique ultrarapide explicite de la molécule de glycine intégrée dans une glace interstellaire sous l'effet d'une collision avec des vents solaires à haute énergie.

  Quelle est l'évolution électronique de la molécule de glycine suite à une collision ?

- Dans le cas des PAH, une étude similaire est réalisée dans le but de comprendre **la dynamique électronique ultrarapide des PAH après collision avec une particule à haute énergie.**

- Nous utiliserons les niveaux de théorie DFT (pour *Density functional theory*) et DFTB en utilisant respectivement les logiciels deMon2k[148] et deMonNano.[149] Il convient de noter qu'en raison des limitations informatiques, les calculs DFT ont parfois été remplacés par des calculs DFTB. Afin d'atteindre les objectifs mentionnés ci-dessus, deux nouvelles implémentations ont été nécessaires dans le logiciel deMonNano. En premier lieu, la modélisation de la molécule de glycine piégée dans une glace interstellaire devient un défi en raison du grand nombre d'atomes présents dans ce système.

  Pour ces raisons, la méthode quantique DFTB sera combinée avec la mécanique moléculaire de classe 1 dans un schéma hybride QM/MM.

- La simulation des processus ultrarapides induits par les collisions à haute énergie nécessite d'aller au-delà du problème de Schrödinger indépendant du temps et de modéliser la dynamique explicite des électrons.

> ***Pour cette raison, l'approche DFTB en temps réel - dépendante du temps (RT-TD-DFTB) sera mise en œuvre dans deMonNano en combinaison avec l'approche DFTB/MM de classe 1.***

Chapitre 2. Méthodes et Outils de Simulation

La mécanique quantique est née au XXe siècle de la nécessité de décrire correctement le comportement des particules atomiques et subatomiques au-delà de la description de la mécanique classique. L'étude de Planck sur la quantification de l'émission de lumière, le développement mathématique de Mileva Mariç et l'affirmation d'Einstein selon laquelle le comportement de la lumière est à la fois corpusculaire et ondulatoire, ont marqué le début de la mécanique quantique. En outre, la suggestion de de Broglie concernant l'aspect ondulatoire du mouvement des électrons et le principe d'incertitude d'Heisenberg ont jeté les bases de la mécanique quantique. La chimie quantique a pour but d'étudier les propriétés chimiques d'un système en utilisant la mécanique quantique. Aujourd'hui, la chimie computationnelle utilise des ressources informatiques pour résoudre les problèmes de chimie quantique. Elle utilise différentes méthodes et approches pour résoudre les postulats de la mécanique quantique

Une introduction à l'équation de Schrödinger et à l'approximation de Born-Oppenheimer est présentée dans ce chapitre. En outre, une vue d'ensemble des principales méthodes de calcul de la structure électronique visant à résoudre ce problème quantique est présentée. D'autres détails sur l'exploration des surfaces d'énergie potentielle du système et la simulation de leur évolution dynamique dans le temps sont décrits.

Un résumé des deux principales méthodes employées dans le cadre de ce travail de thèse est détaillé ici : DFT et DFTB.

Théorie de la Fonctionnelle de la Densité (DFT)

Le théorème essentielle de Kohn-Sham (KS) établit que la densité de l'état fondamental $\rho_0(\mathbf{r})$ d'un **système en interaction** (soumis à un potentiel externe $v(\mathbf{r})$) peut être reproduite avec précision dans un **système sans interaction** soumis à un potentiel $v_{KS}(\mathbf{r})$. Avec cette approche, l'énergie électronique d'un système sans interaction sera calculée à partir d'un ensemble auxiliaire d'orbitales Φ^{KS}. L'énergie électronique totale d'un système $E^{DFT}[\rho(\mathbf{r})]$ est donnée par :

$$\begin{aligned} E^{DFT}[\rho(\mathbf{r})] &= \sum_{i=1}^{N} -\frac{1}{2}\int d\mathbf{r}\phi_i^{KS*}(\mathbf{r})\nabla^2\phi_i^{KS}(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{\rho(\mathbf{r})\rho\left(\mathbf{r}'\right)}{|\mathbf{r}-\mathbf{r}'|} \\ &+ E_{xc}[\rho] + \int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r}) \\ &= T_{KS}[\rho(\mathbf{r})] + E_H[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] + \int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r}) \end{aligned} \tag{S.1}$$

Il faut noter que dans l'approche KS, la densité électronique dans le premier terme de l'Eq. S.1 est donnée en termes d'orbitales de Kohn Sham. $E_H[\rho(\mathbf{r})]$ est l'énergie de Hartree (deuxième terme de l'Eq. S.1) et $v(\mathbf{r})$ est un potentiel externe. Enfin, $E_{xc}[\rho]$ est **l'énergie d'échange-corrélation** contenant les différences entre le système réel d'interaction des électrons et le système sans interaction.

Il est important de noter que l'Eq. S.1 est l'expression exacte de l'énergie DFT totale. D'autre part, la fonctionnelle de l'énergie d'échange-corrélation est le seul terme inconnu dans l'expression de l'énergie DFT totale. La description correcte de cette fonctionnelle reste l'un des principaux défis de la chimie quantique. Plusieurs approches pour estimer E_{xc} ont été développées et se distinguent en quatre familles: **1) Approximations de la densité locale (LDA pour *Local Density Approximation*)** traitant la densité électronique localement comme un gaz d'électrons uniforme ; **2) Approximation du gradient généralisé (GGA pour *Generalized Gradient Approximation*)** ajoutant la dérivée première de la densité ; **3) Meta-GGA** incluant le laplacien de la densité en plus de la dérivée première, et enfin **4) Fonctionnelles hybrides** contenant un pourcentage de la corrélation d'échange HF et une approximation (semi ou totalement empirique) de la fonctionnelle de corrélation d'échange en DFT, *e. g.* PBE0,[164] B3LYP,[165–168] etc.

Théorie de la Fonctionnelle basée sur Liaison Forte (DFTB)

Les limites des méthodes basées sur la fonction d'onde dite *ab initio* ou sur la DFT sont liées à la taille des systèmes simulés et aux exigences de calcul. Lorsque les systèmes sont constitués d'un grand nombre d'atomes, ou lorsqu'il s'agit de réaliser un grand nombre d'énergies ponctuelles, des approches telles que la DFTB deviennent utiles. La DFTB est basée sur la utilisation d'une base reduite d'orbitales atomiques,[179] en plus d'une quantité d'éléments paramétrés susceptibles de réduire le temps de calcul (jusqu'à ∼3 ordres de grandeur plus rapide que la DFT).

En combinant l'expression de l'énergie DFT avec la série d'expansion de Taylor, et la définition de l'opérateur de Kohn-Sham (la dérivation partielle de la fonctionnelle de l'énergie DFT par rapport à la densité), l'équation ci-dessus peut être réécrite comme suit :

$$
\begin{gathered}
\left.\begin{aligned}
E = &\int d\mathbf{r}v(\mathbf{r})\rho_0(\mathbf{r}) + E_{xc}[\rho_0] - \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}'\frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\\
&+\sum_{i=1}^{N} n_i \left\langle \phi_i(\mathbf{r}) \middle| \hat{H}_{KS} \middle| \phi_i(\mathbf{r}) \right\rangle
\end{aligned}\right\}\mathbf{E^{DFTB0}}\\
\left.+\frac{1}{2}\int\int\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}+\left.\frac{\delta^2 E_{xc}[\rho(\mathbf{r})]}{\delta\rho\delta\rho'}\right|_{\rho=\rho_0}\right)\delta\rho\delta\rho'\qquad\right\}\mathbf{E^{2nd}}
\end{gathered}
\tag{S.2}
$$

L'énergie totale E contient différentes contributions énergétiques. Dans la première version de DFTB,[179] les termes de second ordre et d'ordre supérieur sont négligés. Il s'agit de la **méthode DFTB d'ordre zéro (DFTB0)**. Plus tard, en 1998, Elstner *et al.*[180] a introduit la dernière ligne de l'équation S.2 afin d'améliorer l'équilibre des charges dans les systèmes hétéronucléaires, ce qui a donné la **méthode DFTB du second ordre (DFTB2)**. Après plusieurs étapes de dérivation, l'expression de l'énergie totale de la méthode DFTB2 est donnée par :

$$
E^{DFTB2} = E_{rep}[\rho_0(\mathbf{r})] + \sum_{i=1}^{N} n_i \sum_{\mu\nu}^{M} c_{\mu i}c_{\nu i}H^0_{\mu\nu} + \frac{1}{2}\sum_{a=1}^{K}\sum_{b=1}^{K}\Delta q_a\gamma_{ab}\Delta q_b \tag{S.3}
$$

Chapitre 3. Implémentation de Méthode Hybride (QM/MM) dans deMonNano et Étude Comparative.

Les schémas hybrides QM/MM populaires *i.e* la combinaison des méthodes de chimie quantique (QM) et des approches de mécanique moléculaire (MM) (également appelées méthodes de type champ de force), sont apparus dans les années 70 basées sur les travaux de Warshel et Karplus[230] et Warshel et Levitt.[231] Les approches QM/MM recherchent un juste équilibre entre la fiabilité des calculs et l'application à des systèmes de grande taille, et donc une consommation adéquate de temps de calcul. Alors que les méthodes QM sont employées pour donner une description correcte d'un processus chimique se déroulant dans un système moléculaire, les approches MM intègrent les effets environnementaux possibles sur le système moléculaire.

Dans le cadre de l'approche QM, tout type de méthode de calcul de la structure électronique peut en principe être utilisé, en rappelant que plus la méthode est précise (fonction d'onde ou méthodes basées sur la DFT), plus le temps de calcul sera important. Pour cette raison et en raison de la taille importante des systèmes étudiés dans cette thèse, une nouvelle implémentation QM/MM a été réalisée dans le code deMonNano et est présentée dans ce chapitre. Cette approche est basée sur la combinaison de la méthode de liaison serrée DFTB avec les champs de force de classe 1 DFTB/MM (DFTB0-DFTB2-DFTB3/MM). Dans ce qui suit, en l'absence de spécifications, DFTB peut se référer indistinctement à DFTB0, DFTB2 ou DFTB3.

Dans ce chapitre, de nouvelles mises en œuvre d'une **approche hybride de couplage additif DFTB(QM)/MM, dans les schémas d'intégration mécanique et électrostatique** sont présentées. Différents niveaux de théorie **DFTB** (DFTB0, DFTB2 et DFTB3) sont utilisés pour le calcul de l'énergie et des forces dans la partie QM et divers **champs de force potentiels de classe 1** ont été inclus dans la description de la mécanique moléculaire. Actuellement, seuls les champs de force de classe 1 AMBER-FF99SB[258] et OPLS-AA (OPLS)[260] ont été utilisés dans les études de référence, mais cette nouvelle implémentation permet d'utiliser n'importe quel champ de force de classe 1, à condition que l'utilisateur fournisse les paramètres correspondants.

Dans l'approche additive des deux schémas d'intégration mécanique et électrostatique, les interactions entre les sous-régions du système MM et QM sont calculées à partir d'un potentiel de Lenard-Jones. Dans le couplage électrostatique, un terme supplémentaire décrit les interactions de Coulomb entre les charges atomiques du QM et du MM. Les charges atomiques du domaine MM restent constantes, mais celles de la région QM sont optimisées de manière autocohérence dans le cadre du schéma variationnel DFTB, y compris **la polarisation de la région QM par les charges de la région MM**.

Les **gradients énergétiques**, obligatoires pour effectuer la dynamique moléculaire ou pour calculer les spectres vibrationnels, ont également été implémentés pour ce schéma hybride. Des benchmarks ont été réalisés, comparant les gradients d'énergie calculés soit à partir de la méthode analytique DFTB/MM, soit à partir de différences finies. En outre, il a été confirmé que l'énergie totale pendant la simulation de dynamique moléculaire dans un ensemble micro-canonique était conservée pendant les simulations. En outre, il a été confirmé que l'énergie totale pour une simulation de dynamique moléculaire dans un ensemble micro-canonique était conservée pendant les simulations.

Un benchmark plus complet a été rapporté, pour un ensemble de petites molécules sélectionnées (NH_3, NH_2CH_2COOH, C_5H_12 et H_2O) dans un environnement d'aggrégats d'eau ($\sim$ 215 molécules H_2O), en comparant les spectres IR harmoniques calculés aux niveaux DFTB3/MM et DFT/MM. OPLS-AA (modèle TIP3P pour les molécules d'eau) a été utilisé pour paramétrer le potentiel du champ de force. Ce benchmark confirme que **les principaux effets de l'environnement calculés au niveau DFT/MM sont reproduits au niveau DFTB3/MM**.

À l'avenir, la combinaison de la présente implémentation avec des schémas similaires à celui développé par Iftner *et al.*[263] pour coupler la description des atomes d'argon polarisables au niveau MM avec le schéma DFTB nous permettrait de mélanger DFTB avec des champs de force polarisables de manière similaire à ce qui a été fait au niveau DFT[169,252] .

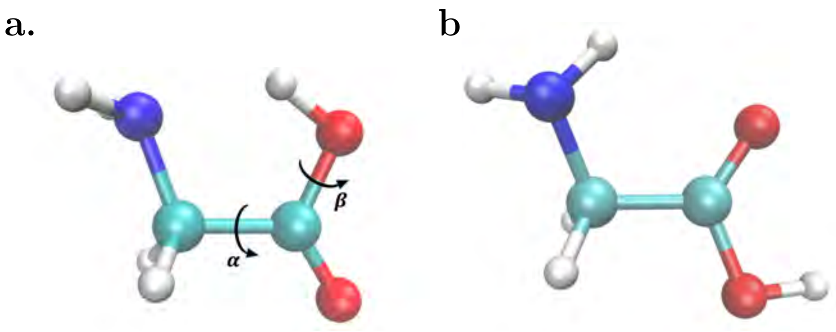
Chapitre 4. Transformations induites par la température de la molécule neutre de glycine intégrée dans la glace interstellaire

Comme nous l'avons déjà présenté dans le premier chapitre, l'acide aminé glycine et son formation sur les grains de glace dans l'espace soulèvent des questions fondamentales sur la chimie de la glycine dans les milieux interstellaires. Dans ce travail, nous avons étudié l'espace conformationnel de la glycine et les mécanismes de tautomérisation associés dans les milieux aqueux dans des conditions astrophysiques pertinentes au moyen de simulations de dynamique moléculaire DFT/MM et DFTB/MM.

Huit isomères N-Gly, les plus stables en phase gazeuse, ont été identifiés à partir des calculs CCSD(T).[55] Dans cet ensemble de conformères de glycine, l'orientation de l'hydrogène carboxylique par rapport à l'atome N a été prise en compte. Pour les besoins de notre travail, seuls quatre des huit isomères rapportés ont été choisis, sans tenir compte de la rotation autour de la liaison C-N. Elle représente diverses rotations du groupe -COOH autour de la liaison C-C et la rotation de l'hydrogène carboxylique, comme le montre la figure S.3.

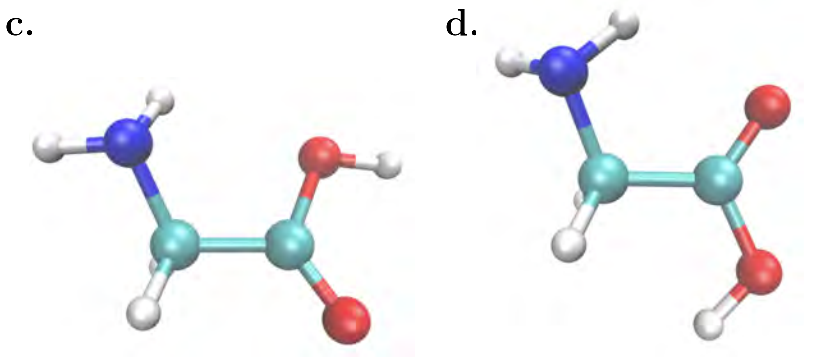
Pour modéliser les glaces interstellaires, nous avons utilisé le modèle de glace LDA (Low Density Amorphous) de Ghesquière *et al.*[123]

En outre, l'évolution chimique de la glycine dépend principalement des interactions avec les molécules d'eau les plus proches (première couche de solvatation). L'évolution dynamique dépend également des molécules d'eau plus éloignées de la glycine. Afin de caractériser cet effet, nous avons effectué des simulations dans des environnements d'eau plus petits. Elles ne correspondent pas aux structures les plus



cis-cis(*cc*)-N-Gly $\alpha \sim [6^\circ]$ $\beta \sim [-1.5^\circ]$ $E_R = 0.0$ kcal.mol^{-1}

trans-trans(*tt*)-N-Gly $\alpha \sim [162^\circ]$ $\beta \sim [177^\circ]$ $E_R = 2.6$ kcal.mol^{-1}



cis-trans(*ct*)-N-Gly $\alpha \sim [-30^\circ]$ $\beta \sim [-178^\circ]$ $E_R = 4.0$ kcal.mol^{-1}

trans-cis(*tc*)-N-Gly $\alpha \sim [166^\circ]$ $\beta \sim [-2.5^\circ]$ $E_R = 7.5$ kcal.mol^{-1}

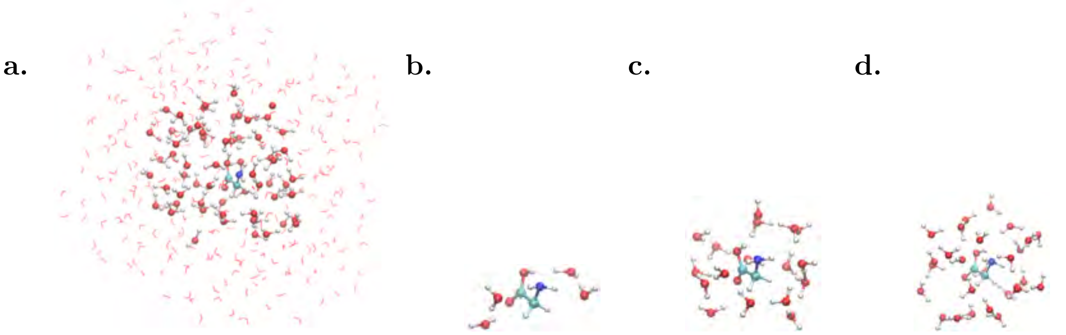
Figure S.3: DFT optimized neutral glycine (N-Gly) isomers. Cis and trans conformational character described by the α and β dihedral angles[284] .

stables des clusters (comme dans d'autres travaux[58, 59, 286, 287]).

Conclusions de l'Exploration DFT et DFT/MM et Implications Astrochimiques.

Cette étude portant sur l'exploration des mécanismes de tautomérisation et des effets de l'environnement hydrique et de la température dans différents systèmes moléculaires de glycine a permis de dégager les tendances générales suivantes :

- Dans les modèles représentatifs de la glace, la glycine neutre est stabilisée à basse température (20 et 250K), mais pas à haute température (450K). Dans ce dernier cas, la transformation de tautomérisation vers la forme Z-Gly a été principalement observée pour le *cc*-N-Gly-LDA, alors que pour les autres



***cc*-Gly - (61/62) H_2O (QM) - 282 H_2O (MM)**   ***cc*-Gly- 4H_2O**   ***cc*-Gly- 17H_2O**   ***cc*-Gly- 27H_2O**

Figure S.4: Representative snapshots of the four different N-Gly embedded molecular systems extracted from the BOMD simulations. The atoms, treated at DFT level are shown as balls and those at MM level as lines. a) *cc*-Gly- 61 H_2O(QM)-282 H_2O(MM) model; b) *cc*-Gly- 4H_2O cluster; c) *cc*-Gly- 17H_2O cluster. d)*cc*-Gly- 27H_2O cluster[284] .

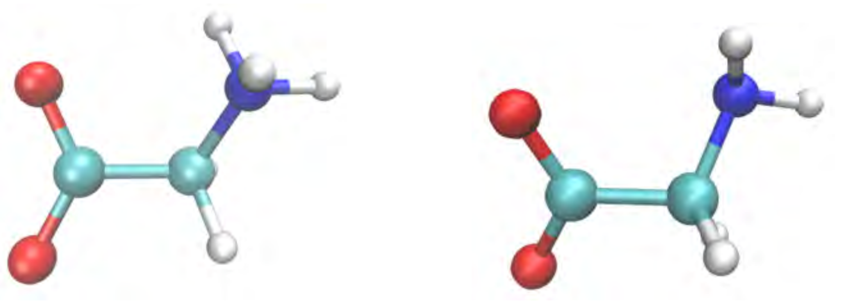
isomères, des formes de délocalisation de protons ont été observées.

- Les tendances observées dans le modèle représentatif de la glace, et en particulier la tautomérisation, ne peuvent être observées qu'en considérant explicitement un grand réseau d'eau autour de la glycine, les simulations avec un plus petit nombre de molécules d'eau conduisant à des conclusions différentes.

- Avec un nombre suffisant de molécules, l'efficacité du processus de tautomérisation (N-Gly → Z-Gly) dépend de la nature des isomères de la glycine. C'est également le cas du mécanisme associé. Par exemple, le mécanisme direct ne s'est produit que pour les isomères *cc*-N-Gly-nH2O, alors que pour les systèmes moléculaires *ct*- et *tc*-N-Gly, le transfert de proton a été assisté par les molécules d'eau dans le cadre d'un mécanisme concerté.

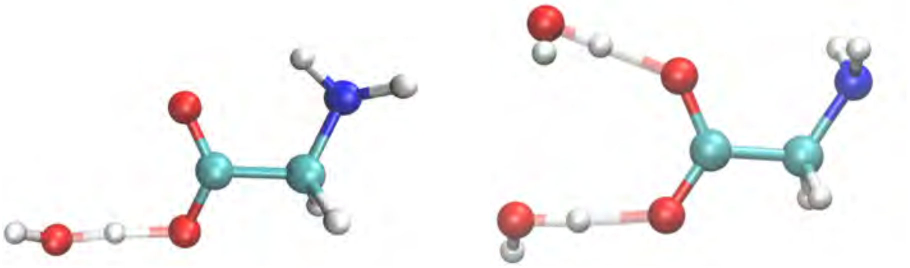
Ce travail renforce l'idée que la prise en compte des différents isomères de la glycine est obligatoire pour obtenir une image complète des mécanismes de tautomérisation (voir Figure S.5). En plus des structures initiales des isomères, l'autre facteur conduisant à la tautomérisation est la représentation de l'environnement. Les dynamiques de l'eau et de la glycine dépendent également de la température. Pour les clusters de taille intermédiaire, l'agrégation eau - eau se produit plus rapidement à $T = 450K$, ce qui stabilise la N-Gly. A cette température et dans le modèle d'eau de grande taille (à partir du modèle LDA-ice), *cc*-N-Gly est entièrement converti en Z-Gly, alors que pour les trois autres isomères, la délocalisation entre H_2O et le groupe -COO- domine.

En outre, nous avons démontré que **le processus de tautomérisation de la**



Zwitterion-Gly
$(H_3\ N^+CH_2COO^-)$

Anion-Gly
$(NH_2CH_2COO^-)$



H delocalized-Gly
$[OH\cdots H\cdots NH_2CH_2COO]^-$

Double H delocalized-Gly
$[OH\cdots H\cdots NH_2CH_2COO\cdots H\cdots OH]^-$

Figure S.5: Illustration de quatre formes différentes de glycine observées le long de la dynamique : glycine zwitterion (panneau supérieur gauche), anion (panneau supérieur droit) et formes de glycine à délocalisation d'un et de deux protons (panneaux inférieurs gauche et droit).

glycine est fortement influencé par la nature de son isomère conformationnel.

Du point de vue de l'astrochimie, notre étude suggère que la glycine dans les glaces interstellaires froides devrait rester neutre jusqu'à au moins 250 K, conformément aux résultats expérimentaux de Iopollo *et al.*[26].

Du point de vue des simulations chimiques, notre étude suggère fortement que la modélisation de la chimie interstellaire en masse nécessite quelques couches d'eau (notre modèle de glace LDA) entourant les réactifs pour éviter l'auto-agrégation des molécules d'eau.

Conclusions de l'Exploration DFTB et DFTB/MM.

Grâce à une nouvelle reparamétrisation des potentiels répulsifs V_{rep}^{CC}, V_{rep}^{NH} et

V_{rep}^{OH} de la DFTB, de grandes améliorations ont été obtenues :

- Optimisations locales de Z-Gly ont montré que les structures minimales de la glycine neutre peuvent être explorées. La rupture de la liaison C-C apparaissant lors des optimisations locales utilisant les paramètres BIO standard est désormais évitée grâce à la nouvelle reparamétrisation.
- Alors que dans les calculs DFTB avec les paramètres BIO très peu de mécanismes de tautomérisation ont été observés, les quatre conformères de glycine (neutre, zwitterionique, anionique et proton délocalisé) et leurs mécanismes de transformation respectifs sont détectés avec cette nouvelle reparamétrization.
- Avec moins de limitations dans les ressources de calcul par rapport à l'approche DFT, toutes les trajectoires DFTB pour les plus grands systèmes modèles de glace ont été complétées jusqu'à 10 ps.

En outre, dans le cas du modèle de glace LDA, à haute température (450 K), une évaporation presque complète des molécules d'eau a été observée, ce qui compromet la précision des résultats.

En conclusion, bien qu'une amélioration de l'étude de la glycine soit observée avec les nouveaux paramètres BIO reparamétrés, quelques réglages des potentiels DFTB V_{rep}^{OH} et V_{rep}^{NH} doivent encore être effectués avant d'élargir l'étude de la glycine. En particulier, comme idée préliminaire, un réglage intermédiaire entre les paramètres BIO standards et les reparamétrages V_{rep}^{OH} et V_{rep}^{NH} utilisés dans ce chapitre pourrait être effectué. Dans les travaux futurs, à partir des potentiels DFTB reparamétrés optimaux, une extension à des tailles de systèmes plus importantes ainsi qu'une amélioration de l'échantillonnage statistique seront envisagées.

Chapitre 5. Implémentation en RT-TD-DFTB méthode dans deMonNano et Première Applications.

Comme indiqué au chapitre 1, des processus de collision à haute énergie peuvent exister dans le MIS ou le milieu circumstellaire, *e.g* les rayons cosmiques (CR) ou les vents stellaires. Ces processus peuvent induire des phénomènes d'excitation électronique. Jusqu'à présent, dans cette thèse, seuls les états stationnaires de l'état fondamental électronique ont été pris en compte, et l'équation de Schrödinger indépendante du temps a donc été résolue. Néanmoins, dans l'étude de ces processus dynamiques ultrarapides, la description des états électroniques excités et l'évolution temporelle dynamique des systèmes électroniques nécessitent d'aller au-delà du cas de l'état fondamental électronique stationnaire et donc de résoudre l'équation de Schrödinger dépendante du temps.

La méthode TDDFT (Time dependent) apparaît comme un excellent outil pour résoudre le problème de Schrödinger dépendant du temps, en utilisant soit la réponse linéaire (LR)-TD-DFT, soit l'approche temps réel (RT)-TD-DFT.

Dans ce chapitre, une implémentation de la méthode **Real Time-Time Dependent-DFTB (RT-TD-DFTB)** dans deMonNano est présentée. Plusieurs niveaux de théorie DFTB (DFTB0, DFTB2, DFTB3) peuvent être utilisés pour la simulation dynamique explicite des électrons, ainsi que l'approche **hybride DFTB(QM)/MM classe 1 FF**.

Dans le cadre de l'approche RT-TD-DFTB, deux nouvelles implémentations du propagateur de Magnus du second ordre (SOMP) couplé au schéma prédicteur-correcteur[320] sont détaillées. Avec ce dernier, une réduction notable du temps de calcul a été observée (∼80% plus rapide pour le PC) tandis que la précision des calculs a été préservée.

Une étude de référence a été réalisée, comparant plusieurs spectres d'absorption obtenus soit à partir des méthodes LR-TD-DFTB,[313] soit à partir de la méthode RT-TD-DFTB dans deMonNano. Dans ce dernier cas, une dernière mise en œuvre consistant en une initialisation par **pulse δ-kick** a été nécessaire afin d'imiter la perturbation linéaire et faible modélisée par l'approche LR-TD-DFTB. Un ensemble de trois petites molécules a été sélectionné : NH_2CH_2COOH, C_2H_2 et H_2O. Dans la comparaison des spectres d'absorption, il a été observé que les principaux pics d'absorption étaient similaires dans les spectres calculés par les méthodes LR-TD-DFTB et RT-TD-DFTB.

En outre, une série de simulations préliminaires de **dynamique électronique induite par une collision ionique** a été réalisée pour deux systèmes moléculaires. Six systèmes moléculaires différents de N-Glycine et de PAH ont été soumis à une collision avec un proton, afin d'étudier l'évolution **des ultrarapides charges**. Plusieurs migrations d'électrons ont été observées sous l'effet du proton.

Comme perspective, nous notons que le pouvoir d'arrêt électronique (ESP) et le dépôt d'énergie pendant la collision peuvent être étudiés en utilisant l'approche RT-TD-DFTB. Nous avons effectué des calculs préliminaires pour comparer l'ESP et le dépôt d'énergie obtenus par les méthodes RT-TD-DFTB et RT-TD-DFT. Deux systèmes simples ont été modélisés : une simple collision H^+-molécule d'eau (1MeV) et une collision H^+-coronène (de 0,01MeV à 1MeV). On observe une sous-estimation des deux énergies obtenues par les méthodes RT-TD-DFTB. D'après une première hypothèse, cette sous-estimation pourrait s'expliquer par l'absence de l'électrons des core dans la méthode DFTB. En outre, améliorations dans la description des orbitales pourraient être réalisées en ajoutant certaines orbitales diffuses ou de polarisation dans la base DFTB. D'un autre côté, le réglage de cette orbitale doit être effectué avec précaution afin de ne pas compromettre le principal avantage de la méthode DFTB, à savoir la rapidité des calculs. À l'avenir, il conviendrait d'effectuer une étude comparative.

Une autre perspective de ce travail consistera à inclure une description correcte des processus d'ionisation après la collision.

Enfin, dans ce chapitre, la dynamique explicite des électrons a été simulée. Afin de capturer les processus à longue échelle, *e.g* la chimisorption, la fragmentation, etc, un couplage entre la dynamique électronique et nucléaire est nécessaire. À l'avenir, ce couplage sera effectué en termes de dynamique moléculaire non adiabatique (NAMD) au moyen du schéma Ehrenfest.[331]

Bibliography

[1] J. Frenzel, A. F. Oliveira, N. Jardillier, T. Heine, G. Seifert. *Semi-relativistic, self-consistent charge Slater-Koster tables for density-functional based tight-binding (DFTB) for materials science simulations.*, **2004-2009**.

[2] P. Ghesquière. *Rôle des glaces interstellaires dans la complexitémoléculaire de l'espace: modélisation par les méthodes de la chimie théorique*, **2015**.

[3] M. Sewiło, S. B. Charnley, P. Schilke, V. Taquet, J. M. Oliveira, T. Shimonishi, E. Wirström, R. Indebetouw, J. L. Ward, J. T. van Loon, J. Wiseman, S. Zahorecz, T. Onishi, A. Kawamura, C. H. R. Chen, Y. Fukui, R. Hamedani Golshan. *ACS Earth and Space Chemistry*, **2019**. *3*, 2088–2109.

[4] A. Bacmann, V. Taquet, A. Faure, C. Kahane, C. Ceccarelli. *Astronomy & Astrophysics*, **2012**. *541*, L12.

[5] C. R. Arumainayagam, R. T. Garrod, M. C. Boyer, A. K. Hay, S. T. Bao, J. S. Campbell, J. Wang, C. M. Nowak, M. R. Arumainayagam, P. J. Hodge. *Chemical Society Reviews*, **2019**. *48*, 2293–2314.

[6] E. F. van Dishoeck, E. A. Bergin. *arXiv preprint arXiv:2012.01472*, **2020**.

[7] E. Landau. *Life and death of a planetary system.*

[8] T. Y. Brooke. *Endeavour*, **1997**. *21*, 101–104.

[9] J. Berkowitz. *The Stardust Revolution: The new story of our origin in the stars.* Rowman & Littlefield, **2022**.

[10] J. P. Gardner, J. C. Mather, M. Clampin, R. Doyon, M. A. Greenhouse, H. B. Hammel, J. B. Hutchings, P. Jakobsen, S. J. Lilly, K. S. Long. *Space Science Reviews*, **2006**. *123*, 485–606.

[11] P. Swings, L. Rosenfeld. *Astrophysical Journal, vol. 86, p. 483-486*, **1937**. *86*, 483–486.

[12] A. McKellar. *Publications of the Astronomical Society of the Pacific*, **1940**. *52*, 187–192.

[13] A. Douglas, G. Herzberg. *Astrophysical Journal, vol. 94, p. 381*, **1941**. *94*, 381.

[14] A. Cheung, D. M. Rank, C. Townes, D. D. Thornton, W. Welch. *Physical Review Letters*, **1968**. *21*, 1701.

[15] A. Cheung, D. M. Rank, C. Townes, D. D. Thornton, W. Welch. *Nature*, **1969**. *221*, 626–628.

[16] L. E. Snyder, D. Buhl, B. Zuckerman, P. Palmer. *Physical Review Letters*, **1969**. *22*, 679.

[17] J. Cernicharo, N. Marcelino, J. Pardo, M. Agúndez, B. Tercero, P. De Vicente, C. Cabezas, C. Bermúdez. *Astronomy & Astrophysics*, **2020**. *641*, L9.

[18] J. Cernicharo, N. Marcelino, M. Agúndez, C. Bermúdez, C. Cabezas, B. Tercero, J. Pardo. *Astronomy & Astrophysics*, **2020**. *642*, L8.

[19] M. Guélin, J. Cernicharo. *Frontiers in Astronomy and Space Sciences*, **2022**. *9*.

[20] C. Guilloteau, T. Oberlin, O. Berné, É. Habart, N. Dobigeon. *The Astronomical Journal*, **2020**. *160*, 28.

[21] P. Team:, O. Berné, É. Habart, E. Peeters, C. Team:, A. Abergel, E. A. Bergin, J. Bernard-Salas, E. Bron, J. Cami, E. Dartois, A. Fuente, J. R. Goicoechea, K. D. Gordon, Y. Okada, T. Onaka, M. Robberto, M. Röllig, A. G. G. M. Tielens, S. Vicente, M. G. e. a. Wolfire. *Publications of the Astronomical Society of the Pacific*, **2022**. *134*, 054301.

[22] O. Berné, M.-A. Martin-Drumel, I. Schroetter, J. R. Goicoechea, U. Jacovella, B. Gans, E. Dartois, L. Coudert, E. Bergin, F. Alarcon. *Nature*, **2023**, 1–3.

[23] J. Black, A. Dalgarno. *Astrophysical Journal Supplement Series, vol. 34, July 1977, p. 405-423.*, **1977**. *34*, 405–423.

[24] D. Smith. *Chemical reviews*, **1992**. *92*, 1473–1485.

[25] E. Herbst. *Frontiers in Astronomy and Space Sciences*, **2021**. *8*, 776942.

[26] S. Ioppolo, G. Fedoseev, K. J. Chuang, H. M. Cuppen, A. R. Clements, M. Jin, R. T. Garrod, D. Qasim, V. Kofman, E. F. van Dishoeck, H. Linnartz. *Nature Astronomy*, **2021**. *5*, 197–205.

[27] N. Biver, D. Bockelée-Morvan. *ACS Earth and Space Chemistry*, **2019**. *3*, 1550–1555.

[28] K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin. *Science advances*, **2016**. *2*, e1600285.

[29] *Nasa's webb finds water, and a new mystery, in rare main belt comet.*

[30] K. KVENVOLDEN, J. LAWLESS, K. PERING, E. PETERSON, J. FLORES, C. PONNAMPERUMA, I. R. KAPLAN, C. MOORE. *Nature*, **1970**. *228*, 923–926.

[31] J. E. Elsila, D. P. Glavin, J. P. Dworkin. *Meteoritics & Planetary Science*, **2009**. *44*, 1323–1330.

[32] K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin, J. D. Keyser, F. Dhooghe, B. Fiethe, S. A. Fuselier, S. Gasc, T. I. Gombosi, K. C. Hansen, M. Haessig, A. Jäckel, E. Kopp, A. Korth, L. L. Roy, U. Mall, B. Marty, O. Mousis, T. Owen, H. Rème, M. Rubin, T. Sémon, C.-Y. Tzou, J. H. Waite, P. Wurz. *Science Advances*, **2016**. *2*, e1600285.

[33] E. T. Parker, H. L. McLain, D. P. Glavin, J. P. Dworkin, J. E. Elsila, J. C. Aponte, H. Naraoka, Y. Takano, S. Tachibana, H. Yabuta, H. Yurimoto, K. Sakamoto, T. Yada, M. Nishimura, A. Nakato, A. Miyazaki, K. Yogata, M. Abe, T. Okada, T. Usui, M. Yoshikawa, T. Saiki, S. Tanaka, S. Nakazawa, Y. Tsuda, F. Terui, T. Noguchi, R. Okazaki, S. ichiro Watanabe, T. Nakamura. *Geochimica et Cosmochimica Acta*, **2023**. *347*, 42–57.

[34] C. Ceccarelli, L. Loinard, A. Castets, A. Faure, B. Lefloch. *Astronomy and Astrophysics*, **2000**. *362*, 1122–1126.

[35] Y.-J. Kuan, S. B. Charnley, H.-C. Huang, W.-L. Tseng, Z. Kisiel. *The Astrophysical Journal*, **2003**. *593*, 848.

[36] L. E. Snyder, F. J. Lovas, J. M. Hollis, D. N. Friedel, P. R. Jewell, A. Remijan, V. V. Ilyushin, E. A. Alekseev, S. F. Dyubko. *The Astrophysical Journal*, **2005**. *619*, 914.

[37] P. D. Holtom, C. J. Bennett, Y. Osamura, N. J. Mason, R. I. Kaiser. *The Astrophysical Journal*, **2005**. *626*, 940.

[38] R. Rubin, G. Swenson Jr, R. Benson, H. Tigelaar, W. Flygare. *Astrophysical Journal, vol. 169, p. L39*, **1971**. *169*, L39.

[39] M. Ferus, V. Laitl, A. Knizek, P. Kubelík, J. Sponer, J. Kára, J. Sponer, B. Lefloch,

G. Cassone, S. Civiš. *Astronomy & Astrophysics*, **2018**. *616*, A150.

[40] S. Zeng, D. Quénard, I. Jiménez-Serra, J. Martín-Pintado, V. Rivilla, L. Testi, R. Martín-Doménech. *Monthly Notices of the Royal Astronomical Society: Letters*, **2019**. *484*, L43–L48.

[41] A. Belloche, K. Menten, C. Comito, H. Müller, P. Schilke, J. Ott, S. Thorwirth, C. Hieret. *Astronomy & Astrophysics*, **2008**. *492*, 769–773.

[42] V. Rivilla, J. Martín-Pintado, I. Jiménez-Serra, S. Zeng, S. Martín, J. Armijos-Abendano, M. Requena-Torres, R. Aladro, D. Riquelme. *Monthly Notices of the Royal Astronomical Society: Letters*, **2019**. *483*, L114–L119.

[43] V. M. Rivilla, M. Sanz-Novo, I. Jiménez-Serra, J. Martín-Pintado, L. Colzi, S. Zeng, A. Megías, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi. *arXiv preprint arXiv:2307.11507*, **2023**.

[44] E. Herbst. *Chem. Soc. Rev.*, **2001**. *30*, 168–176.

[45] R. T. Garrod. *The Astrophysical Journal*, **2013**. *765*, 60.

[46] K. Hadraoui, H. Cottin, S. Ivanovski, P. Zapf, K. Altwegg, Y. Benilan, N. Biver, V. Della Corte, N. Fray, J. Lasue, et al. *Astronomy & Astrophysics*, **2019**. *630*, A32.

[47] S. A. Krasnokutski, C. Jäger, T. Henning. *The Astrophysical Journal*, **2020**. *889*, 67.

[48] G. M. Muñoz Caro, U. J. Meierhenrich, W. A. Schutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W. H. P. Thiemann, A. Brack, J. M. Greenberg. *Nature*, **2002**. *416*, 403–406.

[49] J. E. Elsila, J. P. Dworkin, M. P. Bernstein, M. P. Martin, S. A. Sandford. *The Astrophysical Journal*, **2007**. *660*, 911.

[50] C.-W. Lee, J.-K. Kim, E.-S. Moon, Y. C. Minh, H. Kang. *The Astrophysical Journal*, **2009**. *697*, 428.

[51] S. Maclot, D. G. Piekarski, A. Domaracka, A. Méry, V. Vizcaino, L. Adoui, F. Martín, M. Alcamí, B. A. Huber, P. Rousseau, S. Díaz-Tendero. *The Journal of Physical Chemistry Letters*, **2013**. *4*, 3903–3909.

[52] S. Maclot, D. G. Piekarski, R. Delaunay, A. Domaracka, A. Méry, V. Vizcaino, J.-Y. Chesnel, F. Martín, M. Alcamí, B. A. Huber, L. Adoui, P. Rousseau, S. Díaz-Tendero. *The European Physical Journal D*, **2014**. *68*, 149.

[53] B. Maté, Y. Rodriguez-Lazcano, Ó. Gálvez, I. Tanarro, R. Escribano. *Phys. Chem. Chem. Phys.*, **2011**. *13*, 12268–12276.

[54] K. Iijima, K. Tanaka, S. Onuma. *Journal of Molecular Structure*, **1991**. *246*, 257–266.

[55] R. M. Balabin. *Chemical Physics Letters*, **2009**. *479*, 195–200.

[56] P. D. Godfrey, R. D. Brown. *Journal of the American Chemical Society*, **1995**. *117*, 2019–2023.

[57] C. H. Hu, M. Shen, H. F. Schaefer III. *Journal of the American Chemical Society*, **1993**. *115*, 2923–2929.

[58] O. Rahaman, A. C. T. van Duin, W. A. I. Goddard, D. J. Doren. *The Journal of Physical Chemistry B*, **2011**. *115*, 249–261.

[59] B. Balta, V. Aviyente. *Journal of Computational Chemistry*, **2003**. *24*, 1789–1802.

[60] G. M. Chaban, R. B. Gerber. *The Journal of Chemical Physics*, **2001**. *115*, 1340–1348.

[61] I. Tuñón, E. Silla, M. F. Ruiz-López. *Chemical Physics Letters*, **2000**. *321*, 433–437.

[62] N. Kishimoto. *Chemical Physics Letters*, **2017**. *667*, 172–179.

[63] J. Sun, D. Bousquet, H. Forbert, D. Marx. *The Journal of Chemical Physics*, **2010**. *133*, 114508.

[64] M. Haranczyk, M. Gutowski. *The Journal of Chemical Physics*, **2008**. *128*, 125101.

[65] C. K. Kim, B.-H. Park, H. W. Lee, C. K. Kim. *Organic & Biomolecular Chemistry*, **2013**. *11*, 1407–1413.

[66] F. R. Tortonda, J. L. Pascual-Ahuir, E. Silla, I. Tuñón. *Journal of Molecular Structure: THEOCHEM*, **2003**. *623*, 203–210.

[67] R. Ramaekers, J. Pajak, B. Lambie, G. Maes. *The Journal of Chemical Physics*, **2004**. *120*, 4182–4193.

[68] M. K. Ghosh, T. H. Choi, C. H. Choi. *Theoretical Chemistry Accounts*, **2016**. *135*, 103.

[69] Q. Cui. *The Journal of Chemical Physics*, **2002**. *117*, 4720–4728.

[70] S. Tolosa, A. Hidalgo, J. A. Sansón. *Journal of Molecular Modeling*, **2014**. *20*, 2147.

[71] I. Snook, A. Barnard. *Graphene Nano-Flakes and Nano-Dots: Theory, Experiment and Applications*, **2011**.

[72] L. J. Allamandola, A. G. G. M. Tielens, J. R. Barker. , **1985**. *290*, L25–L28.

[73] A. Léger, J. L. Puget. , **1984**. *137*, L5–L8.

[74] T. S.-Y. Lai, L. Armus, M. Bianchin, T. Diaz-Santos, S. T. Linden, G. C. Privon, H. Inami, T. Bohn, A. S. Evans, K. L. Larson. *arXiv preprint arXiv:2307.15169*, **2023**.

[75] A. Simon, M. Rapacioli. *Chemical Modelling*, **2018**, 195–216.

[76] M. Rapacioli. *Modelling polycyclic aromatic hydrocarbons and their derivatives*, **2013**. 127–152.

[77] M. Rapacioli, S. Cazaux, N. Foley, A. Simon, R. Hoekstra, T. Schlathölter. *Physical Chemistry Chemical Physics*, **2018 (cover of PCCP, selected for the theme issue 2018 PCCP HOT Articles)**. *20*, 22427–22438.

[78] S. Chakraborty, G. Mulas, M. Rapacioli, C. Joblin. *Journal of Molecular Spectroscopy*, **2021**. *378*, 111466.

[79] A. Candian, C. J. Mackie. *Int. J. Quantum Chem.*, **2017**. *117*, 146–150.

[80] C. Joblin, A. G. G. M. Tielens, eds. *PAHs and the Universe: A Symposium to Celebrate the 25th Anniversary of the PAH Hypothesis*, *EAS Publications Series*, volume 46, **2011**.

[81] J. Cernicharo, A. M. Heras, A. G. G. M. Tielens, J. R. Pardo, F. Herpin, M. Guélin, L. B. F. M. Waters. *The Astrophysical Journal*, **2001**. *546*, L123–L126.

[82] B. A. McGuire, R. A. Loomis, A. M. Burkhardt, K. L. K. Lee, C. N. Shingledecker, S. B. Charnley, I. R. Cooke, M. A. Cordiner, E. Herbst, S. Kalenskii, M. A. Siebert, E. R. Willis, C. Xue, A. J. Remijan, M. C. McCarthy. *Science*, **2021**. *371*, 1265–1269.

[83] J. Cernicharo, M. Agú ndez, C. Cabezas, B. Tercero, N. Marcelino, J. R. Pardo, P. de Vicente. *Astronomy &amp Astrophysics*, **2021**. *649*, L15.

[84] A. M. Burkhardt, K. Long Kelvin Lee, P. Bryan Changala, C. N. Shingledecker, I. R. Cooke, R. A. Loomis, H. Wei, S. B. Charnley, E. Herbst, M. C. McCarthy, B. A. McGuire. *The Astrophysical Journal Letters*, **2021**. *913*, L18.

[85] J. E. Elsila, N. P. de Leon, P. R. Buseck, R. N. Zare. *Geochimica et Cosmochimica Acta*, **2005**. *69*, 1349–1357.

[86] H. Sabbah, M. Carlos, P. Jenniskens, M. H. Shaddad, J. Duprat, C. A. Goodrich, C. Joblin. *The Astrophysical Journal*, **2022**. *931*, 91.

[87] J. Cami, J. Bernard-Salas, E. Peeters, S. E. Malek. *Science*, **2010**. *329*, 1180–1182.

[88] O. Berne, G. Mulas, C. Joblin. *Astronomy & Astrophysics*, **2013**. *550*, L4.

[89] G. A. H. Walker, D. A. Bohlender, J. P. Maier, E. K. Campbell. *The Astrophysical Journal*, **2015**. *812*, L8.

[90] F. Pietrucci, W. Andreoni. *J. Chem. Theor. Comput.*, **2014**. *10*, 913–917.

[91] O. Berné, J. Montillaud, C. Joblin. *Astron. Astrophys.*, **2015**. *577*, A133.

[92] C. Joblin, J. Cernicharo. *Science*, **2018**. *359*, 156–157.

[93] I. Cherchneff, J. R. Barker, A. G. Tielens. *The Astrophysical Journal*, **1992**. *401*, 269–287.

[94] P. Pilleri, C. Joblin, F. Boulanger, T. Onaka. *Astronomy & Astrophysics*, **2015**. *577*, A16.

[95] L. Martínez, G. Santoro, P. Merino, M. Accolla, K. Lauwaet, J. Sobrado, H. Sabbah, R. J. Pelaez, V. J. Herrero, I. Tanarro, M. Agúndez, A. Martín-Jimenez, R. Otero, G. J. Ellis, C. Joblin, J. Cernicharo, J. Martín-Gago. *Nature Astronomy*, **2020**. *4*, 97–105.

[96] F. Salama, G. A. Galazutdinov, J. Krełowski, L. J. Allamandola, F. A. Musaev. , **1999**. *526*, 265–273.

[97] V. S. Reddy, S. Ghanta, S. Mahapatra. *Physical Review Letters*, **2010**. *104*, 111102–.

[98] A. Marciniak, V. Despré, T. Barillot, A. Rouzée, M. Galbraith, J. Klei, C.-H. Yang, C. Smeenk, V. Loriot, S. N. Reddy. *Nature Communications*, **2015**. *6*, 7909.

[99] A. Marciniak, V. Despré, V. Loriot, G. Karras, M. Hervé, L. Quintard, F. Catoire, C. Joblin, E. Constant, A. I. Kuleff, F. Lépine. *Nature Communications*, **2019**. *10*, 337.

[100] S. Banhatti, D. B. Rap, A. Simon, H. Leboucher, G. Wenzel, C. Joblin, B. Redlich, S. Schlemmer, S. Brünken. *Physical Chemistry Chemical Physics*, **2022**. *24*, 27343–27354.

[101] W. Karcher, R. Fordham, J. Dubois, P. Glaude, J. Ligthart. *Spectral atlas of polycyclic aromatic compounds: including data on occurrence and biological activity, spectral atlas of polycyclic aromatic compounds, d*, **1985**.

[102] P. J. Garratt. *Advanced Materials*, **1993**. *5*, 770–770.

[103] G. Malloci, C. Joblin, G. Mulas. *Chemical physics*, **2007**. *332*, 353–359.

[104] K. F. Hall, M. Boggio-Pasqua, M. J. Bearpark, M. A. Robb. *The Journal of Physical Chemistry A*, **2006**. *110*, 13591–13599.

[105] A. M. Tokmachev, M. Boggio-Pasqua, M. J. Bearpark, M. A. Robb. *The Journal of Physical Chemistry A*, **2008**. *112*, 10881–10886.

[106] S. Ghanta, V. S. Reddy, S. Mahapatra. *Physical Chemistry Chemical Physics*, **2011**. *13*, 14531–14541.

[107] S. N. Reddy, S. Mahapatra. *The Journal of Physical Chemistry A*, **2013**. *117*, 8737–8749.

[108] A. Simon, M. Rapacioli, M. Lanza, B. Joalland, F. Spiegelman. *Physical Chemistry Chemical Physics*, **2011**. *13*, 3359–3374.

[109] B. Joalland, M. Rapacioli, A. Simon, C. Joblin, C. J. Marsden, F. Spiegelman. *The Journal of Physical Chemistry A*, **2010**. *114*, 5846–5854.

[110] F. Jolibois, A. Klotz, F. Gadéa, C. Joblin. *Astronomy & Astrophysics*, **2005**. *444*, 629–634.

[111] S. R. Castillo, A. Simon, C. Joblin. *International journal of mass spectrometry*, **2018**. *429*, 189–197.

[112] H. Leboucher, J. Mascetti, C. Aupetit, J. A. Noble, A. Simon. *Photochem*, **2022**. *2*, 237–262.

[113] M. Rapacioli, A. Simon, C. C. Marshall, J. Cuny, D. Kokkin, F. Spiegelman, C. Joblin. *The Journal of Physical Chemistry A*, **2015**. *119*, 12845–12854.

[114] S. Zamith, J.-M. l'Hermite, L. Dontot, L. Zheng, M. Rapacioli, F. Spiegelman, C. Joblin. *The Journal of chemical physics*, **2020**. *153*.

[115] V. Pirronello, J. Krelowski, G. Manicò. *Solid state astrochemistry*, volume 120. Springer Science & Business Media, **2012**.

[116] *Evolution of interstellar ices*. Springer, **1999**.

[117] *Interstellar ices*, volume 414, **2009**.

[118] T. Bartels-Rausch, V. Bergeron, J. H. Cartwright, R. Escribano, J. L. Finney, H. Grothe, P. J. Gutiérrez, J. Haapala, W. F. Kuhs, J. B. Pettersson. *Reviews of Modern Physics*, **2012**. *84*, 885.

[119] T. Loerting, K. Winkel, M. Seidl, M. Bauer, C. Mitterdorfer, P. H. Handle, C. G. Salzmann, E. Mayer, J. L. Finney, D. T. Bowron. *Physical Chemistry Chemical Physics*, **2011**. *13*, 8783–8794.

[120] P. V. Hobbs. *Ice physics*. OUP Oxford, **2010**.

[121] R. S. Smith, J. Matthiesen, J. Knox, B. D. Kay. *The Journal of Physical Chemistry A*, **2011**. *115*, 5908–5917.

[122] J.-B. Bossa, K. Isokoski, M. De Valois, H. Linnartz. *Astronomy & Astrophysics*, **2012**. *545*, A82.

[123] P. Ghesquière, T. Mineva, D. Talbi, P. Theulé, J. A. Noble, T. Chiavassa. *Physical Chemistry Chemical Physics*, **2015**. *17*, 11455–11468.

[124] S. A. Sandford, M. Nuevo, P. P. Bera, T. J. Lee. *Chemical Reviews*, **2020**. *120*, 4616–4659.

[125] A. A. Correa. *Computational Materials Science*, **2018**. *150*, 291–303.

[126] H. Rothard, A. Domaracka, P. Boduch, M. E. Palumbo, G. Strazzulla, E. F. Da Silveira, E. Dartois. *Journal of Physics B: Atomic, Molecular and Optical Physics*, **2017**. *50*, 062001.

[127] A. M. Bykov, A. Marcowith, E. Amato, M. E. Kalyashova, J. M. D. Kruijssen, E. Waxman. *Space Science Reviews*, **2020**. *216*, 42.

[128] E. Herbst, R. T. Garrod. *Frontiers in Astronomy and Space Sciences*, **2022**. *8*.

[129] I. R. Cooke, I. R. Sims. *ACS Earth and Space Chemistry*, **2019**. *3*, 1109–1134.

[130] J. Renoud, S. Indrajith, A. Domaracka, P. Rousseau, P. Moretto-Capelle, B. A. Huber, J.-P. Champeaux. *Physical Chemistry Chemical Physics*, **2020**. *22*, 5785–5796.

[131] A. L. F. de Barros, A. Domaracka, D. P. P. Andrade, P. Boduch, H. Rothard, E. F. da Silveira. *Monthly Notices of the Royal Astronomical Society*, **2011**. *418*, 1363–1374.

[132] H. W. Jochims, E. Ruhl, H. Baumgartel, S. Tobita, S. Leach. *Astrophys. J.*, **1994**. *420*, 307–317.

[133] Castellanos, P., Candian, A., Zhen, J., Linnartz, H., Tielens, A. G. G. M. *A&A*, **2018**. *616*, A166.

[134] U. Jacovella, C. Rossi, C. Romanzin, C. Alcaraz, R. Thissen. *ChemPhysChem*, **2023**. *24*, e202200474.

[135] Marciniak, A., Joblin, C., Mulas, G., Mundlapati, V. Rao, Bonnamy, A. *A&A*, **2021**. *652*, A42.

[136] G. RouillÃ©, M. Steglich, P. Hemberger, C. JÃ¤ger, T. Henning. *The Astrophysical Journal*, **2019**. *885*, 21.

[137] A. Simon, J. P. Champeaux, M. Rapacioli, P. M. Capelle, F. X. Gadéa, M. Sence. *Theor Chem Acc*, **2018**. *137*, 106.

[138] J. Zhen, S. R. Castillo, C. Joblin, G. Mulas, H. Sabbah, A. Giuliani, L. Nahon, S. Martin, J.-P. Champeaux, P. M. Mayer. *The Astrophysical Journal*, **2016**. *822*, 113.

[139] R. Siebenmorgen, E. Krügel. *A&A*, **2010**. *511*, A6.

[140] R. Visser, V. C. Geers, C. P. Dullemond, J.-C. Augereau, K. M. Pontoppidan, E. F. van Dishoeck. *A&A*, **2007**. *466*, 229–241.

[141] E. R. Micelotta, A. P. Jones, A. G. G. M. Tielens. *Astron. Astrophys.*, **2010**. *510*, A36.

[142] E. R. Micelotta, A. P. Jones, A. G. G. M. Tielens. *Astron. Astrophys.*, **2010**. *510*, A37.

[143] A. Marciniak, V. Despré, T. Barillot, A. Rouzée, M. C. E. Galbraith, J. Klei, C. H. Yang, C. T. L. Smeenk, V. Loriot, S. N. Reddy, A. G. G. M. Tielens, S. Mahapatra, A. I. Kuleff, M. J. J. Vrakking, F. Lépine. *Nature Communications*, **2015**. *6*, 7909.

[144] J. A. Noble, C. Aupetit, D. Descamps, S. Petit, A. Simon, J. Mascetti, N. Ben Amor, V. Blanchet. *Phys. Chem. Chem. Phys.*, **2019**. *21*, 14111–14125.

[145] Y. Garniron. *Development and parallel implementation of selected configuration interaction methods*, **2018**.

[146] M. Nagaoka, N. Okuyama-Yoshida, T. Yamabe. *The Journal of Physical Chemistry A*, **1998**. *102*, 8202–8208.

[147] H. Miyamoto, M. Aida. *Chemistry Letters*, **2013**. *42*, 598–600.

[148] A. Koster, G. Geudtner, A. Alvarez-Ibarra, P. Calaminici, M. Casida, J. Carmona-Espindola, V. Dominguez, R. Flores-Moreno, G. Gamboa, A. Goursot, T. Heine, A. Ipatov, A. de la Lande, F. Janetzko, J. del Campo, D. Mejia-Rodriguez, J. U. Reveles, J. Vasquez-Perez, A. Vela, B. Zuniga-Gutierrez, D. Salahub. *deMon2k Version 6*, **2018**.

[149] T. Heine, M. Rapacioli, S. Patchkovskii, J. Frenzel, A. Koster, P. Calaminici, H. A. Duarte, S. Escalante, R. Flores-Moreno, A. Goursot, J. Reveles, D. Salahub, A. Vela. *deMonNano, http:demon-nano.ups-tlse.fr*, **2009**.

[150] I. N. Levine, D. H. Busch, H. Shull. *Quantum chemistry*, volume 6. Pearson Prentice Hall Upper Saddle River, NJ, **2013**.

[151] F. Jensen. *Introduction to computational chemistry.* John wiley & sons, **2017**.

[152] A. Szabo, N. S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory.* Dover Edition, **1996**.

[153] F. Louisnard. *Exploration of energetic landscapes and nuclear quantum effects : a parallel-tempering path-integral molecular dynamics approach*, **2022**.

[154] J. C. Slater. *Phys. Rev.*, **1929**. *34*, 1293–1322.

[155] V. Fock. *Z Angew Phys*, **1930**. *61*, 126–148.

[156] R. J. Bartlett, D. M. Silver. *International Journal of Quantum Chemistry*, **1976**. *10*, 185–186.

[157] J. S. Binkley, J. A. Pople. *International Journal of Quantum Chemistry*, **1975**. *9*, 229–236.

[158] *Int. J. Quantum Chem.*, volume 10, **1976**.

[159] J. Čížek. *The Journal of Chemical Physics*, **2004**. *45*, 4256–4266.

[160] O. Sinanoğlu. *The Journal of Chemical Physics*, **1962**. *36*, 706–717.

[161] R. J. Bartlett. *The Journal of Physical Chemistry*, **1989**. *93*, 1697–1708.

[162] R. J. Bartlett. *Coupled-Cluster Theory: An Overview of Recent Developments.* Yarkony, D. R., World Scientific, Singapore, **1995, p 1047-1131**.

[163] P. Hohenberg, W. Kohn. *Phys. Rev.*, **1964**. *136*, B864–B871.

[164] C. Adamo. *J. Chem. Phys*, **1999**, 6158–6170.

[165] R. Colle, O. Salvetti. *Theoretica chimica acta*, **1975**. *37*, 329–334.

[166] R. Colle, O. Salvetti. *The Journal of chemical physics*, **1983**. *79*, 1404–1407.

[167] C. Lee, W. Yang, R. G. Parr. *Phys. Rev. B*, **1988**. *37*, 785–789.

[168] J. P. Perdew, M. Ernzerhof, K. Burke. *The Journal of chemical physics*, **1996**. *105*, 9982–9985.

[169] A. de la Lande, A. Alvarez-Ibarra, K. Hasnaoui, F. Cailliez, X. Wu, T. Mineva, J. Cuny, P. Calaminici, L. López-Sosa, G. Geudtner, I. Navizet, C. Garcia Iriepa, D. R. Salahub, A. M. Köster. *Molecules*, **2019**. *24*.

[170] B. I. Dunlap, J. Connolly, J. Sabin. *The Journal of Chemical Physics*, **1979**. *71*, 3396–3402.

[171] J. Mintmire, B. Dunlap. *Physical Review A*, **1982**. *25*, 88.

[172] J. Mintmire, J. Sabin, S. Trickey. *Physical Review B*, **1982**. *26*, 1743.

[173] M. E. Casida. *Time-Dependent Density Functional Response Theory for Molecules*, **1995**. 155–192.

[174] E. Runge, E. K. U. Gross. *Phys. Rev. Lett.*, **1984**. *52*, 997–1000.

[175] J. Jokar, C. Honerkamp, N. Helbig. *Adiabatic approximations within time-dependent density functional theory for the non-linear regime.* Technical report, Fachgruppe Physik, **2017**.

[176] B. F. Habenicht, N. P. Tani, M. R. Provorse, C. M. Isborn. *The Journal of Chemical Physics*, **2014**. *141*, 184112.

[177] M. E. Casida. *J. Mol. Struct. : Theochem*, **2009**. *914*, 3–18.

[178] K. A. Omar, F. A. Korsaye, R. Tandiana, D. Tolu, J. Deviers, X. Wu, A. Parise, A. Alvarez-Ibarra, F. Moncada, J. N. Pedroza-Montero. *The European Physical Journal Special Topics*, **2023**, 1–27.

[179] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, R. Kaschner. *Physical Review B*, **1995**. *51*, 12947.

[180] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert. *Phys. Rev. B*, **1998**. *58*, 7260–7268.

[181] M. Elstner, G. Seifert. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **2014**. *372*, 20120483.

[182] G. Seifert, D. Porezag, T. Frauenheim. *Int. J. Quantum Chem.*, **1996**. *58*, 185–192.

[183] P. Koskinen, V. Makinen. *Comput. Mat. Sc.*, **2009**. *47*, 237–253.

[184] R. Pariser. *The Journal of Chemical Physics*, **1956**. *24*, 250–268.

[185] B. Aradi, B. Hourahine, T. Frauenheim. *A*, **2007**. *111*, 5678–5684. PMID: 17567110.

[186] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler. *J. Comput. Chem.*, **2001**. *22*, 931–967.

[187] R. Salomon-Ferrer, D. A. Case, R. C. Walker. *WIREs Comput Mol Sci*, **2013**. *3*, 198–210.

[188] H. Berendsen, D. van der Spoel, R. van Drunen. *Comput. Phys. Commun.*, **1995**. *91*, 43 – 56.

[189] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox. *Gaussian~16 Revision C.01*, **2016**. Gaussian Inc. Wallingford CT.

[190] A. Humeniuk, R. Mitrić. *Comput. Phys. Commun.*, **2017**. *221*, 174–202.

[191] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele. *WIREs Comput. Mol. Sci.*, **2014**. *4*, 15–25.

[192] M. Rapacioli, F. Spiegelman, D. Talbi, T. Mineva, A. Goursot, T. Heine, G. Seifert. *The Journal of Chemical Physics*, **2009**. *130*, 244304.

[193] A. E. Reed, L. A. Curtiss, F. Weinhold. *Chem. Rev.*, **1988**. *88*, 899–926.

[194] R. F. Bader. *Atoms in Molecules: a Quantum Theory*. Oxford University Press, **1990**.

[195] U. C. Singh, P. A. Kollman. *Journal of computational chemistry*, **1984**. *5*, 129–145.

[196] B. H. Besler, K. M. Merz Jr, P. A. Kollman. *Journal of computational chemistry*, **1990**. *11*, 431–439.

[197] J. Kalinowski, B. Lesyng, J. Thompson, C. Cramer, D. Truhlar. *A*, **2004**. *108*, 2545–2549.

[198] A. J. Bridgeman, G. Cavigliasso, L. R. Ireland, J. Rothery. *Journal of the Chemical Society, Dalton Transactions*, **2001**, 2095–2108.

[199] E. Michoulier, N. B. Amor, M. Rapacioli, J. A. Noble, J. Mascetti, C. Toubin, A. Simon. *Physical Chemistry Chemical Physics*, **2018**. *20*, 11941–11953.

[200] B. Hourahine, B. Aradi, T. Frauenheim. *J. Phys. Conf. Ser.*, **2010**. *242*, 012005.

[201] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, T. Frauenheim. *The Journal of Chemical Physics*, **2020**. *152*, 124101.

[202] T. A. Niehaus, S. Suhai, F. Della Sala, P. Lugli, M. Elstner, G. Seifert, T. Frauenheim. *B*, **2001**. *63*, 085108–9.

[203] A. Humeniuk, R. Mitrić. , **2015**. *143*, 134120.

[204] V. Lutsker, B. Aradi, T. A. Niehaus. , **2015**. *143*, 184107.

[205] M. P. Kroonblawd, F. Pietrucci, A. M. Saitta, N. Goldman. *Journal of Chemical Theory and Computation*, **2018**. *14*, 2207–2218.

[206] G. Groenhof. *Biomolecular simulations: methods and protocols*, **2013**, 43–66.

[207] M. A. González. *JDN*, **2011**. *12*, 169–200.

[208] L. Meinhold. *Crystalline protein dynamics: A simulation analysis of staphylococcal nuclease*, **2005**.

[209] W. L. Jorgensen. , **1986**. *90*, 1276–1284.

[210] J. O. Jensen, P. N. Krishnan, L. A. Burke. *Chem. Phys. Lett.*, **1995**. *246*, 13–19.

[211] L. Zheng. *Structure, solvation, thermodynamics and fragmentation of molecular clusters*, **2021**.

[212] C. G. BROYDEN. *IMA Journal of Applied Mathematics*, **1970**. *6*, 76–90.

[213] R. Fletcher. *The Computer Journal*, **1970**. *13*, 317–322.

[214] D. Goldfarb. *Mathematics of Computation*, **1970**. *24*, 23–26.

[215] D. F. Shanno. *Mathematics of Computation*, **1970**. *24*, 647–656.

[216] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. *The Journal of Chemical Physics*, **1953**. *21*, 1087–1092.

[217] Z. Li, H. A. Scheraga. *Proc. Natl. Acad. Sci.*, **1987**. *84*, 6611–6615.

[218] Y. Sugita, Y. Okamoto. *Chem. Phys. Lett.*, **1999**. *314*, 141–151.

[219] Y. Sugita, Y. Okamoto. *Chem. Phys. Lett.*, **2000**. *329*, 261–270.

[220] J. C. Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, **2016**.

[221] J. Butcher. *Numerical Analysis and Optimization: NAO-III, Muscat, Oman, January 2014*, **2015**, 37–58.

[222] W. Kutta. *Z. Math. Phys*, **1901**. *46*, 435–453.

[223] K. Diethelm, N. J. Ford, A. D. Freed. *Nonlinear Dynamics*, **2002**. *29*, 3–22.

[224] L. Verlet. *Physical review*, **1967**. *159*, 98.

[225] F. Fraige, P. Langston. *Advanced Powder Technology*, **2004**. *15*, 227–245.

[226] N. S. Martys, R. D. Mountain. *Physical Review E*, **1999**. *59*, 3733.

[227] R. Kosloff. *Annual review of physical chemistry*, **1994**. *45*, 145–178.

[228] M. Barbatti, M. Ruckenbauer, F. Plasser, J. Pittner, G. Granucci, M. Persico, H. Lischka. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2014**. *4*, 26–33.

[229] R. Crespo-Otero, M. Barbatti. *Chemical reviews*, **2018**. *118*, 7026–7068.

[230] A. Warshel, M. Karplus. *Journal of the American Chemical Society*, **1972**. *94*, 5612–5625.

[231] A. Warshel, M. Levitt. *Journal of Molecular Biology*, **1976**. *103*, 227–249.

[232] M. J. Field, P. A. Bash, M. Karplus. *Journal of Computational Chemistry*, **1990**. *11*, 700–733.

[233] F. Maseras, K. Morokuma. *Journal of Computational Chemistry*, **1995**. *16*, 1170–1179.

[234] U. Ryde. *Journal of computer-aided molecular design*, **1996**. *10*, 153–164.

[235] M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, K. Morokuma. *The Journal of Physical Chemistry*, **1996**. *100*, 19357–19363.

[236] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner. *Journal of Molecular Structure: THEOCHEM*, **2003**. *632*, 1–28.

[237] H. M. Senn, W. Thiel. *Angewandte Chemie International Edition*, **2009**. *48*, 1198–1229.

[238] G. Lamoureux, B. Roux. *The Journal of chemical physics*, **2003**. *119*, 3025–3039.

[239] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, M. H. Olsson. *Chemical reviews*, **2006**. *106*, 3210–3235.

[240] A. K. Rappe, W. A. Goddard III. *The Journal of Physical Chemistry*, **1991**. *95*, 3358–3363.

[241] X. Assfeld, J.-L. Rivail. *Chemical physics letters*, **1996**. *263*, 100–106.

[242] J. Gao, P. Amara, C. Alhambra, M. J. Field. *The Journal of Physical Chemistry A*, **1998**. *102*, 4714–4721.

[243] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Aprà, T. L. Windus. *Computer Physics Communications*, **2010**. *181*, 1477–1489.

[244] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng. *Molecular Physics*, **2015**. *113*, 184–215.

[245] M. e. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson. *Gaussian 09 (Revision A. 02)*, **2009**.

[246] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus. *B*, **2001**. *105*, 569–585.

[247] G. Hou, X. Zhu, Q. Cui. *Journal of Chemical Theory and Computation*, **2010**. *6*, 2303–2314. PMID: 20711513.

[248] Y. Nishimoto. *A*, **2016**. *120*, 771–784.

[249] T. Kubař, P. B. Woiczikowski, G. Cuniberti, M. Elstner. *B*, **2008**. *112*, 7937–7947.

[250] M. Lundberg, Y. Sasakura, G. Zheng, K. Morokuma. *J. Chem. Theo. Comp.*, **2010**. *6*, 1413–1427.

[251] A. de La Lande, A. Alvarez-Ibarra, K. Hasnaoui, F. Cailliez, X. Wu, T. Mineva, J. Cuny, P. Calaminici, L. López-Sosa, G. Geudtner. *Molecules*, **2019**. *24*, 1653.

[252] X. Wu, J.-M. Teuler, F. Cailliez, C. Clavaguéra, D. R. Salahub, A. de la Lande. *Journal of Chemical Theory and Computation*, **2017**. *13*, 3985–4002.

[253] P. Cieplak, F.-Y. Dupradeau, Y. Duan, J. Wang. *Journal of Physics: Condensed Matter*, **2009**. *21*, 333102.

[254] T. Mineva, N. Russo. *International journal of quantum chemistry*, **1997**. *61*, 665–671.

[255] P. Amara, M. J. Field. *Theoretical Chemistry Accounts*, **2003**. *109*, 43–52.

[256] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, M. Hodoscek. *International Journal of Quantum Chemistry*, **1996**. *60*, 1189–1200.

[257] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren, J. W. Ponder. *Journal of chemical theory and computation*, **2018**. *14*, 5273–5289.

[258] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case. *Journal of computational chemistry*, **2004**. *25*, 1157–1174.

[259] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami. , **2004**. *120*, 647–659.

[260] A. Alvarez-Ibarra, P. Calaminici, A. Goursot, C. Gómez-Castro, R. Grande-Aztatzi, T. Mineva, D. R. Salahub, J. Vásquez-Pérez, A. Vela, B. Zuniga-Gutierrez. *First Principles Computational Biochemistry with deMon2k*. Elsevier, **2015**. 281–325.

[261] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch. *Journal of computational chemistry*, **2009**. *30*, 1545–1614.

[262] G. F. d. Lima, T. Heine, H. A. Duarte, J. R. Sabin, E. Brändas. *Chapter 5 - Molecular Dynamics of Polypeptides and Their Inclusion Compounds with Cyclodextrin in Aqueous Solution Using DC–SCC–DFTB/UFF Approach*, volume 59. Academic Press, **2010**. 145–180.

[263] C. Iftner, A. Simon, K. Korchagina, M. Rapacioli, F. Spiegelman. *J. Chem. Phys.*, **2014**. *140*, 034301.

[264] G. de M Seabra, R. C. Walker, M. Elstner, D. A. Case, A. E. Roitberg. *The journal of physical chemistry. A*, **2007**. *111*, 5655–5664.

[265] A. W. Götz, M. A. Clark, R. C. Walker. *Journal of computational chemistry*, **2014**. *35*, 95–108.

[266] J. Torras, B. P. Roberts, G. M. Seabra, S. B. Trickey. *Advances in Protein Chemistry and Structural Biology*, **2015**. *100*, 1–31.

[267] S. Metz, J. Kästner, A. A. Sokol, T. W. Keal, P. Sherwood. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2014**. *4*, 101–110.

[268] H. Lin, Y. Zhang, S. Pezeshki, B. Wang, X.-P. Wu, L. Gagliardi, D. Truhlar. *University of Minnesota: Minneapolis, MN, USA*, **2018**.

[269] E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J. Piquemal, G. Andrés Cisneros. *Journal of computational chemistry*, **2016**. *37*, 1019–1029.

[270] N. Gillet, M. Elstner, T. Kubař. *The Journal of Chemical Physics*, **2018**. *149*, 072328.

[271] G. Groenhof. *Introduction to QM/MM Simulations*. Humana Press, Totowa, NJ, **2013**. 43–66.

[272] S. Nosé. *J. Chem. Phys.*, **1984**. *81*, 511.

[273] W. G. Hoover. *Phys. Rev. A*, **1985**. *31*, 1695–1697.

[274] A. F. Oliveira, P. Philipsen, T. Heine. *J. Chem. Theo. Comp.*, **2015**. *11*, 5209–5218. PMID: 26574316.

[275] J. P. Perdew, K. Burke, M. Ernzerhof. *Physical Review Letters*, **1996**. *77*, 3865–3868.

[276] N. Godbout, D. R. Salahub, J. Andzelm, E. Wimmer. *Canadian Journal of Chemistry*, **1992**. *70*, 560–571.

[277] A. M. Köster, J. U. Reveles, J. M. del Campo. *The Journal of Chemical Physics*, **2004**. *121*, 3417–3424.

[278] V. Gutmann, G. Resch. *Lecture Notes on Solution Chemistry*, volume p.97. World Scientific Co. Pte.Ltd, **1995**.

[279] T. M. Miller, B. Bederson, D. R. Bates. *Atomic and Molecular Polarizabilities-A Review of Recent Advances*, volume 13. Academic Press, **1978**. 1–55.

[280] T. N. Olney, N. M. Cann, G. Cooper, C. E. Brion. *Chemical Physics*, **1997**. *223*, 59–98.

[281] Y. ZHAI, Y. L. ZHAO. *Journal of Theoretical and Computational Chemistry*, **2012**. *12*, 1350019.

[282] S. M. Bachrach. *The Journal of Physical Chemistry A*, **2008**. *112*, 3722–3730.

[283] S. Tolosa, A. Hidalgo, J. A. Sansón. *The Journal of Physical Chemistry B*, **2012**. *116*, 13033–13044.

[284] D. T. Maysa Yusef Buey, Tzonka Mineva, M. Rapacioli. *Physical Chemistry Chemical Physics*, **2023**.

[285] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen. *Journal of Computational Chemistry*, **2005**. *26*, 1701–1718.

[286] E. Kassab, J. Langlet, E. Evleth, Y. Akacem. *Journal of Molecular Structure: THEOCHEM*, **2000**. *531*, 267–282.

[287] P. Bandyopadhyay, M. S. Gordon, B. Mennucci, J. Tomasi. *The Journal of Chemical Physics*, **2002**. *116*, 5023–5032.

[288] R. Flores-Moreno, J. V. Ortiz. *The Journal of Chemical Physics*, **2009**. *131*, 124110.

[289] A. Rimola, N. Balucani, C. Ceccarelli, P. Ugliengo. *International Journal of Molecular Sciences*, **2022**. *23*.

[290] W. Xu, Q. Zhu, C. T. Hu. *Angewandte Chemie International Edition*, **2017**. *56*, 2030–2034.

[291] A. Goursot, T. Mineva, R. Kevorkyants, D. Talbi. *Journal of Chemical Theory and Computation*, **2007**. *3*, 755–763.

[292] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling. *Proteins: Structure, Function, and Bioinformatics*, **2006**. *65*, 712–725.

[293] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *Journal of the American Chemical Society*, **1995**. *117*, 5179–5197.

[294] W. S. Ross, C. C. Hardin. *Journal of the American Chemical Society*, **1994**. *116*, 6070–6080.

[295] J. Åqvist. *The Journal of Physical Chemistry*, **1990**. *94*, 8021–8024.

[296] C.-K. Kim, B.-H. Park, H.-W. Lee, C.-K. Kim. *Bulletin of the Korean Chemical Society*, **2011**. *32*, 1985–1992.

[297] S. Pilling, L. A. Mendes, V. Bordalo, C. F. Guaman, C. R. Ponciano, E. F. da Silveira. *Astrobiology*, **2013**. *13*, 79–91.

[298] D. Guerra, L. A. Gomez, A. Restrepo, J. David. *Chemical Physics*, **2020**. *530*, 110645.

[299] L. Wang, S. D. Fried, S. G. Boxer, T. E. Markland. *Proceedings of the National Academy of Sciences*, **2014**. *111*, 18454–18459.

[300] F. Louisnard, T. Mineva, J. Cuny. *Theoretical Chemistry Accounts*, **2022**. *141*, 36.

[301] J. Cuny, J. Cerda Calatayud, N. Ansari, A. A. Hassanali, M. Rapacioli, A. Simon. *The Journal of Physical Chemistry B*, **2020**. *124*, 7421–7432.

[302] N. Cinq, A. Simon, F. Louisnard, J. Cuny. *The Journal of Physical Chemistry B*, **2023**. *127*, 7590–7601.

[303] D. Marx, M. Parrinello. *Zeitschrift für Physik B Condensed Matter*, **1994**. *95*, 143–144.

[304] S. Habershon, B. J. Braams, D. E. Manolopoulos. *The Journal of chemical physics*, **2007**. *127*.

[305] R. Tripathi, L. Durán Caballero, R. Pérez de Tudela, C. Hölzl, D. Marx. *ACS omega*, **2021**. *6*, 12676–12683.

[306] H. Leboucher, A. Simon, M. Rapacioli. *The Journal of Chemical Physics*, **2023**. *158*, 114308.

[307] L. Zheng, J. Cuny, S. Zamith, J.-M. l'Hermite, M. Rapacioli. *Physical Chemistry Chemical Physics*, **2021**. *23*, 27404–27416.

[308] P. M. Dinh, M. Vincendon, F. Coppens, E. Suraud, P. G. Reinhard. *Computer Physics Communications*, **2022**. *270*, 108155.

[309] G. Kresse, J. Furthmüller, J. Hafner. *Physical Review B*, **1994**. *50*, 13181.

[310] G. Kresse, D. Joubert. *Physical review b*, **1999**. *59*, 1758.

[311] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo. *Journal of physics: Condensed matter*, **2009**. *21*, 395502.

[312] R. Rüger, E. van Lenthe, Y. Lu, J. Frenzel, T. Heine, L. Visscher. *J. Chem. Theo. Comp.*, **2015**. *11*, 157–167. PMID: 26574214.

[313] F. Spiegelman, N. Tarrat, J. Cuny, L. Dontot, E. Posenitskiy, C. Martí, A. Simon, M. Rapacioli. *Advances in Physics: X*, **2020**. *5*, 1710252.

[314] X. Li, S. M. Smith, A. N. Markevitch, D. A. Romanov, R. J. Levis, H. B. Schlegel. *Physical Chemistry Chemical Physics*, **2005**. *7*, 233–239.

[315] R. F. Bader. *Chemical Reviews*, **1991**. *91*, 893–928.

[316] A. Castro, M. A. Marques, A. Rubio. *The Journal of chemical physics*, **2004**. *121*, 3425–3433.

[317] A. Gomez Pueyo, M. A. Marques, A. Rubio, A. Castro. *Journal of chemical theory and computation*, **2018**. *14*, 3040–3052.

[318] A. Alvarez-Ibarra, K. A. Omar, K. Hasnaoui, A. de La Lande. *Electron and Molecular Dynamics Simulations with Polarizable Embedding.* Royal Society of Chemistry, **2021**. 117–143.

[319] W. Magnus. *Communications on Pure and Applied Mathematics*, **1954**. *7*, 649–673.

[320] C.-L. Cheng, J. S. Evans, T. Van Voorhis. *Physical Review B*, **2006**. *74*, 155112.

[321] G. Malloci, G. Mulas, C. Joblin. *Astronomy & Astrophysics*, **2004**. *426*, 105–117.

[322] K. Yabana, G. Bertsch. *Physical Review B*, **1996**. *54*, 4484.

[323] R. Sinha-Roy, P. García-González, X. López Lozano, R. L. Whetten, H.-C. Weissker. *Journal of Chemical Theory and Computation*, **2018**. *14*, 6417–6426.

[324] P. A. M. Dirac. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **1927**. *114*, 243–265.

[325] I. N. Sneddon, E. R. Cohen. *Special functions of mathematical physics and chemistry*, **1956**.

[326] M. Walter, H. Häkkinen, L. Lehtovaara, M. Puska, J. Enkovaara, C. Rostgaard, J. J. Mortensen. *The journal of chemical physics*, **2008**. *128*.

[327] F. P. Bonafé, B. Aradi, B. Hourahine, C. R. Medrano, F. J. Hernández, T. Frauenheim, C. G. Sánchez. *Journal of Chemical Theory and Computation*, **2020**. *16*, 4454–4469.

[328] A. de la Lande, S. Denisov, M. Mostafavi. *Phys. Chem. Chem. Phys.*, **2021**. *23*, 21148–21162.

[329] A. Alvarez-Ibarra, A. Parise, K. Hasnaoui, A. de la Lande. *Phys. Chem. Chem. Phys.*, **2020**. *22*, 7747–7758.

[330] B. Gu, D. Muñoz-Santiburcio, F. Da Pieve, F. Cleri, E. Artacho, J. Kohanoff. *Radiation Physics and Chemistry*, **2022**. *193*, 109961.

[331] S. Hammes-Schiffer, J. C. Tully. , **1994**. *101*, 4657–4667.

[332] T. Sommerfeld, M. Ehara. *Journal of Chemical Theory and Computation*, **2015**. *11*, 4627–4633.

[333] M. Meziane. *Étude de la dynamique électronique ultra-rapide suivant l'ionisation de la molécule de caféine par la méthode td-dftb*, **2019**.

[334] M. P. Kroonblawd, R. K. Lindsey, N. Goldman. *Chemical science*, **2019**. *10*, 6091–6098.