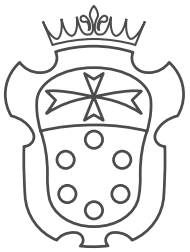

SCUOLA
NORMALE
SUPERIORE

Classe di Scienze

Corso di perfezionamento in Astrochimica

XXXIV ciclo

Development of computational strategies to simulate structures, reactivity and spectroscopic features of molecules at the interstellar ice interface

Settore Scientifico Disciplinare **CHIM/02**

Candidata

Dr.ssa Carmen Baiano

Relatori

Prof. Nicola Tasinato
Prof. Vincenzo Barone

Anno accademico 2022–2023

To the serendipity of being born into the fortunate corner of the world.

To the privilege of nurturing dreams, the freedom to pursue them, and the power to reshape them.

In profound gratitude to the resilient women who have paved the way for my privileges

and to those who will continue the struggle, ensuring that a few people's privilege evolves into the collective freedom of all.

Abstract

Astrochemistry is a multidisciplinary field that involves the study of molecules in space and their formation processes. It requires collaboration between scientists with expertise in astronomy, physics, chemistry, biology, and geology to connect the physical conditions of specific regions in space with the chemistry that takes place there.

Molecules in space are detected through their spectroscopic signatures, and there are several active telescopes that scan the Universe to capture signals from new molecules.

While astronomical observations provide valuable information, laboratory experiments and computational simulations are crucial for understanding the presence of certain molecules in space and explaining their formation. Experimental studies under extreme low temperatures and densities, similar to those found in the interstellar medium (ISM), are challenging. However, computational astrochemistry plays a key role in assisting and guiding both experiments and astronomical observations. Computational simulations provide atomistic perspectives that are difficult to achieve experimentally, offering insights into reaction pathways and spectroscopic fingerprints.

The physical conditions of the ISM impose severe constraints on chemical reactivity. Gas-phase reactions dominate in the low-density regions of the ISM, where collisions between atoms and molecules are infrequent. In this context, we re-investigated the formation process of ethanimine, a prebiotic molecule, in gas-phase. We employed composite schemes for both energy and structural

parameters in order to build an accurate potential energy surface (PES) starting from C_2H_5 and NH as precursors. Rigorous quantum chemical (QC) and kinetics calculations allow for a reliable description of the chemical process. However, the predicted ratio of the two isomers of the ethanimine still needs to be improved with respect to observations. A possible explanation comes from the consideration of reactivity occurring on the surface of interstellar dust grains. About 1 % of the ISM material is made of grains and dust covered by icy mantles at the low temperature of the ISM. Despite the low amount, solid state reactions cannot be overlooked. Ice surfaces can play a pivotal role in the formation of interstellar complex organic molecules (iCOMs). For this reason, we developed a multiscale computational approach to account for the solid icy matrix while using accurate methods for the chemical relevant zone. We tested the accuracy of such strategy for the isomerization of HCN by showing how a realistic modelling of the solid environment is mandatory for a deep understanding of the chemical processes mediated by interstellar ices. We extended this strategy to the simulation of infrared (IR) fingerprints of iCOMs adsorbed on ice.

Astronomical observations of ice spectroscopic fingerprints suggest amorphous structures to be dominant. Laboratory experiments point to low density amorphous (LDA) ice as the most probable structure to be formed in such extreme environment. By using both QC and molecular dynamics (MD) simulations we were able to get reliable models of both crystalline ice (CI) and amorphous solid water (ASW). We studied the adsorption of acetic acid (AA), methyl formate (MF) and glycolaldehyde (GA) on top of both CI and ASW and used the multiscale strategy here developed for the proper simulation of the carbonyl stretching frequency. By employing second-order vibrational perturbation theory (VPT2) and using a reduced dimensionality (RD) approach, we were able to get accurate spectroscopic information on the investigated systems in order to provide insights on the effect of ice morphology on the spectroscopic outcomes and hence on the adsorption mode.

Since ices grow on top of mineral grains and to include the effect of the under-

neath mineral we performed MD analysis of the ice morphology while growing on top of silica models of the interstellar grains. With our approach we are able to get reliable information on the accretion of LDA by providing further insights on ice morphology and structure in the ISM.

Summarizing, the research carried out during this PhD thesis aimed at developing reliable and robust yet feasible computational protocols to realistically describe interstellar ices and the chemical processing occurring on their surfaces. The developed methodology builds on purposely tailored multiscale strategies and on the integration between molecular dynamics simulations and state-of-the-art quantum chemical calculations. Application to selected paradigmatic cases demonstrates the reliability of the computational protocols and paves the route toward accurate simulations of ice-mediated chemistry in the ISM.

Contents

Glossary

AE	Absolute Error
ALMA	Atacama Large Millimeter/submillimeter Array
AO	Atomic Orbital
ASW	Amorphous Solid Water
BJ	Becke-Johnson
BOA	Born-Oppenheimer Approximation
BSSE	Basis Set Superposition Error
CBS	Complete Basis Set
CC	Couple Cluster
ccCA	correlation consistent Composite Approach
CCSD	Coupled Cluster with Single and Double excitations
CCSD(T)	Coupled-Cluster with Single and Double and Perturbative Triple excitations
CCSD(TQ)	Coupled-Cluster with Single and Double and Perturbative Triple and Quadruple excitations
CCSDT	Coupled Cluster with Single, Double and Triple excitations
CCSDT(Q)	Coupled-Cluster with Single, Double and Triple and Perturbative Quadruple excitations

CCSDTQ	Coupled Cluster with Single, Double, Triple and Quadruple excitations
CGenFF	CHARMM Generic Force Field
ChS	Cheap Scheme
CI	Configuration Interaction. Also used for Crystalline Ice
CSF	Configuration State Functions
DBOC	Diagonal Born–Oppenheimer Correction
DF	Density Functional
DFT	Density Functional Theory
DOS	Density Of States
DVPT2	Deperturbed second-order Vibrational Perturbation Theory
FDP	Feller–Peterson–Dixon
FF	Force Field
FR	Fermi Resonance
GAFF	General Amber Force Field
GGA	Generalized Gradient Approximation
GMC	Giant Molecular Cloud
GTO	Gaussian-Type Orbitals
GVPT2	Generalized second-order Vibrational Perturbation Theory
HDA	High Density Amorphous
HEAT	High-accuracy Extrapolated Ab-initio Thermochemistry
HF	Hartree-Fock
iCOM	interstellar Complex Organic Molecule

IFF Interface Force Field

IOM Insoluble Organic Matter

IR InfraRed

IRAM Institute for Radio Astronomy in the Millimeter Range

ISM Interstellar Space Medium

ISO Infrared Space Observatory

jun-ChS Cheap Scheme using the jun-cc-pVnZ basis set

JWST James Webb Space Telescope

KS Kohn-Sham

LAM Large Amplitude Mode

LDA Low Density Amorphous

LSDA Local Spin-Density Approximation

MAE Mean Absolute Error

MBPT Many Body Perturbation Theory

MC Molecular Cloud

MD Molecular Dynamics

MDA Medium Density Amorphous

MIES Metastable Induced Electron Spectroscopy

MM Molecular Mechanics

MO Molecular Orbital

MPn Møller-Plesset, n order of perturbation

MP2 Second Order Møller-Plesset

NOEMA NOrthern Extended Millimeter Array

ONIOM	Our own N-layered Integrated molecular Orbital and Molecular mechanics
oto	orientational tetrahedral order
PAH	Polycyclic Aromatic Hydrocarbon
PBC	Periodic Boundary Conditions
PCM	Polarizable Continuum Model
PES	Potential Energy Surface
PST	Phase Space Theory
PW	Plane Wave
QC	Quantum Chemical
QM	Quantum Mechanical
RD	Reduced Dimensionality
RDF	Radial Distribution Function
RE	Relative Error
RMSD	Root Mean Square Displacement
RRKM	Rice–Rampsberger–Kassel–Marcus
SCF	Self Consistent Field
SE	Schrödinger Equation
Sgr-B2	Sagittarius B2
SIE	Self Interaction Error
STO	Slater-Type Orbitals
TFD	Thomas-Fermi-Dirac

TPD Thermal Programmed Desorption

TST Transition State Theory

UIR Unassigned Infrared

VACF Velocity Autocorrelation Function

VMD Visual Molecular Dynamics

VPT Vibrational Perturbation Theory

VPT2 Vibrational Second Order Perturbation Theory

VTST Variational Transition State Theory

Wn Weizmann-n

ZPE Zero-point energy

ZPVE Zero-point vibrational energy

Chapter 1

Introduction

Humans have always turned their gaze at the sky above their heads wondering about the mysteries behind that fascinating scenery. Today people live in a space station orbiting around the Earth, telescopes collect stunning pictures of galaxies and the black hole at the center of our Galaxy has been recently photographed [1]. Part of the invisible to previous generations has been revealed. The more we see, the more curious we get about the still unexplored Universe and new questions feed our scientific curiosity. We are living in a time where affordable space travels and colonization of new planets are a realistic scenario for a foreseeable future rather than the plot of a sci-fi movie. The even more fundamental question is *"where did the life, as we know it, come from?"*. Though fascinating, this question requires a former definition of life and an understanding of the chemical complexity that brought to the first appearance of primordial life bricks on our planet [2]. Whether this was triggered by molecule delivered from the outer Universe or by the unique physical and chemical properties of the early Earth is still matter of debate [3,4]. For ages the main focus of space science has been the understanding of the physics of the Universe. It was customary to consider the space between stars as a vacant area, not particularly attractive in terms of chemical reactivity. The first record of an interstellar medium (ISM) between the stars dates back to the early 20th century [5,6]. The ISM is a very hostile

environment since temperatures and densities range between ~ 10 and ~ 10^6 K and ~ 10^{-4} to ~ 10^8 particles/cm^3, respectively. Due to such extreme conditions, any process involving atoms aggregation and reactivity seemed impracticable. However, with the discovery of the first molecules a new idea was advanced that, despite the harsh conditions of the ISM, chemistry was active. Understanding how molecules form and react throughout the Universe, why some molecules are detected in specific regions and how they can react in such complex chemical environment are the key questions addressed by astrochemistry [7–9]. In 1937, methylidyne radical ($CH^\bullet$) was one of the first molecule to be detected in the ISM through its strong electronic transition in the optical region [10]. As for methylidyne, $OH^\bullet$ radical has a $^2\Pi$ ground state. Despite this analogy, the optical lines of $OH^\bullet$ escaped the detection because of the interference of the ozone absorption bands in our own atmosphere. Only in the early 60s, with the advent of radio-astronomy, the first lines of $OH^\bullet$ were observed [11]. The radio observation of $OH^\bullet$ encouraged several radio astronomers to consider the possibility of identifying lines of polyatomic molecules in the radio frequency region [12]. The first success was obtained in 1968 with NH_3 [13]. The proof of the existence of polyatomic molecules sealed the role of astrochemistry opening the long sequence of still unanswered questions about polyatomic molecules formation. Shortly thereafter water (H_2O) [14] and formaldehyde (H_2CO) [15] were detected and since then, many more molecules at a nearly linear rate have enriched the chemicals cabinet of the Universe, introducing once for all chemistry to the ISM physics.

Despite the increasing number of detections, unassigned bands remain across the entire electromagnetic spectrum. There is a claim that part of the unidentified infrared emission bands (UIRs) should be assigned to polycyclic aromatic hydrocarbons (PAHs), or at other aromatic carbon structures [16–20]. Despite these hypothesis, yet no individual molecule has been definitely identified from UIR features. At the time of this writing the updated chemical inventory of the ISM counts about 300 molecules ranging in size from two to seventy atoms.

The updated list is available in online databases such as the Cologne Database for Molecular Spectroscopy's *List of Molecules in Space* and the astrochymist's *A Bibliography of Astromolecules*[21–23]. McGuire publishes regularly updated census of Interstellar Circumstellar, Extragalactic, Protoplanetary Disk, and Exoplanetary Molecules and developed a free Python3 software, *astromol*, to get updated information and graphs tracking the number of detection on a given time scale, the size distribution of the molecules identified so far and the observed frequency range that allowed the detection[24,25].

Among the identified molecules, interstellar Complex Organic Molecules (iCOMs) have attracted a lot of attention. iCOMs are molecules with more than six atoms containing carbon[26]. Particularly attractive are iCOMs with prebiotic potential[27]. Prebiotic molecules are thought to be involved in the processes leading to the origin of life. Usually, molecules in this category are species with structural elements in common with living organisms. A lot of the interstellar prebiotic molecules were first detected toward Sgr-B2(N-LMH). For example, Hollis *et al.* first detected glycolaldehyde (CH_2OHCHO) which is considered the smallest sugar[28]. The tri-carbon sugar, glyceraldehyde, still escape the detection in the ISM[29] and has been only found in meteoric samples[30–32] and in laboratory experiments of interstellar ice analogues[33–35]. Acetamide, CH_3CONH_2, is the largest interstellar molecule with a peptide bond to be successfully detected[36]. In 2008, a direct precursor of the simplest amino acid glycine, amino acetonitrile (NH_2CH_2CN) has been found[37].

Figuring out such complex chemical picture and the interstellar chemistry behind it requires a strong interplay between scientists with different expertise. Knowledge of astronomy, physics, chemistry, biology and geology are crucial to connect the dots between the physical conditions of specific regions and the chemistry that can there take place. Interstellar molecules are detected through their spectroscopic signatures. Currently, there are many active telescopes scanning the Universe trying to capture signals of new molecules. They are usually grouped according to the frequency range that can be explored[38]. As

discussed above the advent of radio-astronomy has positively affected our ability to detect molecules. Indeed, most of them have been identified through their rotational fingerprints. The Institute for Radio Astronomy in the Millimeter Range (IRAM) 30 meter in the Spanish Sierra Nevada is one of the largest single dish radio-telescope and is that with the highest number of detection so far[25]. The performance of a radio telescope in distinguishing details in the observed sky depends on its angular resolution which in turns depends on the wavelength of observations divided by the size of the antenna. The smaller the angular resolution the finer the views of the sky will be. Radio telescopes observe long wavelengths, so even for our largest antennas we will still get an angular resolution similar to that of our unaided eye observing the sky. Higher resolution (comparable to that of optical telescopes) requires a radio telescope's antenna size much, much larger. This mathematical hustle can be overcome by combining the views of a group of antennas spread over a large area operating together as one gigantic telescope. Two successful examples using this principle are: the Atacama Large Millimeter Array (Atacama Large Millimeter/submillimeter Array (ALMA))[39,40] located at 5000 m in height in the Atacama desert in Chile which comprises 66 antennas over 16 kilometers and the Northern Extended Millimeter Array (NOEMA) which is the most powerful radio telescope in the Northern Hemisphere operating at millimeter wavelengths. ALMA is a premier telescope for studying the first stars and galaxies that emerged billions of years ago. We find them at great cosmic distances, with most of their light stretched out to millimeter and submillimeter wavelengths by the expansion of the Universe. In the more nearby Universe, ALMA provides an unprecedented ability to study the processes of star and planet formation. The advent of ALMA shows how important the technological progress is, being spectroscopy the main tool available to investigate the chemical features in the ISM. The radio-frequency range is not the only one that experienced this huge progress, we have also recently experienced this advance with the first pictures of the Universe from the James Webb Space Telescope (JWST) which

has been launched in December 2021[41]. The JWST carries three mid-infrared (IR) instruments. This frequency region was also targeted by Spitzer and Infrared Space Observatory (ISO), however the bigger mirrors of the JWST will definitely make it more sensitive. Small molecules like water, methane, and carbon dioxide have vibrational modes that absorb mid-infrared light. Some of these molecules can be used as life trackers. The foremost example in this sense is water whose presence in the atmosphere of exoplanets (i. e. planets that orbit stars not belonging to our Galaxy) is taken as a prerequisite for hosting life. With these new tools, astrochemists do expect a significant leap in their ability to detect molecules in the Universe and understand their chemical reactivity[42].
Besides astronomical observations, how do astrochemists explain the presence of certain molecules? Which reaction pathways may be in place? If we think about setting up an experiment in a laboratory to answer to these questions, the extreme low temperatures and densities require instruments that allows operating at such harsh conditions. Moreover, collecting samples to be analysed to disentangle the debated identifications is clearly not an option. This scenario is a fertile environment for a strong scientific interplay. In this sense, computational astrochemistry is key in assisting and guiding both the experiments and the astronomical observations. Moreover, the predicting powers of computational simulations do provide atomistic perspective that can be hardly achieved in any other experimental field.

One active research areas in astrochemistry falls in the wider field of computational spectroscopy[43,44]. Accurate strategies have been developed for the prediction of spectroscopic fingerprints following clues to the identification of new molecules. These procedures are extremely useful to disentangle the complex features of the astronomical spectra. Whether a molecule has been identified or we are on a track to its detection, the further step is the study of active processes that may be responsible for its formation.

The physical conditions of the ISM do impose severe constraints to the chemical reactivity. In the coldest and rarefied regions, the probability that a

collision between atoms and molecules occurs is extremely low. These same reasons prevent three-body collisions. Furthermore, because of the very low temperature involved, the reactions need to be exothermic. The most common reactions in gas-phase are ion-neutral molecules and neutral-neutral molecules reactions with no energy barrier. Activation barriers are negligible when very reactive chemicals such as ions and radicals are involved. Though the formation of a single product is possible, the energy generated from the newly formed bond will not have open channels to be dissipated and the reactions will not proceed unless some radiative process occurs. The ubiquitous presence of radiation is responsible for the evolution of many photo-activated processes which can not be ruled out from the full picture[45,46].

So far, we have only focused our attention on gas-phase processes since gas accounts for the 99 % of the total mass of the ISM. Nevertheless, gas phase processes cannot account for the total amount of observed molecular hydrogen. Indeed, it is the remaining 1 % mass, the dusty Universe, the one responsible for the presence of H_2[47,48].

Dust and grains can catalyse reactions by favouring molecular encounters with respect to the gas-phase where the low densities and temperatures prevent collisions[49,50]. The proper catalytic mechanism is still unknown and the morphology and compositions of grains seem to have an impact on the actual chemistry. Diffusion and trapping of molecules on solids surfaces or in the bulk could also affect the observed reactivity.

Apart from this, solid cores shield molecules from the ubiquitous radiation field preventing their degradation before any chemistry could take place. We do not have much information about the structure and composition of the cosmic dust. Moreover a collection of samples for a comprehensive analysis is not feasible and the only information comes from spectral features mostly localised in the mid-IR[51]. Silicates, carbonaceous material and silicon carbide (SiC) are the best candidates as chemical components of cosmic grains. Usually, their IR spectrum includes a strong spectral feature around 9.7 μm, that corresponds to the silicate

Si−O stretching mode. There is also a broad feature at 18 μm presumed to be the O−Si−O bending mode in silicates. The absence of these bands in the outflows from carbon-rich stars where all of the oxygen is locked in CO and silicates cannot form support the existence of cosmic silicates. The bandshape of the observed Si−O stretching band suggests that these silicates have mostly an amorphous structure because crystalline silicates usually present sharper IR features, as reported by laboratory measurements[52,53]. Cosmic silicates are mostly identified with the family of olivines with general formula $Mg_{2x}Fe_{2-2x}SiO_4$ and pyroxene with general formula $Mg_xFe_{1-x}SiO_4$ with x between 0 and 1.

Toward the coldest regions, molecules from the gas-phase freeze out on the mineral core generating an icy mantle[50,54]. The existence of a thick layer of ice is well supported by the detection of the IR bands of the constituent molecules. The strongest band is around 3 μm assigned to the O−H stretching modes in solid H_2O that generally accounts for the 60 —70 % of the icy matrix. Additional weaker features are attributed to small organics such as NH_3, CH_3OH and CH_4. Ices with these components are usually called *polar ices*, since H_2O is dominant. In regions where molecular hydrogen is not as abundant with respect to other gas-phase atoms, C, N and O preferentially stick on the grain surfaces forming *apolar ices* consisting of CO, CO_2, N_2 and O_2[55,56]. Making a structural model out of the observed spectral features is quite challenging. Some of these features can be very weak and overlap to each other and the composition of the icy mantle can alter their position and shape complicating the process. Many experimental factors including deposition temperature, the angle of the impinging flux and the method of deposition affect the H_2O ice structure and properties[57].
Amorphous solid water (ASW) is the dominant H_2O phase in astrophysical environments and is formed at temperatures below $\sim$ 130 K. To complicate the picture further, X-Ray and electron diffraction measurements have shown that ASW exists in two distinct phases. At low temperatures ($\sim$ 10 K) a high density amorphous (HDA with density of 1.1 g cm^{-3}) form exists. At 40 —70 K HDA undergoes an irreversible phase transition to a low density amorphous

phase (LDA with density of 0.94 g cm^{-3}). Above 130 K, ASW transforms into a super-cooled metastable liquid, prior to its hexagonal crystalline ice (CI) form at 160 K[58]. All of these phases differ significantly in porosity and surface area, thus the changes due to the ASW–CI transition can result in the trapping or release of volatile species adsorbed or diffused in/on the ice matrix. The usual hexagonal ice found on Earth is formed above 170 K. It is likely that some H_2O ices have undergone mild processing and recooling and will exist in both amorphous and crystalline forms[59]. The spectroscopic fingerprints of both cosmic dusts and ices fall in the exploration window of the JWST thus, we do expect improvements in our understanding of cosmic particle composition.

UV radiation, cosmic radiation, thermal processing and ions bombardment are active ways of altering ice morphology. Experiments performed on interstellar ice analogues prove that the photo-processing drives a rich chemistry with the formation of many organics[60–62]. In regions where temperatures are high enough to induce the sublimation of the icy mantles, the organics trapped in the bulk will be released and new gas-phase processes can be triggered with the formation of higher molecular weight species. Comets and asteroids spend most of their lifetime in a very cold environments and their surfaces are covered by icy mantles. Bright icy patches have been spotted on the surface of the Comet 67P/Churyumov-Gerasimenko[63]. Cometary ices are made from the gases and ices found in the protoplanetary disc and the prestellar core. It is assumed that prestellar ices may have been inherited by comets[64].

Most of the chemicals discussed so far (organics, ices, dust and refractory materials) originated from nucleosynthesis in the inner cores of stars[65]. The material distribution on a cosmic scale is provided by ejecting mechanisms. A solar-type protostar formation starts at low temperatures (~ 10 K) and number density of 10^3 - 10^4 cm^{-3} with the formation of a dense prestellar core in a giant molecular cloud (GMC). In these regions, gas molecules collide and stick to the surfaces of dust grains. Star formation occurs with the gravitational collapse of such core and the formation of a protostar. Subsequently, the formation of

an accretion disk and a bipolar outflow occurs. As the temperatures increase above 100 K, the water-rich ice mantles sublimate, injecting molecules into the gas phase forming the so-called hot corino regions rich in iCOMs. As long as the star slowly contracts, dust grains coagulate and planetesimals are formed in the disk. The temperature increases until hydrogen burns and stops any further contraction: a new main-sequence star (according to the Hertzsprung–Russell diagrams) is born, possibly surrounded by a planetary system. For a main-sequence star, the nuclear energy becomes sufficient to prevent the star from collapsing. Although it was commonly believed that Solar System objects were made of only minerals, metals, and ices, we now know that they all also contain organic materials. The *in situ* analysis of the cometary dust during the Rosetta mission highlighted the presence of multiple organics[66,67]. Other species detected in meteorites, such as the Marchinson meteorite, may have resulted from space weathering and metamorphism occurring later[68]. All the prebiotic material found in meteorites and other objects is entirely of abiotic origin. Most stars in our Galaxy including the Sun are main sequence stars and go through the formation stages mentioned above. As consequence large amounts of mineral and organic solids have been produced by ordinary stars over the last ten billion years and spread throughout the Galaxy. Possible links between Solar System organics and stellar organics can be tested by isotopic ratios. For example, the isotopic signatures of insoluble organic matter (IOM) in the Murchison meteorite seems to be presolar in origin[69]. The chemistry of asteroids and comets may also be an indicator of the chemistry on the primordial Earth and in particular of the evolution of primordial life forms[3,70]. The Earth has a large reservoir of organics, almost all have biological origins.

The Oparin-Haldane hypothesis[71] and the Miller-Urey experiment[72] proved possible that simple ingredients in a suitable environment, the primordial soup, originated life on the early Earth. This endogenous scenario has ruled for more than 50 years. However, the variety of organics formed during the stellar evolution and their wide presence among Solar System objects, has provided

some supporters for an alternative, exogenous delivery hypothesis. In this scenario, comets, asteroids, interplanetary dust particles and micrometeorites act as exogenous delivery agents. Shortly after the formation of the Earth, a large number of minor bodies impacted on its surface during the late heavy bombardment about 3.9 Gyr ago. Comets experienced a migration from their original location between Uranus and Neptune. According to our knowledge on the actual composition of comets and asteroids, we expect that their loads occasionally contained prebiotic species that were delivered to our young planet with some degree of alteration [73]. We cannot rule out the possibility that the planetesimals that aggregated to form the primordial Earth may have contained some star dust. In such case, these organics could well survive the heat and shock conditions during the Earth-formation process and be trapped in the Earth to be released under specific conditions, for example those in the hydrothermal vents. These prebiotic materials could form the ingredients of the first steps to life in the early Earth. Though these two scenarios may seem in contrast to each other, the two are not necessarily mutually exclusive. The complexity of the real world may be the result of a complex series of events relatable to a mixed exogenous/endogenous scenario in which the terrestrial chemical inventory was first enriched from outer delivery but it was only thanks to the unique physical conditions of our Planet that the complex evolutionary process started. This debate is still open and many questions need to be answered. Was the sparkle of a random unique event or a complex mechanism of multiple processes? Was a sole event in the vast Universe or something that happened many times with life spread somewhere else far from our sight? These are the ancestors questions that recall for millions of other phenomena to be understood before the entire genealogical tree of life molecules could be fully drawn. The final understanding depends on our ability to answer smaller related questions and this thesis is a journey in our attempt to develop methods and understandings that may be of use to the entire community active in astrochemistry. With this intent, at the start of my PhD we set our goal to be the investigation of the chemical processes

at the gas-ice interface in the interstellar medium. During the last four years, while exploring this topic, many issues and new questions have been raised and the work has spread over different areas with one main *fil rouge*: the interstellar ice role in the chemistry of the ISM.

1.1 Manuscript organization and timeline

The Introduction presented above is a brief summary of some of the main milestones since the birth of astrochemistry. In this intriguing research area, my PhD is a four year work mostly based on the use of state-of-the-art computational chemistry tools in order to develop accurate, yet, cost-effective protocols to explore the chemistry at the interstellar ices interface. The work may be divided in three main lines: (I) the thermodynamic and kinetic investigation of processes occurring either in gas-phase or on interstellar water ice surfaces, (II) the set up of a procedure for the prediction of vibrational features of organics adsorbed on interstellar ice analogues, (III) the investigation of the ice morphology and its effect on the adsorption of small organics. In what follows, I will present in depth the timeline of the work performed in the last four years and the organization of this manuscript.

This thesis comprises seven chapters. This is the Introduction (Chapter **1**) where the background and the main topics related to astrochemisty are presented. The second chapter (Chapter **2**) presents a summary of the theoretical background underlying the methods used in this thesis. In particular, theoretical descriptions of methods for accurate thermochemical, spectroscopic and kinetic data are presented. A picture of common methods for both gas-phase and solid state simulations can also be found. The research work about the definition of effective computational protocols and their applications is presented in the next chapters. Chapter **3** provides a test case for the investigation of gas-phase formation pathways of astrochemical interest using the paradigmatic case of ethanimine as example. The focus is then shifted to the importance of defining a proper

strategy for the simulation of interstellar water ice (Chapter **4**, Chapter **5** and Chapter **6**). In these three chapters, we describe a procedure for the accurate simulation of processes at the interstellar ice interface (Chapter **4**). This procedure is then implemented for the proper simulation of IR fingerprints of iCOMs adsorbed on ice (Chapter **5**). Finally, the focus is on the ice morphology and the effect of mineral grains on the growing icy mantle (Chapter **6**). The last chapter (Chapter **7**) summarizes the conclusions of this work and provide a short discussion about the future studies that could extend our current knowledge.

As scientists we are aware of the role that collaboration has in the researcher life. This work benefits from the collaboration of people with different expertise and in particular, I would like to mention the work of Dr. Jacopo Lupi who performed all the kinetic calculations that the reader will find in this thesis. The work devoted to the study of amorphous structures would have not been possible without the support of HPC-Europa3, the University of Edinburgh and the expertise of Dr. Valentina Erastova.

A Gantt chart is provided for a visual representation of the research timeline and milestones (Figure **1.1**).

The timeline for the completion of the PhD thesis can be summarized as follows:

1. First Year (11/2018 —11/2019):
    - The first year of the PhD program was primarily focused on education and familiarization with gas-phase system simulations.
    - During this time, I acquired the necessary knowledge and skills for conducting research in the field.
    - The work presented in Chapter **3** is a result of the efforts made during this year.
    - This work resulted in a publication in an international journal[74].
2. Second Year (11/2019 —11/2020):

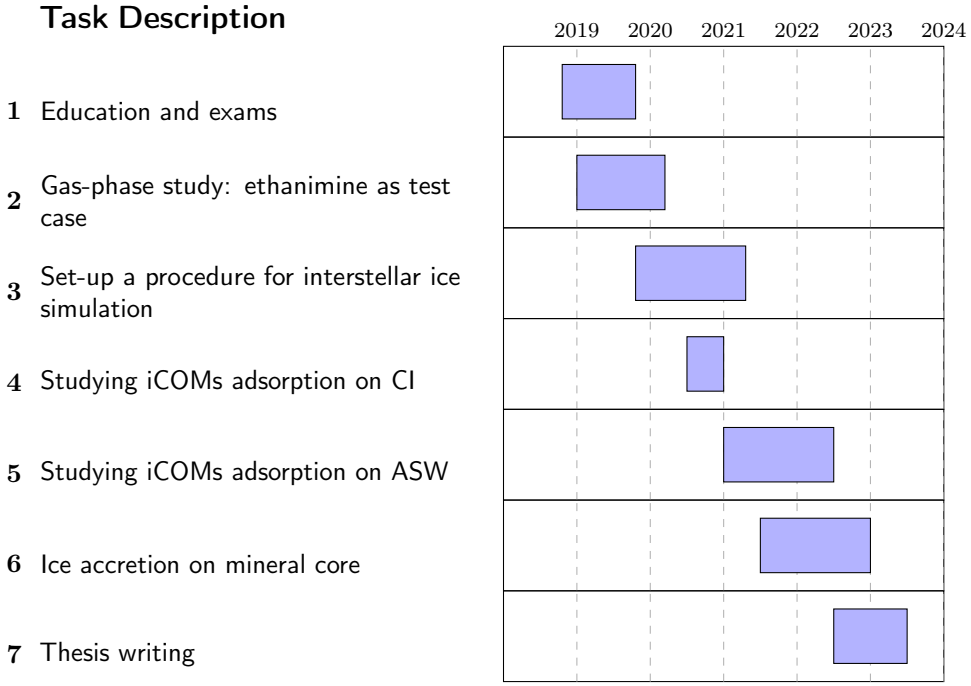

Figure 1.1 Gantt chart for this four years PhD work.

- The second year marked the beginning of the main research line, which focused on simulating processes on ice surfaces.
- The primary goal during this year was to develop an accurate yet cost-effective simulation strategy for studying processes at the ice surface.
- This work resulted in a publication in an international journal [75].

3. Third Year (11/2020 —11/2021):

- This year involved considerations about the ice morphology.
- My project received funding within the HPC-Europa3 program, and I relocated to Edinburgh to work on assessing the amorphous nature of ice.
- The adsorption of iCOMs on ASW was studied during this period.

4. Fourth Year (11/2021 —11/2022):

- The focus shifted towards investigating the IR features of adsorbed iCOMs. We focused on the effect of ice morphology by comparing IR features of adsorbed iCOMs on both CI and ASW.
- In December, a new project received funding within the HPC-Europa3 program, and we started analyzing the effect of mineral cores on the morphology of the ice mantle.

5. Extra Months (11/2022 —03/2023):

   - These additional months were primarily dedicated to the analysis of the final results and the writing of the thesis.

Chapter 2

Methods

The main goal of astrochemistry is to shed light on the chemistry running in the Universe. Whether it is about laboratory experiments, telescopes data management or computational simulations, astrochemist's job is to develop a full knowledge of the active chemical processes. Thanks to the continuous advancement in hardware resources and the development of efficient algorithms, theoretical and computational chemistry has become a powerful tool applicable to many problems of interest in technology and life science, including astrochemistry and environmental chemistry.

The available codes make use of methods that in their mathematical formulation required some approximations in order to simplify the complex intrinsic nature of the real world we intend to describe. This approach makes our models usable and not uselessly demanding if not required. Speaking in chemical terms, simplifications are usually introduced when working with multi-components phases, variable physical parameters such as temperature and density or challenging systems in terms of their morphology and composition such as defective and amorphous solids. Depending on the chemical system under study and the target property and/or application, the optimal method in terms of accuracy and computational effort needs to be selected.

If our interest is in the chemistry of the ISM, we may probably want to

answer the following questions:

- Which molecules populate the observed environment and what kind of chemistry could potentially take place?
- How the detected molecules could have formed in the first place?
- According to our knowledge about the first two points, can we spot new molecules among the still unidentified spectroscopic features?

This chapter reviews and presents computational methods that have been used to develop suitable computational protocols to tackle these topics. The first two points require the study of the thermodynamics and kinetics of potential chemical pathways that could eventually take place in either gas or solid phase. Some methods will be presented with a brief description of the mathematical formulation along with the benefits and drawbacks coming from their usage. We will then, focus on the importance of the accuracy of our results and how this can be balanced out with the computational burden and the size of the system under study. The reader will be guided through the methods used for thermodynamics and kinetics. Since thermodynamically favoured reactions could be overthrown by an hostile kinetics a full understanding of the thermodynamics and kinetics of a chemical process is mandatory for a proper chemical modelling. The last point is all about matching predicted spectroscopic features with experimental and astronomical observations. We will present the methods of computational spectroscopy applied to the studied systems in this thesis which are of use to dissect the complexity of recorded observational and/or experimental spectra. We will focus on vibrational spectroscopy since most of the work during this PhD has been devoted to the understanding of organics adsorption on interstellar ice surfaces. In the end, methods to scale-up the system size will be presented in order to build structural models resembling a bit more the real ones or at least what we know about their structure and composition so far. Reading this chapter a knowledge of the methods used along this PhD research will be acquired and

the reader will go smoothly through the last chapter diving into the specific studied cases and the knowledge acquired from the results of this work.

2.1 Electronic structure methods

With the development of theories in Quantum Mechanics (QM) and the technological advances in supercomputers, computational chemistry developed rapidly. Nowadays, the most modern QM theories have been implemented in both free and licensed software and running calculations and experiments side-by-side is the new paradigm of science. All the research areas (from the pharmaceutical to green chemistry) benefit from the use of computational chemistry tools and astrochemistry is not far behind. In the next subsections some of such tools will be presented with focused discussion on the methods used in this thesis work.

2.1.1 Hartree-Fock

A full picture of the static properties of a molecule requires the resolution of the time-independent Schrödinger equation (SE)[76]:

$$\hat{H}\Psi = E\Psi. \tag{2.1}$$

$\hat{H}$ is the Hamiltonian of the system i.e., the sum of its kinetic energy and its potential energy and Ψ is the electronic wavefunction. The above equation is usually evaluated within the Born-Oppenheimer approximation (BOA)[77]. The BOA allows the factorization of the nuclear and electronic problems and the electronic Schrödinger equation can be solved for a fixed set of nuclear positions retrieving the *potential energy surface* (PES). The exact solution of such equation for a system with more than two electrons is a challenging task. Many different methods have been proposed with the intent to unravel the knot of this mathematical procedure. In particular, *ab-initio* methods solve the above equation without any parameter coming from the fitting of experimental values.

In this section, we will briefly introduce the Hartree-Fock (HF) method, as the starting point of wavefunction based approaches. We will recall a few points in the HF theory which will make easier and more fluent the further discussion about more advanced procedures which have been used for the presented work.

In the HF method, the wavefunction is given by a single Slater determinant of N spin-orbitals. The HF procedure offers only an approximate solution to the Schrödinger equation since the electron-electron correlation is approximated by an average interaction between the electrons. Though in a sufficiently large basis set the HF wavefunction is able to account for $\sim$ 99 % of the total energy, the remaining $\sim$ 1 % can be very important in the description of many chemical phenomena. The difference between the exact non-relativistic energy of the system (ε_0) and the Hartree-Fock energy (E_0) is called correlation energy (E_{corr}):

$$E_{corr} = \varepsilon_0 - E_0. \tag{2.2}$$

HF makes use of the variational principle hence, the HF energy is the upper bound to the exact energy[78–80]. In the next sections, approaches to recover electron correlation (the so called post-HF methods) will be described. While many procedures have been developed, we will focus on those used for the research work presented in this dissertation.

2.1.2 Post-HF methods

In the attempt to recover part of the electronic correlation, post-HF methods have been developed, which give more accurate results than HF calculations, although the added accuracy comes with the price of added computational cost. They normally use the HF wavefunction as starting point for improvements and the generic, derived multideterminant trial function will have the following expression:

$$\Psi = a_0\phi_{HF} + \sum_{i=1} a_i\phi_i, \tag{2.3}$$

where the coefficient a_0 provides the contribution of the ground-state HF wavefunction and the higher order a_i coefficients retrieve the contribution of the electronic correlation. We will see these terms in details for various methods. Following this approach the first method developed to recover electron correlation is the configuration interaction (CI) [81]. CI describes the exact wavefunction as a linear combination of configuration state functions (CSFs) built from spin-orbitals. If all the CSFs were included (full-CI) and the basis were complete, the electronic Schrödinger equation would be correctly solved. Even for small molecules and moderately sized one-electron basis sets, the number of N-electron determinants for a full-CI is huge and the computational demand becomes insurmountable. Thus, truncated CI methods in which only excitation till a specific order are considered are much more of use. The major problems of truncated CI methods are size-inconsistency and size-inextensivity. Indeed, the energy of two infinitely separated particles is not double the energy of the single particle (size-inconsistency) and the predicted energy does not scale proportionally with the number of electrons (size-extensivity). A different method to treat electronic correlation comes with the Many-Body Perturbation Theory (MBPT). The general idea of perturbation methods is that real systems (perturbed problems) are the results of a small perturbation acting on an unperturbed system for which a solution is already available. In this case, the Hamiltonian operator for the actual system can be partitioned into two contributions: a reference term ($\hat{H}_0$) corresponding to the Hamiltonian of the model (unperturbed) system and the perturbation ($\hat{H}'$) describing its alterations:

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}', \tag{2.4}$$

where λ is a dimensionless parameter that goes from 0 to 1 mapping $\hat{H}_0$ into $\hat{H}$. As the perturbation increases, the energy and wavefunction must change continuously and they can be expanded in a power series of the λ parameter that, in turn, keeps track of the order of the perturbation.

$$E = E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \ldots, \tag{2.5}$$

$$\Psi = \Psi_0 + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \ldots, \tag{2.6}$$

where $E^{(0)}$ is the eigenvalue for Ψ_0 which is the ground-state eigenfunction of $\hat{H}_0$ and the terms having the superscript n are the *nth-order correction* to the zeroth-order term. Using Equation (2.4) into Equation (2.1), we may write:

$$(\hat{H}_0 + \lambda \hat{H}')|\Psi\rangle = E|\Psi\rangle, \tag{2.7}$$

by resorting to Equation (2.5) and Equation (2.6), terms can be collected according to powers of λ and the terms for the different orders of perturbation can be retrieved:

$$\lambda^0 : \hat{H}_0 \Psi_0 = E^{(0)} \Psi_0, \tag{2.8}$$

$$\lambda^1 : \hat{H}_0 \Psi^{(1)} + \hat{H}' \Psi_0 = E^{(0)} \Psi^{(1)} + E^{(1)} \Psi_0, \tag{2.9}$$

$$\lambda^2 : \hat{H}_0 \Psi^{(2)} + \hat{H}' \Psi^{(1)} = E^{(0)} \Psi^{(2)} + E^{(1)} \Psi^{(1)} + E^{(2)} \Psi_0. \tag{2.10}$$

In general, at the nth order:

$$\lambda^n : \hat{H}_0 \Psi^{(n)} + \hat{H}' \Psi^{(n-1)} = \sum_{i=0}^{n} E^{(i)} \Psi^{(n-i)}. \tag{2.11}$$

Equation (2.8) is the known zeroth-order solution while Equation (2.9) involves the first-order corrections to the wavefunction and energy.

The solution to the unperturbed problem generates a complete set of functions (ϕ) and the unknown corrections to the wavefunction can be expanded in terms of such functions.

$$\begin{aligned}\Psi^{(1)} &= \sum_i a_i \phi_i, \\ \Psi^{(2)} &= \sum_i b_i \phi_i,\end{aligned} \tag{2.12}$$

with a_i and b_i expansion coefficients. The derived expressions for both energy and wavefunction till the second-order are:

$$E^{(0)} = \langle \Phi_0 | \hat{H}_0 | \Phi_0 \rangle, \tag{2.13}$$

$$E^{(1)} = \langle \Phi_0 | \hat{H}' | \Phi_0 \rangle, \tag{2.14}$$

$$E^{(2)} = \sum_i a_i \left\langle \Phi_0 | \hat{H}' | \Phi_i \right\rangle = \sum_{i \neq 0} \frac{\left\langle \Phi_0 | \hat{H}' | \Phi_i \right\rangle \left\langle \Phi_i | \hat{H}' | \Phi_0 \right\rangle}{E_0 - E_i}, \tag{2.15}$$

$$a_j = \frac{\left\langle \phi_j | \hat{H}' | \phi_0 \right\rangle}{E_0 - E_j}, \tag{2.16}$$

$$b_j = \sum_{i \neq 0} \frac{\left\langle \Phi_j | \hat{H}' | \Phi_i \right\rangle \left\langle \Phi_i | \hat{H}' | \Phi_0 \right\rangle}{(E_0 - E_j)(E_0 - E_i)} - \frac{\left\langle \Phi_j | \hat{H}' | \Phi_0 \right\rangle \left\langle \Phi_0 | \hat{H}' | \Phi_0 \right\rangle}{(E_0 - E_j)^2}. \tag{2.17}$$

The first-order correction to the energy and wavefunction (Equations (2.14) and (2.16)) only require the unperturbed energy and wavefunction. The second-order corrections need both the unperturbed and first-order corrections (see Equations (2.15) and (2.17)). A specific case of MBPT is the Møller-Plesset perturbation theory known as MPn method where n denotes the order of the perturbative treatment[82].

The final expression for the second-order correction to energy is the following:

$$E^{(2)} = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\langle \Phi_0 | H' | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | H' | \Phi_0 \rangle}{E_0 - E_{ij}^{ab}}, \tag{2.18}$$

which in MP2 becomes:

$$E(MP2) = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{(\langle\phi_i\phi_j|\phi_a\phi_b\rangle - \langle\phi_i\phi_j|\phi_b\phi_a\rangle)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{2.19}$$

Expressions for higher order terms follow similarly, although with greater algebraic and computational complexity.

MP2 is an M^5_{basis} method with M number of basis functions, but fairly inexpensive since only two-electron integrals over two occupied and two virtual molecular orbitals (MOs) are calculated which means that MP2 energy can sometimes be calculated at a fair price and generally accounts for about 80 —90 % of the electron correlation. MPn methods are the only MBPT methods to be size extensive. Perturbation methods are based on the inclusion of all corrections at different type (single, double, triple and so on) to a given order (2, 3, 4, *etc.*). MPn methods of higher order than MP2 are rarely used and coupled-cluster (CC) methods are usually preferred. These include all corrections of a specific type to infinite order. The CC method assumes an exponential ansatz for the wavefunction:

$$\Psi_{CC} = e^{\hat{T}}\Phi_{HF}. \tag{2.20}$$

The cluster operator, $\hat{T}$, acts on the wavefunction generating all possible excitation of a specific kind. It can be written as a Taylor expansion of the type:

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + ... = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{T}^k. \tag{2.21}$$

Grouping terms according to the excitation order Equation (2.21) can be recast in the following form:

$$\begin{aligned} e^{\hat{T}} = 1 + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + (\hat{T}_3 + \hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3) \\ +(\hat{T}_4 + \hat{T}_3\hat{T}_1 + \frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_2\hat{T}_1^2 + \frac{1}{24}\hat{T}_1^4) + ... \end{aligned} \tag{2.22}$$

In the above equation, the first term on the right-hand side generates the HF

reference and the second term all the single excitations. The first parenthesis collects terms of doubly excited states which can be connected ($\hat{T}^2$) or disconnected ($\hat{T}_1^2$). The second parenthesis collects all the triple excitations ($\hat{T}^3$ or $\hat{T}_2\hat{T}_1$ or $\hat{T}_1^3$) and so on. So far, the treatment of the CC methods is exact, however, the full applicability is restricted to small systems with few atoms. Moreover, it scales quite fast with both the system size and the basis set which makes such method extremely demanding for practical applications. Hence, the CC operator ($\hat{T}$) must be truncated. The CCSD, CCSDT, CCSDTQ, *etc.*, are obtained by truncating the series in Equation (2.21) to the double, triple, quadruple, *etc.*, excitations, respectively. The way the method is built fix the issues about the size extensivity.

The chemical accuracy that can be reached encounters obstacles rooted in the steep scaling of CC methods. The CCSDT method scales as M^8 and is also impractical for most systems. Alternatively to a full treatment of the excitations at the nth order, contributions till a specific excitation order can be estimated through perturbation theory. The most common example is CCSD(T) where triple excitations are estimated as in MP4[83]. Higher order methods such as CCSD(TQ) and CCSDT(Q) are possible but, again, so demanding to be used for anything but very small systems. While the scaling of CCSD(T) is still steep (M^7), the one-step (noniterative) evaluation of the effects of triple excitations provides the best compromise between accuracy and computational applicability, and the method has come to be known as the *gold standard* of quantum chemistry. Most of the methods mentioned here are single-determinant reference methods. In order to describe the multireference character in the wavefunction, the $\mathcal{T}_1$ diagnostic procedure has been developed. In general, singles amplitudes (t_1) can be taken as indication of how good the HF single-determinant function is as reference. The $\mathcal{T}_1$ diagnostic for a system with N_{ele} electrons is defined as follows:

$$\hat{T}_1 = \frac{1}{\sqrt{N_{ele}}}|t_1|. \tag{2.23}$$

As a rule of thumb, if $\mathcal{T}_1 < 0.02$ the CCSD(T) is expected to give good results otherwise multi-reference methods should be considered.

2.1.3 Basis sets

A basis set is a set of functions (called basis functions) which are linearly combined to model MO. These functions are typically atomic orbitals (Atomic Orbital (AO)) centered on atoms, but can, theoretically, be any function, for example plane waves are frequently used in solid-state computational chemistry and physics. From a general point of view, there are two types of atom-centered basis functions that are employed in quantum chemical calculations: Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs). STOs have the following mathematical form in atom-centered polar coordinates:

$$R(r) = Nr^{n-1}e^{-\zeta r}Y_l^m(\theta,\psi), \tag{2.24}$$

where n is a natural number that plays the role of principal quantum number, $n = 1, 2, \ldots$, capital N is a normalizing constant, r is the distance of the electron from the atomic nucleus, ζ is the Slater orbital exponent and $Y_l^m(\theta,\psi)$ spherical harmonics. More than one STO can be used to represent one atomic orbital. The larger the basis set, the higher the computational time. Unfortunately, STOs are not well suited in numerical work because some of the integrals (particularly three- and four- center integrals) are difficult, if not impossible, to evaluate. For this reason, GTOs were introduced into molecular orbital calculations.

A GTO is obtained as:

$$G_{nlm}(r,\theta,\psi) = N_n \underbrace{r^{n-1}e^{-\alpha r^2}}_{\text{radial part}} \underbrace{Y_l^m(\theta,\psi)}_{\text{angular part}}, \tag{2.25}$$

with α exponential factor. Unfortunately, Gaussian functions do not match the shape of an atomic orbital very well. In particular, they are flat rather than steep near the atomic nucleus at $r = 0$, and they fall off more rapidly at large

values of r. To compensate for this problem, linear combinations of Gaussians are formed to approximate the radial part of a STO. Each STO is then replaced with a number of Gaussian functions with different values for the exponential parameter.

$$\phi(x, y, z; \alpha, i, j, k) = \sum_{a=1}^{M} c_\alpha \phi(x, y, z; \alpha_a, i, j, k), \tag{2.26}$$

where M is the number of Gaussians used in the linear combination, and the coefficients c are chosen to optimize the shape of the basis function sum and ensure normalization. When a basis function is defined as a linear combination of Gaussians, it is referred to as a *contracted* basis function, and the individual Gaussians are called *primitive* Gaussians. The degree of contraction depends on the total number of primitives used. Contracted GTOs permit analytical evaluation of all of the four-center integrals. In 1969, for the first time, Hehre, Stewart, and Pople systematically determined optimal contraction coefficients and exponents for mimicking STOs with contracted GTOs for a large number of atoms in the periodic table[84]. They called these different basis sets STO-MG, which stands for Slater-Type Orbital approximated by M Gaussians.

The STO-3G is represented by a linear combination of three primitive Gaussian functions and is called a single-ζ basis set which means that there is only one basis function for each orbital from core through valence. Thus for H and He, there is only a 1s function. For Li to Ne, there are five functions, $1s$, $2s$, $2p_x$, $2p_y$, and $2p_z$, *etc.* This number is the absolute minimum required, and it is certainly nowhere near the infinite basis set limit. A basis set with two functions for each atomic orbital is called a double-ζ basis set. Of course, we could decontract further, and we would have higher multiple-ζ basis sets. Modern examples of such basis sets are the cc-pVnZ sets of Dunning and co-workers, where the acronym stands for 'correlation-consistent polarized Valence (Double/Triple/*etc.*) Zeta'[85].

From a chemical standpoint, a proper description of a chemical bond could

benefit from higher flexibility in the valence basis functions than in the core. This led to the development of so-called split-valence or valence-multiple-ζ basis sets. In such basis sets, core orbitals are still represented by a single (contracted) basis function, while valence orbitals are split into many functions. Amongst the most widely used split-valence basis sets are those of Pople *et al.* that include 3-21G, 6-21G, 4-31G, 6-31G, and 6-311G[86–88]. The first number in this nomenclature indicates the number of primitives used in the contracted core functions. The numbers after the hyphen indicate the numbers of primitives used in the valence functions. Two numbers indicate a valence-double-ζ basis, three a valence-triple-ζ. The scheme used by Pople and co-workers and other researchers in the field, relied on the variational principle. That is, exponents and coefficients were optimized so as to match the minimum energy over predefined test sets. One feature of the Pople basis sets is that they use a so-called *segmented* contraction. This implies that the primitives used for one basis function are not used for another of the same angular momentum. An alternative method to carrying out a segmented contraction is to use a so-called *general* contraction[89]. In a general contraction, there is a single set of primitives that are used in all contracted basis functions, but they appear with different coefficients in each. The general contraction scheme has some technical advantages over the segmented one. One advantage in terms of efficiency is that integrals involving the same primitives, are calculated only once, and stored for later reuse as needed. Split-valence basis sets using general contractions are the cc-pVDZ, cc-pVTZ. The *correlation-consistent* part of the name implies that the exponents and contraction coefficients were variationally optimized not only for HF calculations, but also for calculations including electron correlation. However, s and p functions centered on the atoms do not provide sufficient mathematical flexibility to adequately describe the wavefunction. This flexibility is almost always added in the form of basis functions corresponding to one quantum number of higher angular momentum than the valence orbitals. Thus, for a first-row atom, the most useful polarization functions are d GTOs, and for hydrogen, p GTOs. Balanced double-ζ basis sets

should include d functions on heavy atoms and p functions on H, triple-ζ basis sets should include 1 set of f and 2 sets of d functions on heavy atoms, and 1 set of d and 2 sets of p functions on H, *etc.* This is the polarization prescription adopted by the cc-pVnZ basis sets of Dunning and co-workers already mentioned above. When a basis set does not have the flexibility necessary to allow a weakly bound electron to localize far from the remaining density, significant errors in energies and other molecular properties can occur. To address this limitation, standard basis sets are often *augmented* with diffuse basis functions. In the Dunning family of cc-pVnZ basis sets, diffuse functions on all atoms are indicated by prefixing with *aug*. Moreover, one set of diffuse functions is added for each angular momentum already present. Thus, aug-cc-pVTZ has diffuse f, d, p, and s functions on heavy atoms and diffuse d, p, and s functions on H and He. It has been realized for a long time that basis sets for electronic structure calculations must contain diffuse functions for quantitative accuracy. It has also been shown that for some molecular properties diffuse functions have a relatively small effect (it is the case of ionization potentials and stationary point geometries). It is best to avoid using diffuse basis functions when they are unnecessary because adding diffuse functions to a basis set not only increases the cost of the calculation but can cause problems with self-consistent field (SCF) convergence and can increase basis set superposition error (BSSE). In this sense, Truhlar developed minimal augmentation with a full variety of expansion coefficients and diffuse functions that can be used to keep the computational cost reasonable[90].

2.1.4 Composite schemes

It is now well-recognized that accurate methods such as CCSD(T) have to be evaluated with large basis sets containing high angular momentum basis functions to attain convergence in the calculated energies. The most commonly used systematic basis sets in this approach belong to the correlation-consistent family discussed in the previous section. Ideally, explicit calculations on progressively larger basis sets have to be carried out, and the results have to be extrapolated

to the basis set limit. Important effects such as zero-point vibrational corrections, scalar relativistic effects, *etc.* can then be included to yield accurate results for small molecules. However, a direct evaluation of all the important effects with large basis sets is frequently impractical, and a composite approach involving a series of calculations is almost invariably carried out to incorporate all the key interactions.

In the previous section some of the most common post Hartree-Fock methods have been presented. Accurate QM procedures come with a high price and making them more affordable requires simplifications which reduce the intrinsic accuracy of the calculation. Two are the main sources of inaccuracy in QM methods: the wavefunction and the basis set truncation. A proper evaluation of all the possible excitations and the inclusion of all the possible determinants in Equation (2.3) with an infinite basis set would provide an exact solution to the SE. However, this approach is impossible to be implemented, hence, truncation in both the number of determinants for the electron correlation and the basis set are mandatory. The higher the order of such truncation the more accurate and by consequence demanding the procedure will be. So, in terms of wavefunction truncation, the computational demand and the accuracy for the methods we have discussed will increase in the following order:

$$\text{HF} < \text{MP2} < \text{CISD} < \text{MP4} < \text{CCSD(T)} < \text{Full CI}.$$

Same for the basis set size:

$$\text{SD} < \text{TZ} < \text{QZ} < \text{5Z} < \text{6Z} < ... < \infty\text{Z}.$$

Combining methods and basis sets the behaviour reported in Figure **2.1** can be sketched: the accuracy increases moving in the direction of the arrow and exact results are obtained for full CI computations in the limit of an infinite basis set, i.e. the so-called complete basis set (CBS).

The higher we move in the sequences above, the more difficult and computationally demanding becomes the set-up of an applicable procedure. Composite methods aim for high accuracy by combining the results of several calculations.

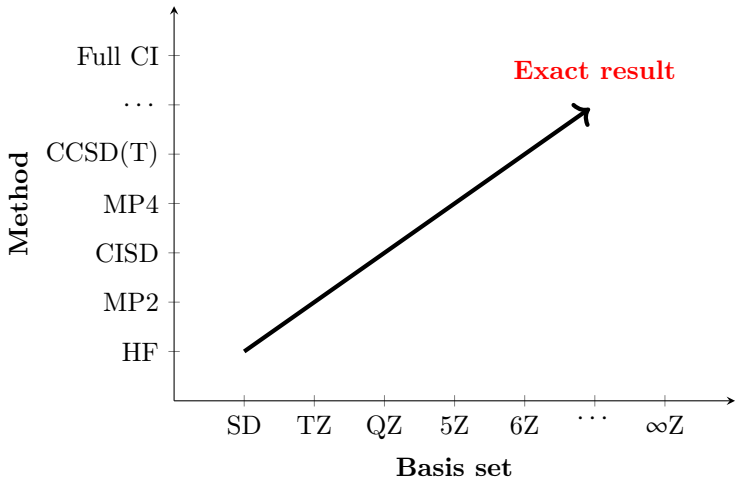

Figure 2.1 Convergence to exact solution with respect to selected methods and basis sets.

They combine methods with a high level of theory and a small basis set with methods that employ lower levels of theory with larger basis sets. They aim for chemical accuracy, i.e. within 1 kcal/mol. The key assumption is the *additivity* approximation that different contributions can be individually computed, and then summed together to predict their cumulative effects. While such multilevel methods have been employed for rather a long time, only in the late 1980s the approach has been systematized so as to define a model chemistry with broad applicability. One of the earliest methods to yield accurate thermochemical information for molecules was the composite Gaussian-1 theory developed by Pople *et al.* in 1989[91]. Since then, several families of composite methods have been developed: the Weizmann-n (Wn) theory[92], the Feller–Peterson–Dixon (FDP) procedure[93], the high-accuracy extrapolated ab-initio thermochemistry (HEAT) protocol[94–96], the multi-coefficient methods of Truhlar and coworkers[97] and the correlation consistent Composite Approach (ccCA) developed by Wilson and coworkers[98].

2.1.4.1 Composite schemes for energy extrapolation

The exact results to the CBS limit would require the use of an infinite basis set which, of course, is impracticable. To address this problem, extrapolation procedures have been developed with the intent to recover part of the lost

accuracy due to the basis set truncation and push the accuracy to the desired level. The basis set extrapolation approach requires a knowledge of the basis set convergence behaviour and rates towards the basis set limit for a specific method. This allows the definition of the appropriate extrapolation technique. In order to be of practical use, an extrapolation procedure needs a hierarchical sequence of basis sets that approach the basis-set limits in controlled and systematic manner. The cc-pVnZ basis set family was designed such that all basis functions in shells recover roughly the same amount of correlation energy (for example the 2d and 1f functions of the cc-pVTZ sets for the atoms of the second row recover approximately the same amount of correlation energy, as the 3d, 2f, and 1g functions of the cc-pVQZ sets). This design results in a hierarchical family of basis sets that has the property that they systematically approach the CBS limit. Hence, the energies obtained from a series of basis sets (double, triple, and quadruple-ζ quality sets) could be used to extrapolate the energy of the system to the CBS limit.

With this in mind, the total energy to the CBS limit (E_{CBS}) is defined as the sum of the electronic HF-SCF and correlation energies extrapolated individually at the CBS limit, E_{CBS}^{HF-SCF} and E_{CBS}^{corr} respectively:

$$E_{CBS} = E_{CBS}^{HF-SCF} + E_{CBS}^{corr}. \tag{2.27}$$

The above expression refers to the electronic energy calculated from the non-relativistic electronic Hamiltonian within the BOA. Further improvements to Equation (2.27) add corrective terms accounting for higher order excitation, scalar relativistic effects, spin–orbit interactions and the so-called diagonal Born–Oppenheimer correction. In what follows we will briefly discuss the HEAT protocol and more recent composite procedures which have been selected for the studied cases in this thesis. The total HEAT energy (E_{HEAT}) is[94]:

$$E_{HEAT} = E_{CBS}^{HF-SCF} + \Delta E_{CBS}^{CCSD(T)} + \Delta E^{CCSDT} + \Delta E^{CCSDTQ} + \Delta E_{Rel} + \Delta E_{ZPE} + \Delta E_{DBOC} + \Delta E_{SO}. \tag{2.28}$$

E_{CBS}^{HF-SCF} and $\Delta E_{CBS}^{CCSD(T)}$ correspond respectively to E_{CBS}^{HF-SCF} and E_{CBS}^{corr} (see Equation (2.27)). ΔE^{CCSDT} and ΔE^{CCSDTQ} retrieve part of the accuracy lost by evaluating the correlation term at CCSD(T) level, ΔE_{Rel} includes scalar relativistic correction, ΔE_{ZPE} includes zero-point energy corrections, ΔE_{DBOC} accounts for the contribution of the nuclear kinetic energy operator and ΔE_{SO} accounts for spin-orbit coupling. The extrapolation of the HF-SCF energy uses the exponential formula suggested by Feller[99] upon the analysis of the convergence behaviour to the CBS limit of electronic energy in the case of water:

$$E_{basis}^{HF} = E_{CBS}^{HF} + a \exp(-bX). \tag{2.29}$$

The parameters, a, b, and the extrapolated HF-SCF energy are determined uniquely from the energies evaluated using three basis set of increasing cardinal number X in the aug-cc-pCVXZ family. This is one of the major peculiarities of the HEAT protocol. For the correlation energy, the two-points extrapolation formula by Halkier is used[100].

$$\Delta E_{basis}^{CCSD(T)} = \Delta E_{CBS}^{CCSD(T)} + \frac{a}{X^3}, \tag{2.30}$$

with a and $\Delta E_{CBS}^{CCSD(T)}$ parameters to be determined. Due to the slow convergence of the correlation energy with the basis set size which is usually considered the bottleneck in the accuracy of composite methods, many extrapolation formula have been proposed along the years[101–105]. Here we will only present that proposed by Halkier *et al.*[100] which is the one selected for the HEAT protocol and all of the composite approaches we are going to discuss into details later in this section. Here, both $E_{basis}^{CCSD(T)}$ and $\Delta E_{CBS}^{CCSD(T)}$ are not the total CCSD(T)

energy but only the correlation energy. The two unknown parameters, a and the estimated CCSD(T) correlation energy at the CBS limit can be obtained by the calculation of the correlation energy with two basis sets within the aug-cc-pCVXZ series.

With CCSD(T) the effect of triple excitations is only evaluated in a perturbative manner. For cases where triples corrections are large, ΔE^{CCSDT} becomes relevant. However, it is not possible to perform full CCSDT calculations using the large basis sets used in the extrapolated CCSD(T) energies defined above and used in Equation (2.30). Since correlation effects beyond CCSD(T) do not benefit particularly by the usage of bigger basis set, cc-pVTZ and cc-pVQZ basis are used within the HEAT protocol. Implicitly assumed here is that effects due to diffuse functions and core correlation are already given sufficiently well by the extrapolated HF-SCF and CCSD(T) energies. The formula for the CCSDT-CCSD(T) energy difference as evaluated in the frozen core approximation (fc) is:

$$\Delta E^{CCSDT} = E_{TQ}^{CCSDT}(fc) - E_{TQ}^{CCSD(T)}(fc), \tag{2.31}$$

where TQ denotes that the corresponding contribution has been obtained by the two points extrapolation formula in Equation (2.30) using the frozen core CCSDT and CCSD(T) energies obtained with the cc-pVTZ and cc-pVQZ basis sets. Despite its computational complexity and cost, even CCSDT may not provide sufficient description of the correlation energies. However, coupled-cluster calculations beyond CCSDT have only been generally possible for small molecules. The use of basis set beyond a polarized valence double-ζ basis does not improve sensibly the results in the evaluation of corrections for higher order excitations. Contribution of quadruple excitation (as difference between CCSDTQ and CCSDT) converges rapidly with the basis set and the ΔE^{CCSDTQ} term accounting for the effects of higher-level correlation effects beyond CCSD(T) is estimated with the cc-pVDZ basis set in the frozen core approximation:

$$\Delta E^{CCSDTQ} = E^{CCSDTQ}_{cc-pVDZ}(fc) - E^{CCSDT}_{cc-pVDZ}(fc). \tag{2.32}$$

Zero-point vibrational energies for all species apart from atoms are determined from anharmonic force fields calculated at the all electron CCSD(T) level of theory with the cc-pVQZ basis sets:

$$\Delta E_{ZPE} = \sum_i \frac{\omega_i}{2} + \sum_{i \geq j} \frac{\chi_{ij}}{4}, \tag{2.33}$$

where ω_i are the harmonic frequencies and χ_{ij} the anharmonicity constants (see Section **2.3**). As anticipated above, ΔE_{DBOC} accounts for the contribution of the nuclear kinetic energy operator. The simple potential energy surface picture of a molecule is maintained although this surface becomes mass dependent. For atoms, this correction—known as the diagonal Born–Oppenheimer correction (DBOC) accounts for the finite mass of the nucleus, and it can be calculated as:

$$\Delta E_{DBOC} = \langle \psi_e(\mathbf{r};\mathbf{R}) | \hat{T}_n | \psi_e(\mathbf{r};\mathbf{R}) \rangle, \tag{2.34}$$

where $\hat{T}_n$ is the nuclear kinetic energy operator and ψ_e ($\mathbf{r}$;$\mathbf{R}$) is the normalized electronic wavefunction obtained at the set of nuclear positions $\mathbf{R}$. Despite its simple form, it is not straightforward to calculate the DBOC. Within the HEAT protocol, the HF-SCF wavefunctions with aug-cc-pVTZ are considered sufficiently accurate. The remaining two terms in Equation (2.28) are ΔE_{SO} and ΔE_{Rel}. A detailed description of the implications of the non-relativistic theory in the HEAT protocol and further variation of the original procedure in the protocol can be found in literature[94–96].

The HEAT protocol usually achieves a sub-chemical accuracy of ~ 1 kJ/mol for thermochemical data. Despite the accuracy, we cannot indulge on the computational effort associated to the CCSD(T) and CCSDTQ calculations which are generally possible only for small systems. More recent composite methods take inspiration from the HEAT protocol but aim at reducing the computation

effort. The main idea is to make use of the accuracy of methods such as CCSD(T) (same as HEAT) reducing the demanding cost that comes with large basis sets to extend the applicability to small to medium size systems. In the third paper published about the HEAT protocol, Harding *et al.* explored its performances in comparison with different modified versions[96]. What is interesting from their study is that the more sophisticated treatments of individual contributions do not necessarily improve results. To move toward larger systems, ΔE_{Rel}, ΔE_{DBOC} and ΔE_{SO} can be dropped. The errors arising from such approximation are usually quite small and the bottleneck, as already mentioned, still lies in the extrapolation of the correlation energy. Stepping a bit further, the energy can be expressed as follows:

$$E_{CBS} = E_{CBS}^{HF-SCF} + \Delta E_{CBS}^{corr} + \Delta E_{CV}. \tag{2.35}$$

In the above equation the HF-SCF and the correlation energy are obtained correlating valence electrons only and the core correlation is included separately as difference between all-electron (ae) and frozen core energy:

$$\Delta E_{CV} = E^{CCSD(T)}(ae) - E^{CCSD(T)}(fc). \tag{2.36}$$

This approach is computationally cheaper and the errors coming from such procedure are usually smaller than 0.5 kJ/mol which is acceptable for many applications. Such procedure is labelled as CCSD(T)/CBS+CV and it has been implemented as a black-box procedure in the CFOUR software[106].

So far, the scheme we have presented requires two CCSD(T) calculations with basis sets of increasing size in order to extrapolate the correlation energy at the basis set limit. This implies a calculation at the CCSD(T)/cc-pVQZ level that can become prohibitively expensive for medium size systems. Hence, extrapolation schemes have been proposed in which the CCSD(T) calculations with larger basis set are replaced by MP2[107]. This scheme has been found to be successful and has been extensively used, for example, in providing reference energies for

weak intermolecular interactions[108,109]. Among the others, we will discuss the cheap scheme (ChS) which has been successfully applied to thermodynamics and kinetics for processes of astrochemical and environmental relevance. According to the ChS the total electronic energy is:

$$E_{ChS} = E_{CBS}^{HF-SCF} + E_{corr}^{CCSD(T)/VTZ} + \Delta E_{CBS}^{MP2_{corr}} + \Delta E_{CV}^{MP2}, \tag{2.37}$$

where:

$$\Delta E_{CBS}^{MP2_{corr}} = E_{CBS}^{HF-SCF} + \Delta E_{CBS}^{MP2_{corr}} - E_{VTZ}^{MP2_{corr}}. \tag{2.38}$$

The Hartree-Fock CBS limit energy (E_{CBS}^{HF-SCF} in Equation (2.38)) is retrieved from the extrapolation formula in Equation (2.29) by using the Dunning's correlation-consistent basis sets cc-pVXZ with X=T, Y=Q and Z=5.

The extrapolated correlation energy at MP2 level ($E_{CBS}^{MP2_{corr}}$) is obtained employing the same extrapolation procedure as in Equation (2.29) with cc-pVTZ and cc-pVQZ basis sets.

Recently, a new version of the ChS, termed jun-Cheap (jun-ChS) has been proposed in which the cc-pVTZ and cc-pVQZ basis sets are replaced by the calendar augmented jun-cc-pVXZ counterparts proposed by Truhlar and collaborators[90].

2.1.4.2 Composite schemes for geometry extrapolation

The procedures introduced so far refer to the electronic energy extrapolation. Composite schemes can be extended to properties other than energy. The reliability of the additivity approximation depends on the specific property and in particular, it becomes much less accurate the less the investigated property depends on the energy. Of course, since the nature of such assumption is purely speculative any extension of an extrapolation procedure to a new property requires a proper validation. Due to the influence of geometry on the resulting energy, the correct determination of accurate structural parameters is of paramount importance. As such, Gauss *et al.* proposed an additivity scheme analogous to the one discussed for energy at the energy gradient level,

the so-called *gradient scheme* [110,111]. Following the same reasoning that brought to the definition of Equation (2.27) we can define a similar expression for the energy gradient which on the basis of geometry optimization is:

$$\frac{dE_{tot}}{dx} = \frac{dE_A^{CCSD(T)}}{dx} + \frac{d\Delta E_B^{CCSDT}}{dx} + \frac{d\Delta E_C^{CCSDTQ}}{dx} + \frac{d\Delta E_D^{core}}{dx}. \tag{2.39}$$

The terms on the right hand side are calculated with progressively smaller basis sets for highly demanding calculations. In the original work, A=cc-pV6Z, B=cc-pVTZ, C=cc-pVDZ and D=cc-pCVTZ. For most of the calculations, the third term accounting for quadruple excitations can be neglected and using extrapolation procedure to account for the basis set truncation error, we get the final expression:

$$\frac{dE_{tot}}{dx} = \frac{dE_{CBS}^{HF-SCF}}{dx} + \frac{d\Delta E_{CBS}^{CCSD(T)}}{dx} + \frac{d\Delta E^{core}}{dx}. \tag{2.40}$$

The first two terms on the right hand side account respectively for the electronic and correlation energy, both extrapolated to the CBS limit, and the third term takes into account core correlation. The same formulas discussed for the energy extrapolation are used for the evaluation of such terms. The *gradient scheme* reaches accuracy in the range of 0.01 pm for bond lengths.
While the gradient scheme is theoretically well justified, a *geometry scheme* with the direct extrapolation of structural parameters has been proposed. Here, multiple optimizations are required to estimate the CBS limit.

Literature is full of works demonstrating the validity of such kind of approach [112]. In the attempt to generalise such procedure to systems of different size, the ChS for geometry has also been validated and the best-estimate of a given structural parameter, r_{best}, can be obtained as:

$$r_{best} = r^{CCSD(T)/pVTZ} + \Delta r_{CBS} + \Delta r_{CV}[+\Delta r_{aug}], \tag{2.41}$$

with

$$\Delta r_{CBS} = \frac{Y^3 r^{MP2/pVXZ} - X^3 r^{MP2/pVXZ}}{Y^3 - X^3} - r^{MP2/pVXZ}, \tag{2.42}$$

$$\Delta r_{CV} = r^{MP2/pCVTZ,a.e.} - r^{MP2/pCVTZ,f.c.}, \tag{2.43}$$

analogous terms to those described for the energy evaluation and retrieving corrections for the structural parameter evaluation at the CBS level and core-valence contributions.

$$\Delta r_{aug} = r^{MP2/aug-cc-pVTZ,f.c.} - r^{(MP2/pVTZ,f.c.)}. \tag{2.44}$$

The last term (Δr_{aug}) should recover the errors due to the dropping of diffuse function when switching from aug-cc-pVXZ to cc-pVXZ basis set. However, such term is empirical and its reliability needs to be properly investigated [113].

2.1.5 Density Functional Theory

Density Functional Theory (DFT) lies its roots in the proof by Hohenberg and Kohn that the ground state properties are uniquely defined by the electron density (ρ) [114]. The potential, which describes the interaction between nuclei and electrons, implies one and only electron density which in turn contains all the information required for the description of our chemical system. While wavefunction methods depends on $4N$ coordinates (three spatial coordinates plus one spin coordinate for each electron), the electron density depends on three spatial coordinates. The improvements in terms of computational effort are straightforward. So, while the complexity of the wavefunction increases exponentially with the system size, the scaling of DFT boosted a rapid growth in the usage of electron density based methods. Research fields like solid state chemistry and bio-inspired simulations did benefit most from DFT since they often manage large systems. As the theory name suggests, the terms in the total

energy can be described as functionals of the electron density. Within the BOA the total DFT energy is given by the following expression:

$$E_e[\rho(r)] = T[\rho(r)] + V_{ne}[\rho(r)] + J[\rho(r)] + K[\rho(r)], \tag{2.45}$$

where $T[\rho(r)]$ is the kinetic energy of electrons, $V_{ne}[\rho(r)]$ is the nuclear-electron attraction energy, $J[\rho(r)]$ is the classical electron-electron repulsion energy and $K[\rho(r)]$ is the non-classical (quantum) electron-electron interaction energy. While expressions for $V_{ne}[\rho(r)]$ and $J[\rho(r)]$ are well-known, the big issue in such picture is the absence of an explicit functional form describing the other two terms. Most of the research in the development of DFT methods aim at the definition of optimal expressions for such terms, and over the years it has led to a wide variety of density functionals [115–117]. This section is not going to be a detailed description of the DFT formalism, rather, the intent is to provide a brief summary of the density functionals evolution along the years with the specific intent to make easier the understanding of the methods selected in this thesis. The first formalism for the unknown functionals dates back to 1930 with the Thomas-Fermi-Dirac (TFD) model. It approximates the system to a uniform electron gas made of non interacting particles. The expressions for $T[\rho(r)]$ and $K[\rho(r)]$ are here reported:

$$T[\rho] = C_F \int \rho^{5/3}(r) dr, \tag{2.46}$$

$$K[\rho(r)] = -C_x \int \rho^{4/3}(r) dr, \tag{2.47}$$

with $C_F = \frac{3}{10}(3\pi^2)^{(2/3)}$ and $C_x = \frac{3}{4}(\frac{3}{\pi})^{1/3}$. The assumption of non-interacting particles is quite strong and errors can vary in the range 15 —50 %. Researchers working in this field have struggled looking for a better description of both $T[\rho(r)]$ and $K[\rho(r)]$. Despite their efforts, DFT at that stage cannot compete with the wavefunction methods. If detailed functional expressions could be

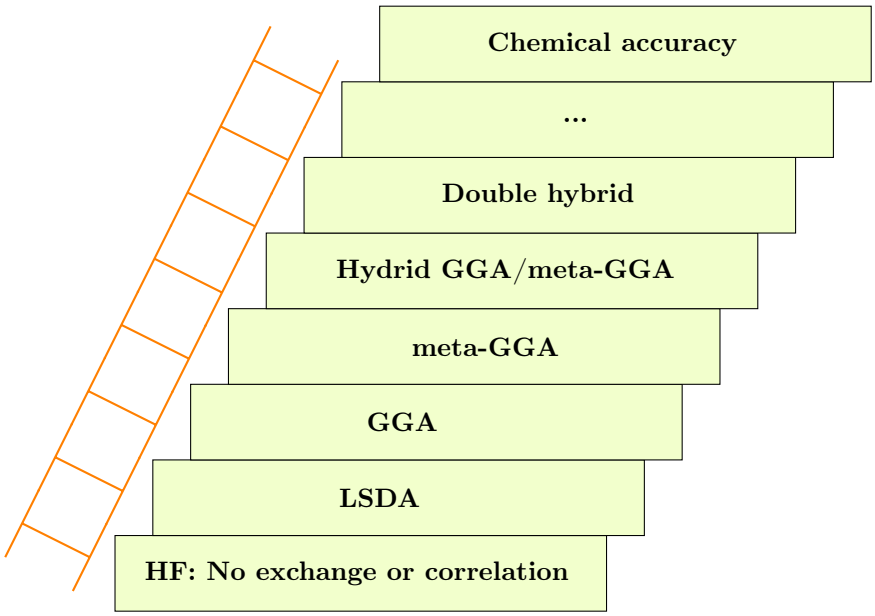

Figure 2.2 Jacob's ladder indicating the increasingly accurate methods for unknown exchange-correlation functional of DFT.

derived a full orbital-free DFT procedure could be finally developed. However, to circumvent the TFD model inaccuracy the Kohn-Sham (KS) method has introduced orbitals. In the KS approach the kinetic energy term is split in two components[118]. The first one can be calculated exactly since it is the kinetic energy of a system of non-interacting particle:

$$T_s = \sum_{i=i}^{N_{elec}} \langle \phi_i | -\frac{1}{2}\nabla^2 | \phi_i \rangle. \tag{2.48}$$

The difference between T_s and the exact kinetic energy is included as an additional correction term in $K[\rho(r)]$ which together are referred to as the *exchange-correlation* term (E_{xc}).

Many expressions for the exchange-correlation functional have been developed. To be cited is the pioneering work of Perdew and Becke in the development of non-empirical and semi-empirical density functionals. The increasing accuracy of DFT functionals is pictured in the Perdew's methaphorical Jacob's ladder representation (see Figure **2.2**) which summarises the improvements of DFT till the reach of the heaven of chemical accuracy[119].

Each rung of the ladder above the HF contains new physical terms which contribute to the development of more accurate models.

1. First rung. The simplest method is the local spin-density approximation (LSDA). These functionals depend exclusively on the electron density. They are exact for a uniform electron gas but fail to predict a lot of molecular properties. The poor description of the electron density in LSDA results in an erroneous prediction of structural parameters with an over binding also of weakly bound compounds.

2. Second rung. The first improvements to LSDA functionals comes from the introduction of the density gradient ($\nabla\rho(r)$) which recovers part of the lost description of the electron density. This approach is called Generalized Gradient Approximation (GGA)[120] and popular exchange functionals are B88[121], PBE[120], revPBE[122], PW91[123] while popular GGA correlation functionals are P86[124], LYP[125], PW91[123], PBE[120]. These components can be combined to form GGA exchange-correlation functionals. The most popular one is PBE exchange-correlation functional resulting from the PBE exchange and the PBE correlation term. As opposite to combining them, semi-empirical exchange-correlation functionals have also been parametrised (it is the case of HCTH suite of functionals and many others)[126,127]. Many other GGA functionals have been developed and described in the literature. In chemistry, moving from LDA to GGA brought a massive improvement of roughly an order of magnitude in binding energies.

3. Third rung. Meta-GGA functionals introduce either the kinetic energy density or the Laplacian of the density in the functional expression[128,129].

4. Fourth rung. Hybrid GGA/meta-GGA. DFT functionals introduced so far are affected by the self interaction error (SIE). The Hartree-Fock method is free of self-interaction error because the Coulomb and the exchange terms cancel each other for an electron interacting with itself. In DFT the exact HF exchange is substituted by the exchange-correlation functional (E_{xc})

making most functionals affected by the SIE. Hybrid functionals face this issue by replacing the local exchange functional (E_x^{DFT}) with the exact exchange functional (the HF one, E_x^{HF}) while still using a local correlation functional (E_c^{DFT}):

$$E_{xc} = c_x E_x^{HF} + (1 - c_x) E_x^{DFT} + E_c^{DFT}, \tag{2.49}$$

where c_x is the percentage of exact HF exchange used. The first of such functionals was developed in the early '90 by Becke: the BHLYP introduces 50 % of HF exchange in the functional expression [130]. Almost simultaneously B3LYP, one of the most popular functionals ever, has been developed [125,131]. With the intent to improve performaces with respect to common GGAs, a degree of parametrization in the formulation of functionals has been introduced [132]. An example is the implementation in the Gaussian software of B3LYP:

$$E_{xc}^{B3LYP} = 0.2E_x^{HF} + 0.8E_x^{LDA} + 0.72\Delta E_x^{B88} + 0.81E_c^{LYP} + 0.19E_c^{VWN}, \tag{2.50}$$

where the constants have been determined by Becke *via* fitting to the G1 molecule set.

The formula above can be extended to meta-GGAs to give global hybrid meta-GGA density functionals. Truhlar has developed a number of global hybrid meta-GGAs density functionals, including PW6B95 [133], M06 [134], M06-2X [134] and MN15 [135]. The fraction of exact exchange across these functionals varies from 27 % (M06) to 54 % (M06-2X). Other noteworthy global hybrid meta-GGAs include the semi-empirical BMK [136] functional with 42 % exact exchange and it has been proposed for applications to chemical kinetics. Despite the undeniable improvements of such functionals in their local counterparts for bonded interactions and kinetics, they only

partially address the self-interaction issue. A more rigorous approach to this problem is through range-separation, where the exact exchange contribution is split into a short-range component ($E_{x,sr}^{HF}$) and a long-range component ($E_{x,lr}^{HF}$)[137,138]. $E_{x,sr}^{HF}$ can be optionally scaled ($c_{x,sr}$) to give a non-zero fraction of short-range exact exchange, while the scaling coefficient ($c_{x,lr}$) of $E_{x,lr}^{HF}$ is usually set to one to ensure that the exchange functional is one-electron SIE-free in the long range. The corresponding local exchange functional should also be partitioned in the same way, with the short- and long-range components scaled by $(1 - c_{x,sr})$ and $(1 - c_{x,lr})$ respectively. The functional form for a typical range-separated hybrid (RSH) functional is:

$$E_{xc}^{RSH} = c_{x,sr}E_{x,sr}^{HF} + c_{x,lr}E_{x,sr}^{HF} + (1 - c_{x,sr})E_{x,sr}^{DFT} + (1 - c_{x,lr})E_{x,lr}^{DFT} + E_{c}^{DFT}. \tag{2.51}$$

Among range-separated hybrid GGA functionals there are the semi-empirical ωB97 and ωB97X functionals from Chai and Head-Gordon, which have 0 % and 15.77 % short-range exact exchange, respectively, and tend to 100 % exact exchange in the long-range[139,140].

5. Fifth rung. Double-hybrid functionals rely on a mixing of standard generalized gradient approximations (GGAs) for exchange and correlation with Hartree-Fock (HF) exchange and a perturbative second-order correlation part (PT2) that is obtained from the Kohn-Sham (GGA) orbitals and eigenvalues[141].

   The most common example is the B2PLYP functional[142].

$$E_{xc}^{B2PLYP} = aE_{HF} + (1 - a)E_{x}^{GGA} + bE_{c}^{GGA} + (1 - b)E_{c}^{PT2}. \tag{2.52}$$

   Several reparametrizations were subsequently proposed. Different GGA and meta-GGA exchange and correlation functionals have been considered

along the years. It turns out, pretty much any good exchange functional is adequate. However, the LYP correlation functional outperforms all the others. The further evolution is that of spin-component scaling double-hybrids functionals where the same spin and opposite spin perturbational terms are multiplied by different scaling coefficients (c_s and c_o, respectively). The same-spin correlation energy (also called triplet or parallel spin) is associated with long-range, nondynamical effects; opposite spin (also singlet or antiparallel spin) is associated with short-range, dynamical correlation. DSD-DFT functionals have been developed with this scheme and the general expression is:

$$E_{xc}^{DSD-DFT} = (1 - c_x)E_x^{DFT} + c_x E_x^{HF} + c_c E_c^{DFT} + c_o E_o^{DFT} + c_s E_s + c_D E_D, \tag{2.53}$$

with E_D being the dispersion correction (following in this text). DSD-BLYP offers significantly enhanced robustness and depending on the properties results can improve up to 14 %. A downside of DSD-DFT compared to the simpler DHs is the introduction of two additional empirical parameters. Multiple exchange and correlation functionals have been combined in order to identify the best performers and all of them involved the P86 correlation functional in conjunction with PBE, mPW, and HSE exchange functionals. DSD-PBEP86 has been selected on availability grounds, as PBE is available in all major quantum-chemistry packages[143–145].

There is still something that needs to be addressed in our discussion. Local and hybrid exchange-correlation functionals cannot properly account for long-range dynamic correlation. LSDA, GGA, meta-GGA, and hybrid functionals exhibit an exponential long-range decay, instead of the proper R^{-6} decay. Several approaches to tackle this issue have been proposed. Among them, the DFT-D method developed by Grimme, relies on a damped, atom–pairwise dispersion interaction potential that is added to the standard DFT energy. Four generations

of DFT-D have been developed by Grimme thus far: D1, D2, D3 and D4[146–148]. The D3 can be used either with the original damping function, D3(0), or the Becke-Johnson damping function, D3(BJ). The main advantage of DFT-D3 is its computational efficiency and robustness across the periodic table, which makes this approach particularly valuable for the treatment of large systems.

2.2 Kinetics

The models described in the previous section offer accurate procedures for the prediction of the thermodynamics of a chemical process. Beyond the structures and the energetics, thoroughly selected methods are essential to retrieve reaction rates. For reactions in the cold ISM a detailed investigation of the thermodynamics along with reaction rates is mandatory. In this section we will provide a short description of the kinetic methods applied to the test cases that can be found in the next chapter.

2.2.1 Transition State Theory

Transition State Theory (TST) is a widely employed theoretical framework for calculating reaction rates. TST considers the concept of a transition state, which represents an intermediate state between reactants and products with the highest potential energy[149].

TST considers the reaction as a dynamical process that involves the passage through a transition state region on the potential energy surface. The rate constant (k) for a unimolecular reaction can be calculated as:

$$k = \frac{k_B T}{h} \exp(-\frac{\Delta G^{\ddagger}}{RT}). \tag{2.54}$$

In this equation, k_B is the Boltzmann constant, T is the temperature in Kelvin, h is the Planck constant, $\Delta G^{\ddagger}$ is the Gibbs free energy of activation, R is the gas constant, and $\Delta G^{\ddagger}/RT$ represents the dimensionless activation energy. The

most interesting achievement of TST is that the calculation of the partition functions of reactants and transition state leads to the reaction rate [150]. TST can only be applied if the structure of the transition state and its vibrational levels are known, because they are used for the calculation of the rotational and vibrational partition function of the transition state respectively. TST is widely applied in the field of astrochemistry to investigate the kinetics of reactions occurring in the ISM. By providing insights into the energetics and barriers involved in chemical reactions, TST helps in understanding the formation and destruction mechanisms of molecules in space, contributing to our knowledge of astrochemical processes and the evolution of interstellar environments. It's worth noting that there are also modified versions of TST, such as variational transition state theory (VTST) [151] and tunneling correction methods [152], which account for quantum mechanical effects and provide more accurate descriptions of reaction rates, especially for reactions with heavy atoms or at low temperatures.

2.2.2 Rice–Rampsberger–Kassel–Marcus theory

Rice-Ramsperger-Kassel-Marcus (RRKM) theory is an extension of TST that provides a more detailed treatment of unimolecular reactions [153,154]. In this treatment, an energised molecule, A^*, is distinguished from an activated molecule, $A^\ddagger$. An energized molecule refers to a molecule that possesses excess energy compared to its ground state. However it should acquire sufficient vibrational vibrational energy before it could be converted to an activated molecule, and consequently its rate of conversion into $A^\ddagger$ may depend on the critical energy for the conversion, ϵ^*. An activated molecule specifically refers to a molecule that has surpassed the energy barrier, or activation energy, required for a particular chemical reaction to occur. Marcus extended the original RRK formalism in such a way that the individual vibrational frequencies of the energised species and activated complexes are considered explicitly including the rotations and different normal modes of vibration, together with zero-point energy contributions. The expression for the rate of conversion into products, k, is:

$$k\left(\varepsilon^{*}\right)=\frac{s^{\ddagger}}{h}\frac{\sum P\left(\varepsilon^{*}_{\text{active}}\right)}{N\left(\varepsilon^{*}_{\text{active}}\right)F_{\text{r}}}, \tag{2.55}$$

where $s^{\ddagger}$ is a statistical factor and $\Sigma P\left(\varepsilon^{*}_{\text{active}}\right)$ the number of vibrational-rotational quantum states of the activated molecules, corresponding to all energies up to and including $\varepsilon^{\ddagger}_{\text{active}}$ where only the so-called *active modes* are considered. $N\left(\varepsilon^{*}_{\text{active}}\right)$ is the density of states having energy between ε^{*} and $\varepsilon^{*}+\mathrm{d}\varepsilon^{*}$, i.e., the number of states per unit energy range. Finally, F_{r} is a factor introduced to correct for rotations that may be different in the activated molecule and the energised molecule. By considering the density of states and energy transfer between different degrees of freedom, RRKM theory offers a more accurate description of the reaction rate. Rate coefficients for reactions within the RRKM framework can be obtained using the 1D master equation system solver (MESS) code[155].

2.2.3 Phase Space Theory

Phase Space Theory (PST) is particularly useful for studying barrierless elementary reactions. PST assumes that the interaction between the reacting fragments is isotropic and does not affect their internal motions[156]. This holds true when the highest energy barrier of a reaction is encountered when the fragments are considerably distant from each other, enabling them to exhibit unhindered rotational motion and undisturbed vibrational modes. PST has been used in computational kinetics applied to astrochemistry to explore the chemical processes occurring in the interstellar medium, where low-temperature phenomena dominate. Its ability to describe reactions with isotropic attractive potentials, incorporating long-range interactions and neglecting internal motions, makes PST a valuable tool for investigating the chemical complexity and evolution of interstellar environments.

2.3 Vibrational spectroscopy

The discovery of the chemical complexity of the ISM boosted an extension of the spectroscopic tools to investigate such chemical environment. Nowadays, experimental as well as computational spectroscopy play a pivotal role in the study and understanding of the ISM and planetary atmospheres. In this section the discussion starts by introducing the vibrational problem and its simplest resolution in the framework of the harmonic oscillator. Due to the indisputable relevance of anharmonicity in the treatment of vibrations we will then present second-order vibrational perturbation theory (VPT2) and introduce solutions to tackle the problem of resonances plaguing the perturbative expressions. Since the computational cost of anharmonic calculations scales rapidly with the system size we will present reduced-dimensionality (RD) strategies applicable to medium to large size systems.

2.3.1 The nuclear problem

This discussion starts considering the total energy of a molecule in the BOA [77]. As further approximation, we consider valid the Eckart-Sayvetz conditions [157,158] which allow the factorization of the nuclear wavefunction (Ψ) in terms of its translational (ψ^t) and vibro-rotational (ψ^{vr}) components:

$$\psi = \psi^t \psi^{vr} \approx \psi^t \psi^r \psi^v. \tag{2.56}$$

By setting the molecule orientation according to Eckart-Sayvetz conditions the vibro-rotational coupling is minimized but not null. Hence, according to the needed accuracy some treatment of such term may be required. In what follow, we will consider the last approximation above by neglecting the vibro-rotational coupling. In its simplest form the vibrational potential energy can be described through a harmonic potential:

$$V(\boldsymbol{R}) = V(\boldsymbol{R}_{eq}) + \frac{1}{2}\sum_{A,B}\frac{\partial^2 V}{\partial \boldsymbol{R}_A \partial \boldsymbol{R}_B}(\boldsymbol{R}_A - \boldsymbol{R}_{A,eq})(\boldsymbol{R}_B - \boldsymbol{R}_{B,eq}), \tag{2.57}$$

where $\boldsymbol{R}_i$ and $\boldsymbol{R}_{i,eq}$ are the Cartesian coordinates vectors and the equilibrium Cartesian vectors of the ith nucleus and the summation runs over all the nuclei. In the above equation, the vibration problem requires the resolution of a $3N$ dimensional wavefunction. It is convenient to cast the problem in such a way to decouple the normal modes, with the vibrational wavefunction built as the product of one-dimensional wavefunctions. To make this possible, we need to switch to a new set of coordinates. Let's consider the mass-weighted matrix of force constants ($\boldsymbol{F}$) which is by definition symmetric, so it can be written in the form,

$$\boldsymbol{F} = \boldsymbol{L}\boldsymbol{\Omega}\boldsymbol{L}^T, \tag{2.58}$$

with $\boldsymbol{L}$ an orthogonal matrix, and $\boldsymbol{\Omega}$ a diagonal matrix with the eigenvalues of the mass-weighted force constants matrix. The matrix of eigenvectors $\boldsymbol{L}$ can be used to define a new set of coordinates as,

$$\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{M}^{1/2}\left(\boldsymbol{R} - \boldsymbol{R}_{\mathrm{eq}}\right), \tag{2.59}$$

where $\boldsymbol{M}$ is the diagonal matrix of atomic masses. In this new framework, 6 of the coordinates (5 for linear molecules) have null eigenvalues, as related to the translational and rotational motions, and are not included. The remaining coordinates (3N − 6 or 3N − 5) are the vibrational modes. With these coordinates, the vibrational Hamiltonian is decoupled in the following form:

$$\hat{H}^v(\boldsymbol{Q}) = \sum_i \hat{H}_i^v(Q_i) = \frac{1}{hc}\sum_i\left[\frac{\hbar^2}{2}P_i + \frac{\Omega_{ii}}{2}Q_i\right], \tag{2.60}$$

where we have introduced the conjugate momenta, $P_i = -i\hbar, \partial/\partial Q_i$, and, $\Omega_{ii} = 4\pi^2 c^2 \omega_i^2$, with ω_i vibrational wavenumber.

Dimensionless normal coordinates $\boldsymbol{q}$ and $\boldsymbol{p}$ are usually introduced and related to the corresponding dimensional counterparts by:

$$q_i = \sqrt{\frac{2\pi c\omega_i}{\hbar}} Q_i \qquad p_i = \sqrt{\frac{1}{hc\omega_i}} P_i. \tag{2.61}$$

Hence, vibrational Hamiltonian within the harmonic approximation takes the following form:

$$\hat{H}_v(\boldsymbol{q}) = \sum_i \hat{H}_i^v(q_i) = \frac{1}{2} \sum_i \omega_i (p_i^2 + q_i^2), \tag{2.62}$$

where the summation runs over the 3N - 6 (3N - 5 for linear molecules) normal modes. The energy of the vibrational levels (ϵ_r) is:

$$\epsilon_r = \sum_i (v_i + \frac{1}{2})\omega_i, \tag{2.63}$$

where v_i is the vibrational quantum number associated to mode i taking values 0, 1, 2, ...

The harmonic approach here presented does not reproduce correctly the potential energy surface. In particular, as the distance increases the potential energy tends to the molecular dissociation energy. This behaviour is not accounted for in the harmonic description. Moreover, the model so far discussed considers vibrational levels equally spaced while the spacing progressively diminishes as the vibrational energy increases. As consequence the harmonic approach tends to overestimate the vibrational energy and such bad behaviour gets worse the higher the vibrational energy is.

2.3.2 Second-order Vibrational Perturbation Theory

The stated limitations in the harmonic approximation prevent to achieve quantitative accuracy. Descriptions of molecular vibrations beyond the harmonic approximation are necessary. The deviation from the harmonic model is called anharmonicity. Methods of treating anharmonicity include vibrational pertur-

bation theory (VPT) and variational approaches, both based on Taylor series expansions of the potential energy. The vibrational self-consistent field method is an alternative, which variationally improves the zeroth-order wavefunctions. Further corrections are then typically made to the vibrational wavefunctions, which are analogous to the MP2, CI, and CC treatments described above for electron correlation. VPT2 is a widely used method for dealing with anharmonicity in molecular vibrations[159]. As pointed out in Section **2.1.2**, according to perturbation theory the Hamiltonian is separated into a zeroth-order part, the exact eigenfunctions of which are known, and a perturbation that is assumed to be small. The starting point of VPT is based on the Watson's Hamiltonian which makes use of dimensionless normal coordinates[160]. In a similar manner to Equation (2.4) in Section **2.1.2**, the vibro-rotational Hamiltonian is described through two terms: the zeroth-order part corresponds to the harmonic oscillator Hamiltonian, while the perturbation is the anharmonic potential. The effects of vibrational angular momentum are also included, again as a perturbation.

$$H^{vr} = \frac{\hbar^2}{2hc}\sum_{\alpha,\beta}(J_\alpha - \pi_\alpha)\mu_{\alpha\beta}(J_\beta - \pi_\beta) - \frac{\hbar^2}{8hc}\sum_{\alpha}\mu_{\alpha\alpha} + \frac{1}{2}\sum_{i=1}^{M}\omega_i p_i^2 + V, \quad (2.64)$$

where J and π are the rotational and vibrational angular momentum operators and $\mu_{\alpha\beta}$ are the elements of the inverse of the modified inertia tensor, also called the rotational tensor. The summations run over $\alpha, \beta = a, b, c$, i.e. the principal axis of inertia in the Eckart frame and the M normal modes. The second term on the right hand side is known as the 'Watson term' and is null for linear molecules. Both the rotational tensor and the potential energy are expanded as Taylor series around the equilibrium geometry:

$$\begin{aligned}\mu_{\alpha\beta}(\boldsymbol{q}) = \mu_{\alpha\beta}(\boldsymbol{q}^{eq}) - \sum_{i=1}^{M}\mu_{\alpha\alpha}^{eq}\left(\frac{\partial I_{\alpha\beta}}{\partial q_i}\right)_{eq}\mu_{\beta\beta}^{eq}q_i \\ +\frac{3}{4}\sum_{\gamma}\sum_{i,j=1}^{M}\mu_{\alpha\alpha}^{eq}\left(\frac{\partial I_{\alpha\gamma}}{\partial q_i}\right)_{eq}\mu_{\gamma\gamma}^{eq}\left(\frac{\partial I_{\gamma\beta}}{\partial q_j}\right)_{eq}\mu_{\beta\beta}^{eq}q_i q_j + o(\boldsymbol{q}^3),\end{aligned} \quad (2.65)$$

$$V(\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{M}\omega_i q_i^2 + \frac{1}{6}\sum_{i,j,k=1}^{M}\left(\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k}\right)_{eq} q_i q_j q_k + \frac{1}{24}\sum_{i,j,k,l=1}^{M}\left(\frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l}\right)_{eq} q_i q_j q_k q_l + o(\boldsymbol{q}^5), \tag{2.66}$$

where $I_{\alpha,\beta}$ are the inertia tensor elements. The third-order and fourth-order derivatives in Equation (2.66) are usually referred as cubic (f_{ijk}) and quartic (f_{ijkl}) force constants, respectively. Analytic implementations for the evaluation of these terms are not available and hence they are computed numerically. Let's consider the case in which the system is in the rotational ground state. In such case, the Hamiltonian can be reformulated by neglecting the contributions from the angular momentum in Equation (2.64) and truncating Equation (2.65) to the zeroth-order. This approach still includes some vibro-rotational coupling coming from the following term:

$$\frac{\hbar}{2hc}\sum_{\alpha\beta=x,y,z}\mu_{\alpha\beta}^{eq}\pi_\alpha\pi_\beta = \sum_{\alpha=x,y,z}B_\alpha^{eq}\sum_{i,j,k,l=1}^{M}\zeta_{ij,\alpha}\zeta_{kl,\alpha}\sqrt{\frac{\omega_j\omega_l}{\omega_i\omega_k}}q_i p_j q_k p_l, \tag{2.67}$$

where B_α^{eq} is the equilibrium rotational constant along the $\alpha = a, b, c$ axis and $\boldsymbol{\zeta}$ is the Coriolis matrix of the vibro-rotational coupling. By using Equation (2.65) and Equation (2.66) to the 0-th and 4-th order, respectively, the vibrational Harmiltonian is:

$$H^v = \frac{1}{2}\sum_i \omega_i(p_i^2 + q_i^2) + \frac{1}{6}\sum_{i,j,k} f_{ijk}q_i q_j q_k + \frac{1}{24}\sum_{i,j,k,l} f_{ijkl}q_i q_j q_k q_l + \sum_\alpha B_\alpha^{eq}\sum_{i,j,k,l}\zeta_{ij,\alpha}\zeta_{kl,\alpha}\sqrt{\frac{\omega_j\omega_l}{\omega_i\omega_k}}q_i p_j q_k p_l. \tag{2.68}$$

The vibrational energies can be obtained from the above Hamiltonian using either a variational[161,162] or perturbative approach[159,163,164]. In what follows we will only discuss the VPT2 since it has been used in this work as it represents

a good trade-off between accuracy and computational cost.

VPT2 provides one analytical expression for the calculation of the energy of any vibrational states. The energy of a vibrational lever (ϵ_v) is:

$$\epsilon_v = \epsilon_0 + \sum_i v_i\omega_i + \sum_i \sum_{j\geq i} \chi_{ij}[v_i v_j + \frac{1}{2}(v_i + v_j)], \tag{2.69}$$

with ϵ_0 being the anharmonic zero-point vibrational energy:

$$\begin{aligned}\varepsilon_0 = &\sum_i \frac{\omega_i}{2} + \sum_{i,j} \frac{\mathrm{f}_{iijj}}{32} - \sum_{i,j,k} \left[\frac{\mathrm{f}_{iik}\mathrm{f}_{jjk}}{32\omega_k} + \frac{{\mathrm{f}_{ijk}}^2}{48\left(\omega_i + \omega_j + \omega_k\right)} \right] \\ &- \sum_\tau \frac{B_\tau^{\mathrm{eq}}}{4} \left[1 - \sum_{i<j} \left\{\zeta_{ij,\tau}\right\}^2 \frac{\left(\omega_i - \omega_j\right)^2}{\omega_i\omega_j} \right],\end{aligned} \tag{2.70}$$

and $\boldsymbol{\chi}$ the matrix of anharmonic constants whose diagonal and off-diagonal terms can be cast in the following way:

$$16\chi_{ii} = \mathrm{f}_{iiii} - \frac{5{\mathrm{f}_{iii}}^2}{3\omega_i} - \sum_{j\neq i} \frac{\left(8\omega_i^2 - 3\omega_j^2\right)\mathrm{f}_{iij}^2}{\omega_j\left(4\omega_i^2 - \omega_j^2\right)},$$

$$\begin{aligned}4\chi_{ij} = &\mathrm{f}_{iijj} - \frac{2\omega_i\mathrm{f}_{iij}^2}{\left(4\omega_i^2 - \omega_j^2\right)} - \frac{2\omega_j\mathrm{f}_{ijj}^2}{\left(4\omega_j^2 - \omega_i^2\right)} - \frac{\mathrm{f}_{iii}\mathrm{f}_{ijj}}{\omega_i} - \frac{\mathrm{f}_{jjj}\mathrm{f}_{iij}}{\omega_j} \\ &+ \sum_{k\neq i,j} \left[\frac{2\omega_k\left(\omega_i^2 + \omega_j^2 - \omega_k^2\right)\mathrm{f}_{ijk}^2}{\left(\omega_i + \omega_j + \omega_k\right)\left(\omega_i - \omega_j - \omega_k\right)\left(\omega_j - \omega_i - \omega_k\right)\left(\omega_k - \omega_i - \omega_j\right)} - \frac{\mathrm{f}_{iik}\mathrm{f}_{jjk}}{\omega_k} \right] \\ &+ \frac{4\left(\omega_i^2 + \omega_j^2\right)}{\omega_i\omega_j} \sum_\tau B_\tau^{\mathrm{eq}} \left\{\zeta_{ij,\tau}\right\}^2.\end{aligned} \tag{2.71}$$

The zero-point vibrational energy is useful to compute reaction energies, but does not play a role in the transition energies for the simulation of vibrational spectra, which can be obtained as it follows:

$$\nu_v = \varepsilon_v - \varepsilon_0 = \sum_i v_i\omega_i + \sum_i \sum_{j\geq i} \chi_{ij} \left[v_i v_j + \frac{1}{2}\left(v_i + v_j\right) \right]. \tag{2.72}$$

A critical issue of VPT2 is the potential presence of resonant terms, the so-called Fermi resonances (FRs). Such terms originate from singularity in the Equation (2.71) and are classified in two groups: Fermi resonances of type I $(\omega_i \approx 2\omega_j)$ and of type II $(\omega_i \approx \omega_j + \omega_k)$. The terms responsible for FRs of type I and II in Equation (2.71) are respectively:

$$\frac{\mathrm{f}_{iij}^2}{\left(4\omega_i^2 - \omega_j^2\right)}, \tag{2.73}$$

$$\frac{{\mathrm{f}_{ijk}}^2}{\left(\omega_i + \omega_j + \omega_k\right)\left(\omega_i - \omega_j - \omega_k\right)\left(\omega_j - \omega_i - \omega_k\right)\left(\omega_k - \omega_i - \omega_j\right)}. \tag{2.74}$$

The problem of resonances can be treated by removing from the perturbative expressions the resonant terms resulting in the so-called deperturbed second-order vibration perturbation theory (DVPT2). Different procedures have been proposed[165–167]. One possible procedure which has been used in this thesis as implemented in the Gaussian software[168] is based on the Martin's test[167] and it makes use of a two step evaluation procedure:

STEP 1: $\left|2\omega_i - \omega_j\right| \leq \Delta\omega^{1-2}$.

STEP 2: ${\mathrm{f}_{iij}}^4/256\left|2\omega_i - \omega_j\right|^3 \geq K^{1-2}$.

The first step evaluates which frequency needs to be discarded by considering the frequency difference. The second step includes an extra evaluation based on the value of the cubic force constant (f_{iij}). As a result, the capacity of identifying correctly the resonances, and ultimately the quality of the DVPT2 results, will depend on the chosen thresholds, $\Delta\omega^{1-2}$ and especially K^{1-2}. Terms flagged as resonant are then removed from the definition of χ. Of course, DVPT2 avoids a proper treatment of resonances by discarding potential resonant terms. These can be re-introduced a posteriori through a variational treatment on top of DVPT2 energies. This approach is known as generalized VPT2 (GVPT2) which

considers an interaction matrix with DVPT2 energies along the diagonal and coupling terms as off-diagonal entries[165,169]. The diagonalization of such matrix provides the correct vibrational energies and states.

2.3.3 Intensities

So far, the discussion has been focused on the definition of vibrational energies, however, when working with spectroscopy we cannot indulge on the importance of intensities which are a primary property of the vibrational spectrum. IR intensities can be defined in terms of wavenumber dependent absorption coefficients $\epsilon(\tilde{\nu})$.

$$\epsilon(\tilde{\nu}) = \frac{8\pi^3 N_A \tilde{\nu}}{3000\ln(10)hc(4\pi\epsilon_0)} \sum_s \rho_s \sum_f |\langle\mu\rangle_{sf}|^2 \delta(\tilde{\nu}_f - \tilde{\nu}_s - \tilde{\nu}) [\mathrm{dm}^3\ \mathrm{mol}^{-1}\ \mathrm{cm}^{-1}], \tag{2.75}$$

where $\tilde{\nu}$ is the wavenumber of the incident radiation, N_A is the Avogadro's number, ϵ_0 is the vacuum permittivity and the summations over s and f run over all initial and final states, respectively; ρ_s is the Boltzmann population of the initial state s, δ is the Dirac's delta function and $\langle\boldsymbol{\mu}\rangle_{sf}$ is the transition electric dipole moment. Within the BOA the latter can be written as

$$\langle\boldsymbol{\mu}\rangle_{f,s} = \frac{\langle\psi_s\,|\boldsymbol{\mu}|\,\psi_f\rangle}{(\sqrt{\langle\psi_f\mid\psi_f\rangle\,\langle\psi_s\mid\psi_s\rangle})^{1/2}}, \tag{2.76}$$

where $\boldsymbol{\mu}$ is the dipole moment, ψ_s and ψ_f are the wavefunction of the initial and final states, respectively, and the normalization condition is stated explicitly. As for vibrational energies, in which the effects of mechanical anharmonicity can be accounted through perturbation theory upon expanding the potential energy in a Taylor series, electrical anharmonicity ($\boldsymbol{\mu}^e$) can be introduced by expanding the dipole moment in a Taylor series of the normal coordinates around the molecule's equilibrium configuration:

$$\begin{aligned}\boldsymbol{\mu}^e(q) =& \boldsymbol{\mu}^e_{eq} + \sum_i \left(\frac{\partial \boldsymbol{\mu}^e}{\partial q_i}\right)_{eq} q_i + \frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 \boldsymbol{\mu}^e}{\partial q_i \partial q_j}\right)_{eq} q_i q_j + \\ & \frac{1}{6}\sum_{i,j,k}\left(\frac{\partial^3 \boldsymbol{\mu}^e}{\partial q_i \partial q_j \partial q_k}\right)_{eq} q_i q_j q_k + \circ\left(\boldsymbol{q}^4\right) \\ =& \boldsymbol{\mu}^e_0 + \sum_i \boldsymbol{\mu}^e_i q_i + \frac{1}{2}\sum_{i,j}\boldsymbol{\mu}^e_{ij} q_i q_j + \frac{1}{6}\sum_{i,j,k}\boldsymbol{\mu}^{\mathrm{e}}_{ijk} q_i q_j q_k + o\left(\boldsymbol{q}^4\right).\end{aligned} \tag{2.77}$$

By using harmonic oscillator wavefunction and truncating this expansion to the second-order yields the so-called double-harmonic approximation. As for the wavefunction also for the transition dipole moment are available both variational and perturbative approaches in order to account for the anharmonicity. We will focus on the VPT2. The full derivation of the VPT2 as applied to the dipole moment surface is quite cumbersome so we will just remark the main difference with respect to the derivation of the energy expression: the absence of a single expression for any transition. Indeed, different transitions (fundamentals, first overtones, second overtones, binary combinations, ...) do require different equations. For completeness, we report the final equations for fundamental and two-quanta vibrational transitions, respectively [170]:

$$\begin{aligned}&\langle \boldsymbol{\mu}^{\mathrm{e}}\rangle_{0,1_i} = \frac{1}{\sqrt{2}}\boldsymbol{\mu}^{\mathrm{e}}_i + \frac{1}{4\sqrt{2}}\sum_{j=1}^{N}\boldsymbol{\mu}^{\mathrm{e}}_{ijj} - \\ &\frac{1}{8\sqrt{2}}\sum_{j,k=1}^{N} f_{ijkk}\boldsymbol{\mu}^{\mathrm{e}}_j\left[\frac{1}{\omega_i+\omega_j} - \frac{1-\delta_{ij}}{\omega_i-\omega_j}\right] - \\ &\frac{1}{8\sqrt{2}}\sum_{j=1}^{N}\sum_{k=1}^{N}\left\{f_{ijk}\boldsymbol{\mu}^{\mathrm{e}}_{jk}\left(\frac{1}{\omega_i+\omega_j+\omega_k} - \frac{1}{\omega_i-\omega_j-\omega_k}\right)\right. \\ &\left.+\frac{2f_{jkk}}{\omega_j}\boldsymbol{\mu}^{\mathrm{e}}_{ij}\right\} + \frac{1}{2\sqrt{2}}\sum_{j,k=1}^{N}\left(\sum_{\tau=x,y,z} B^{\mathrm{eq}}_{\tau}\zeta_{ik,\tau}\zeta_{jk,\tau}\right)\times \\ &\boldsymbol{\mu}^{\mathrm{e}}_j\left\{\frac{\sqrt{\omega_i\omega_j}}{\omega_k}\left(\frac{1}{\omega_i+\omega_j} + \frac{1-\delta_{ij}}{\omega_i-\omega_j}\right) - \frac{\omega_k}{\sqrt{\omega_i\omega_j}}\left(\frac{1}{\omega_i+\omega_j}\right.\right. \\ &\left.\left.-\frac{1-\delta_{ij}}{\omega_i-\omega_j}\right)\right\} + \frac{1}{16\sqrt{2}}\sum_{j,k,l=1}^{N} f_{ikl}f_{jkl}\times\end{aligned}$$

$$
\begin{aligned}
&\boldsymbol{\mu}_j^{\mathrm{e}} \Bigg\{ \frac{4\omega_j \left(\omega_k + \omega_l\right) \left(1 - \delta_{ij}\right) \left(1 - \delta_{ik}\right) \left(1 - \delta_{il}\right)}{\left(\omega_j^2 - \omega_i^2\right) \left[\left(\omega_k + \omega_l\right)^2 - \omega_i^2\right]} \\
&+ \frac{\left(\omega_k + \omega_l\right) \left[\left(\omega_k + \omega_l\right)^2 - 3\omega_i^2\right] \delta_{ij} \left(1 + \delta_{ik}\right) \left(1 - \delta_{il}\right)}{\omega_i \left[\left(\omega_k + \omega_l\right)^2 - \omega_i^2\right]^2} \\
&+ \frac{4\omega_j \left(3\omega_k + 4\omega_i\right) \left(1 - \delta_{ij}\right) \left(1 - \delta_{ik}\right) \delta_{il}}{\omega_k \left(\omega_j^2 - \omega_i^2\right) \left(\omega_k + 2\omega_i\right)} \Bigg\} \\
&+ f_{ijk} f_{llk} \boldsymbol{\mu}_j^{\mathrm{e}} \Bigg\{ \frac{\delta_{ij}}{\omega_i \omega_k} \left(1 + \frac{2\delta_{ik}\delta_{il}}{9}\right) \\
&+ \frac{4\omega_j \left(1 - \delta_{ij}\right) \left(1 - \delta_{ik}\right) \left(1 - \delta_{il}\right)}{\omega_k \left(\omega_j^2 - \omega_i^2\right)} \\
&+ \frac{4\omega_j \delta_{ik} \left(1 - \delta_{ij}\right)}{\omega_i \left(\omega_j^2 - \omega_i^2\right)} \left(1 + \frac{2\delta_{ij}}{3}\right) \Bigg\} \\
&\langle \boldsymbol{\mu}^{\mathrm{e}} \rangle_{0,\left(1+\delta_{ij}\right)_i \left(1-\delta_{ij}\right)_j} = \frac{1}{2\left(1 + \delta_{ij}\right)} \times \\
&\left[\boldsymbol{\mu}_{ij}^{\mathrm{e}} + \frac{1}{2} \sum_{k=1}^{N} f_{ijk} \boldsymbol{\mu}_k^{\mathrm{e}} \left(\frac{1}{\omega_i + \omega_j - \omega_k} - \frac{1}{\omega_i + \omega_j + \omega_k} \right) \right],
\end{aligned}
\tag{2.78}
$$

where δ_{ij} is the Kronecker's delta while the remaining terms have already been introduced. As for the energy, resonances can affect these expressions. Fermi resonances can be treated with the same procedure described in Section **2.3.2**. The same resonant terms identified during the energy evaluation are discarded in the intensity. Another type of resonances comes from the 1-1 resonances [171] which originate when $\omega_i \approx \omega_j$. This term is missing in the VPT2 energies and a proper treatment needs to be included in the DVPT2 procedure in an similar way to what we have described above for the energies.

2.3.4 Reduced dimensionality anharmonic schemes

While VPT2 offers a good balance between cost and accuracy, the evolution of the anharmonic force field requires 2N harmonic frequency calculations. This can be prohibitive if a high level of electronic structure calculation is required, or if the molecular system is large. To reduce the computational cost different strategies can be employed. One way is to combine higher level methods for the harmonic

part, and a cheaper one for the anharmonic contributions. This procedure is usually quite effective in reducing the computational cost of the anharmonic calculations[172,173]. However, there are cases in which the computational cost is still too high and further reduction of the quality of the methods employed in the anharmonic calculations is not possible. A potential solution is to perform a partial anharmonic treatment, by selecting specific frequency ranges that are of interest for a particular system[174]. Within this procedure, called reduced-dimensionality (RD) approach, only anharmonic data related to the modes of interest are computed. This strategy is particularly helpful to isolate and discard modes not accurately treated at VPT2 level such as large amplitude motions (LAMs). The problem here is that all modes are in theory coupled through the third and fourth derivatives of the energy. If we only carried out the numerical differentiation along mode i, and none other, then the following quantities are known, f_{ijk} and f_{iijj} for any value of j and k (see Equation (2.71)). However, f_{jkl} is not known if none of j, k and l are equal to i. In this situation, two terms in χ (Equation (2.71):

$$\Delta_j = \frac{\mathrm{f}_{jjj}\mathrm{f}_{iij}}{\omega_j} \quad \text{and} \quad \Delta_{jk} = \frac{\mathrm{f}_{iik}\mathrm{f}_{jjk}}{\omega_k} \tag{2.79}$$

cannot be evaluated. One option is to calculate the 2D matrix $\boldsymbol{\kappa}$ where the element $\kappa_{ij} = f_{iij}/\omega_j$. Large terms of κ_{ij} suggest a coupling between the modes i and j and that the terms in Equation (2.79) should not be discarded. The resulting procedure can be quite complex and needs to be properly modulated for the studied system. In general, a mode-by-mode approach is required in which starting from one mode of interest the matrix $\boldsymbol{\kappa}$ is calculated. Each mode corresponding to a large value of $\boldsymbol{\kappa}$ will be activated and included in the anharmonic treatment. The matrix $\boldsymbol{\kappa}$ will be recalculated at each step till no significant coupling remains between active and inactive modes. This procedure has been successfully applied for the prediction of vibrations of molecules adsorbed on surfaces and for bio-molecules where a full anharmonic treatment is precluded

by the system size[174].

Reduced-dimensionality schemes can also be used to remove modes poorly treated at the VPT2 level. This is the case of large amplitude motions. Examples or bad behaviour of VPT2 in predicting anharmonic contributions are torsions with a periodic energy profile[175], out-of-plane motions involving nitrogen atoms or internal rotations[176].

2.4 Approaches for condensed matter chemistry

Solid state chemistry developed quite rapidly and was originally developed to answer solid state physics problems, with special focus on the band structure, the Fermi surface shape and their electrical behaviour among the other. Since the morphological features of the solid affect directly the chemistry, a proper procedure accounting for the solid structure is required.

2.4.1 Quantum Chemical simulations

Preliminary attempts at introducing the QM viewpoint into solid state modeling date to the late 1960s through the generalization of the HF equations for crystalline systems with a local basis set. Shortly before 1980, algorithms were developed that demonstrated reliability in computing the band structure, relatively accurate binding energies, equilibrium geometries, and elastic properties. CRYSTAL88 was the only periodic *ab-initio* all-electron program based on the HF Hamiltonian and the use of atom-centered Gaussian functions at that time[177]. Afterward, the evolution in the field was rapid, and now several *ab-initio* codes are available to users[178–180]. In simulating a solid-state system, the following aspects need to be considered.

1. The model. Two different models are available. Most of the chemistry such as a point defect in solids or adsorptions on a crystalline system can be simulated either with a finite cluster by assuming that the cluster is big enough such that border effects are negligible, or with a periodic boundary

condition (PBC) approach.

2. The Hamiltonian. Most of calculations are performed in the framework of DFT which means that a discussion about the appropriate functional needs to be assessed.

3. The basis set. Codes based on plane waves, local functions, and mixed (local functions in atomic spheres, plane waves in the interstices) or numeric basis sets are available.

2.4.1.1 The Cluster approach

The cluster approach consists of extracting from the bulk solid a finite molecular system which is used to represent the extended crystal[181]. The basic idea is to extract sites of interest for the chemical process under study without resorting to PBC and, therefore, treating the solid as a macromolecule. Within the cluster approach molecular codes can be used which can be very handy since they are much more advanced than solid state simulation codes with many more methods implemented opening up to tools that would be otherwise unpractical. In cases where the PES is quite complex such as for concerted mechanisms or multidimensional reactive paths, the molecular approach offers better techniques to localize transition states. Full anharmonic treatments as well are available only for molecular systems. The two problems coming with the cluster approach are: (1) dangling bonds, (2) limited size. While the former requires a proper construction and saturation of the broken bonds on the edges of the cluster, the latter can be overcome by hybrid methods, which also offer the possibility of using refined post-HF methods or composite schemes. As anticipated above, when constructing the cluster, one should also deal with dangling bonds at the edges of the cluster, which should be usually saturated by hydrogen atoms or hydroxy groups. The limited size of the cluster may be overcome by embedding approaches like the Our Own N-layered Integrated Molecular Orbital and Molecular Mechanics (ONIOM) method proposed by Morokuma and co-workers[182]. The main idea is to model the surface with a

very large cluster but, to make the simulation less demanding, different regions are identified and treated at different levels of theory depending on their role in the problem being studied. The smaller and most representative region within the large cluster (for instance where the adsorption occurs) is treated through a higher level method chosen to improve the chemical description of the active region. For the purpose, DFT, post-HF methods or even composite schemes can be used, while for the low-level portion, that is not actively taking part in the chemistry but is relevant for a proper simulation of the chemical environment, semi-empirical methods or molecular mechanics (MM) are usually employed. Embedded techniques allow for partly solving the "edge effects" problem of the cluster because, since the size of the cluster is large, the edges are far from the chemically relevant zone, avoiding unrealistic interactions.

2.4.1.2 The Periodic approach

In the PBC approach the crystalline solid system is reproduced by repeating its unit cell in the 3D space. Indeed, crystals are translation invariant by definition. A perfect crystal is composed of a three-dimensional array of atoms, ions, or molecules, where a specific spatial pattern is replicated uniformly throughout the entire crystal structure. The collection of points repeated at regular intervals of length $\mathbf{a_1}$, $\mathbf{a_2}$ and $\mathbf{a_3}$ constitutes the lattice and the lattice parameters define the periodicity of the crystalline structure. Lattice parameters and angles between the lattice vectors are called cell parameters and define the crystalline system which the specific crystal belongs. There are 14 different types of lattices that define all the conceivable systems in the ordinary space and are known as Bravais lattices. Seven of these are called *primitive* since the unit cell contains one lattice point. A vector $\mathbf{g}$ joining any two lattice points is a lattice vector. Every lattice vector can always be expressed by the lattice parameters and three integer coefficients n_1, n_2, and n_3:

$$\mathbf{g} = n_1\mathbf{a_1} + n_2\mathbf{a_2} + n_3\mathbf{a_3}. \tag{2.80}$$

Filling the unit cell of a lattice with atoms or molecules in a well-defined geometrical arrangement and applying the translation pattern permits the creation of an ideal crystal. In addition to the set of translations, crystals typically display point symmetry. Point and translation symmetries combine to form a space group. In ordinary space, 230 space groups exist and each of them refers to only one particular Bravais lattice. The complete characterization of space groups can be found in the International Tables of Crystallography [183]. Henceforth, we will refer to the ordinary space as the direct space to distinguish it from the reciprocal space. For every direct lattice a geometric construction of the reciprocal lattice is possible. The basis vectors of the reciprocal lattice are orthogonal to the space of the corresponding direct lattice basis vectors with unequal indices. Solid state simulations codes work in the reciprocal space. In the reciprocal space as in the real space each vector can be expressed as linear combination of the basis vectors (b_1, b_2, b_3) with integer coefficients (K) according to Equation (2.80):

$$\mathbf{K} = K_1\mathbf{b_1} + K_2\mathbf{b_2} + K_3\mathbf{b_3}. \tag{2.81}$$

Within the reciprocal space the Brillouin zone is particularly relevant. It can be obtained by connecting one reciprocal lattice point to all its nearest neighbors and letting orthogonal planes pass through their midpoints. The volume within these planes includes all points that are closer to that reciprocal lattice point than to any other lattice point.

2.4.1.3 Block Theorem and PBC

A real crystal is a macroscopic object of finite size composed of a vast but finite number of atoms. Since the ratio of the number of atoms at the surface to the total number of atoms in the crystal is very small, if the surface is neutral a macroscopic crystal mostly exhibits properties and features of the bulk material, and unless attention is deliberately focused onto the crystal boundary, surface effects can be neglected. The potential energy of such a crystal must be a periodic function with the same periodicity as the lattice, so that for a translation by

any direct lattice vector $\mathbf{g}$, the potential energy does not change:

$$V(\mathbf{r}-\mathbf{g}) = V(\mathbf{r}). \tag{2.82}$$

Same applies to the Schrödinger equation:

$$\hat{H}(\mathbf{r})\Psi(\mathbf{r}) = E\Psi(\mathbf{r}), \tag{2.83}$$

which must also be translation invariant so that after a translation by $\mathbf{g}$ the following equation holds:

$$\hat{H}(\mathbf{r}-\mathbf{g})\Psi(\mathbf{r}-\mathbf{g}) = E\Psi(\mathbf{r}-\mathbf{g}). \tag{2.84}$$

Eigenfunctions of the above equation must obey the Bloch theorem stating that:

$$\Phi(\mathbf{r}+\mathbf{g};\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{g}}\Phi(\mathbf{r};\mathbf{k}). \tag{2.85}$$

As Φ verify the Block theorem, it is called the Bloch function and is a function of the position (r) and the wave vector (k). AO and plane waves (PW) are common choices to build Bloch functions. Both choices would be equivalent, in principle, if an infinite basis set is considered, but they are not equivalent in the practical case of a finite basis set. The use of AOs is better linked to the chemical experience of molecular codes and is particularly suitable to the description of molecular or covalent crystals. On the contrary, plane waves are more suitable to the case of metals and, in general, to the description of delocalized electrons. Another advantage of using PW over AO is that the mathematics involved is usually much easier.

2.4.1.4 Modeling Surfaces and Interfaces

In nature, crystals are not infinite but finite macroscopic three-dimensional (3-D) objects enclosed by surfaces. Many chemical relevant processes occur at the interface between a solid system and the environment. Modeling surfaces is then of great interest. A surface can be created by cutting a crystal through

a specific plane. Upon cutting, two semi-infinite crystals are generated in the direction orthogonal to the surface, and periodicity in that same direction is lost. To overcome this issue, further approximations are required. Alternatively, clusters, embedded clusters, or slabs are possible approaches. The slab model consists of a film made by a few atomic layers enclosed by two surface planes, possibly related by symmetry. For sufficiently thick slabs, this strategy can provide an accurate description of the ideal surface. The reliability of the model must be checked by considering convergence of properties like geometry, energy with the number of atomic layers in the slab. In actual calculations, two different schemes can be used:

1. imposing 2-D periodic boundary conditions. The slab model is two-dimensional;

2. forcing a 3-D periodicity. In this case an array of slabs is reproduced along one direction. Each array is separated from the others with a vacuum zone. The size of the vacuum zone is thoroughly selected in order to avoid fictitious interactions between the slabs. When we use a plane wave basis set, which requires a 3-D Fourier representation of many quantities, this is the only model that can be adopted. On the contrary, when a local basis set is adopted, both schemes can be implemented.

The most common surfaces are those with low Miller indices, because their creation requires a smaller amount of energy. However, not all crystalline surfaces are physically stable or of interest. A slab composed of atomic layers parallel to the selected plane, can be classified according to the stoichiometry of the layer and the net dipole of the surface. Surfaces showing a net dipole are usually unstable and can only be stabilized through some mechanism to remove the macroscopic field (i.e., by reconstruction, molecular adsorption, and so on). In covalent solids, the creation of a surface comes with cutting covalent bonds, which means that dangling bonds would be present at the surface. The resulting instability is partly reduced either by creating new bonds, with a reconstruction

of the surface, or chemisorbing atoms from the environment. Within the slab model approach, the surface formation energy (ΔE^n_{surf}) is computed as:

$$\Delta E^n_{\text{surf}} = \frac{(E^n - nE_{\text{bulk}})}{2A}, \tag{2.86}$$

where E^n is the energy of an n-layer slab, E_{bulk} is the energy of the bulk material, and A is the area of the primitive unit cell. The factor 1/2 accounts for the existence of two limiting surfaces. ΔE^n_{surf} is then the energy per unit area required to form the surface from the bulk, and it is intrinsically a positive quantity (if not, the bulk would exfoliate). As more layers are added in the calculation by increasing the slab thickness ($n \to \infty$), ΔE^n_{surf} will converge to the surface formation energy per unit area. In general, when a surface is cut out of a perfect crystal, the atoms near the surface will move from their bulk positions to minimize the surface energy. The atomic displacement generated by the creation of the surface could either keep the symmetry of the slab (relaxation) or alter the periodicity of the slab (reconstruction).

2.4.2 Molecular Dynamics

Accurate results for chemical properties can be obtained using a proper QM method. However, the steep scaling of the computational cost with the system size limits their application window. As widely discussed in this manuscript, a thorough selection of a method requires a trade-off between size and accuracy. When it comes to biological molecules or materials simulation the dimension of the systems biases the selection of a proper method. MD simulations investigate the dynamics of a molecular system by treating atoms as spheres with fixed charge, size and softness and solving Newtonian equation of motions. When the proper method is adopted it allows the simulation of bigger systems of 10^5 —10^6 atoms/particles covering timescales from nanoseconds to microseconds. Depending on the computer power such number can be even higher and specific methods like machine learning MD [184] and coarse-grained MD [185,186] can allow

to reach even further extended scales. A general MD algorithm can be summed up in the following steps:

1. Definition of input conditions, including a good guess for the system geometry and the force field (discussed later in this text).
2. Force calculation via Newton's equations.
3. Update atoms positions and velocities and any other property of interest.
4. Iterate the previous two steps till the time limit is reached.
5. Analysis and post-processing of results.

In *ab-initio* MD, forces acting on atoms are calculated via QM (mostly DFT), the electronic structure of the system is not lost and excited states analysis as well as reactions can be easily studied. As expected QM always comes with a high price and such procedures cannot be accessible for large size systems and long simulation times. Classical MD algorithms describe the chemical systems as a collection of springs and balls and the potential energy of the entire system is modelled with the so-called force field. Force fields collect a series of parameters that allow the description of our system; they need to be selected properly since the results of the simulation highly depends on the proper use of the force field. Since the jungle of available force fields is quite intricate, we decide to move this discussion to a dedicated section later on in this text.

The MD protocol is available in many softwares and our discussion will refer to GROMACS, which has been used in this work [187,188]. As stated in the previous point 1, the entire procedure requires a starting geometry. In the best case scenario it can be obtained from available crystallographic data or alternatively a good guess can be produced and then minimized before proceeding with the further steps of the dynamic protocol. With step 2, the Newton's equation of motion are solved for a system of N interacting particles:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \qquad i = 1 \dots N \tag{2.87}$$

where m_i and r_i are the masses and positions respectively and forces ($\mathbf{F}_i$) are the negative derivative of the potential energy $V(\mathbf{r_i})$:

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i}. \tag{2.88}$$

The movements of the atoms are obtained by numerically integrating the Newton's equations. One of the most common MD integrator in many codes is the *leap-frog* algorithm which uses positions at time t and velocities at time $t - \frac{1}{2}\Delta t$, while positions and velocities are updated by using the forces value at time t ($\mathbf{F}(t)$) determined by positions at time, t:

$$\begin{aligned} \mathbf{v}\left(t + \frac{1}{2}\Delta t\right) &= \mathbf{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}\mathbf{F}(t), \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \Delta t \mathbf{v}\left(t + \frac{1}{2}\Delta t\right). \end{aligned} \tag{2.89}$$

The above equations are modified in case of either temperature or pressure coupling, as discussed in the next section. In the case of an extremely accurate integration with temperature and/or pressure coupling, the velocity Verlet algorithm should be preferred, in which, positions and velocities at time, t, are used to integrate the equations of motion; velocities at the previous half step are not required:

$$\begin{aligned} \mathbf{v}\left(t + \frac{1}{2}\Delta t\right) &= \mathbf{v}(t) + \frac{\Delta t}{2m}\mathbf{F}(t), \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \Delta t \mathbf{v}\left(t + \frac{1}{2}\Delta t\right), \\ \mathbf{v}(t + \Delta t) &= \mathbf{v}\left(t + \frac{1}{2}\Delta t\right) + \frac{\Delta t}{2m}\mathbf{F}(t + \Delta t). \end{aligned} \tag{2.90}$$

or:

$$\begin{aligned} \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \Delta t \mathbf{v} + \tfrac{\Delta t^2}{2m}\mathbf{F}(t), \\ \mathbf{v}(t + \Delta t) &= \mathbf{v}(t) + \tfrac{\Delta t}{2m}[\mathbf{F}(t) + \mathbf{F}(t + \Delta t)]. \end{aligned} \tag{2.91}$$

Given an initial set of positions and velocities, leap-frog and velocity Verlet

algorithms will not give identical results, as the former interprets the velocities as corresponding to $\mathbf{v}(-\frac{1}{2}\Delta t)$, while the latter interprets them as corresponding to the timepoint $t = 0$.

2.4.3 Temperature coupling

The collection of coordinates and velocities will generate the so-called *trajectory*. While direct use of MD gives rise to the NVE (constant number, constant volume, constant energy) ensemble, most quantities that we wish to calculate are actually from a constant temperature ensemble (NVT or NPT). In this case particle positions and velocities need to be re-scaled according to the operational temperature. Many schemes have been proposed and we will discuss some of the most common: the Berendsen themostat[189], the stochastic velocity-rescaling scheme[190] which are of use for this dissertation. The reader should be aware that other thermostats are available and the proper one needs to be selected according to the type of study. In ensembles other than the microcanonical ensemble the total energy is no longer conserved.

2.4.4 The Berendsen thermostat

The Berendsen thermostat operates by weakly coupling the system to a heat bath with given temperature T_0[189]. The system temperature is adjusted towards the desired value by rescaling the particle velocities. A deviation of the system temperature from T_0 is corrected as follows:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau}. \tag{2.92}$$

The deviation of the system temperature from the desired temperature T_0 decays exponentially with a time constant, τ. In order to move the system dynamics towards the desired value, the velocities are multiplied by a factor λ:

$$\lambda = \left[1 + \frac{n_{\mathrm{TC}}\Delta t}{\tau_T}\left\{\frac{T_0}{T\left(t - \frac{1}{2}\Delta t\right)} - 1\right\}\right]^{1/2}. \tag{2.93}$$

The parameter τ_T and the time constant, τ, of the temperature coupling are related according to the this expression:

$$\tau = 2C_V\tau_T/N_{df}k_B, \tag{2.94}$$

where C_V is the total heat capacity of the system, k_B is Boltzmann's constant, and N_{df} is the total number of degrees of freedom. Using big values of the scaling factor λ can cause instabilities in the simulation, hence its value is maintained in the range of $0.8 <= \lambda <= 1.25$. The kinetic energy is adjusted at each step by:

$$\Delta E_k = (\lambda - 1)^2 E_k. \tag{2.95}$$

The conserved energy can be obtained by subtracting all the changes in the kinetic energy along the dynamics simulation to the total energy.

The scaling process of the system temperature is applied continuously throughout the simulation, leading to a gradual relaxation of the system towards the desired temperature. As a result, the Berendsen thermostat can effectively control the temperature of a system, however, it does not properly account for the statistical fluctuations and correlations between the system's kinetic and potential energy. This means that one does not generate a proper canonical ensemble, and the sampling will be incorrect. This error scales with $1/N$, so for very large systems averages will not be affected. It can lead to incorrect equilibration timescales, incorrect behavior of time-dependent properties, and inaccurate sampling of the phase space. Therefore, for rigorous and accurate simulations, alternative thermostats should be considered.

2.4.5 The stochastic velocity rescaling thermostat

The velocity-rescaling thermostat adds an extra stochastic term to a Berendsen thermostat[190]. This extra term ensures a correct kinetic energy distribution. A correct canonical ensemble is simulated while still using a first-order decay of temperature deviations.

2.4.6 Force fields

As anticipated previously, MD based on MM uses a sum of pair-wise potentials based on classical mechanics. The analytical form of such potential is what is usually called Force Field (FF) and in general it is composed of a few terms:

$$
\begin{aligned}
U &= U_{bonded} + U_{non-bonded}, \\
U_{bonded} &= U_{stretching} + U_{bending} + U_{dihedral} + U_{imp.dihedrals}, \\
U_{stretching} &= \sum_{i \neq j} k_{ij}(R_{ij} - R_{ij,eq})^2, \\
U_{bending} &= \sum_{i \neq j \neq k} k_{ijk}(\theta_{ijk} - \theta_{ijk,eq})^2, \\
U_{dihedral} &= \sum_{i \neq j \neq k \neq l,n} [1 + cos(n\phi_{ikjl} - \delta_n)], \\
U_{improperdihedrals} &= \sum_{i \neq j \neq k \neq l} k_{ijkl}(\phi_{ijkl} - \phi_{ijkl,eq})^2, \\
U_{non-bonded} &= \underbrace{\sum_{i \neq j} \frac{q_i q_j}{r_{ij}}}_{\text{Coulomb}} + \underbrace{\sum_{ij} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]}_{\text{vdW}}.
\end{aligned}
\tag{2.96}
$$

In the above equation, the potential energy (U) is the result of bonded and non-bonded interactions (U_{bonded} and $U_{non-bonded}$ respectively). U_{bonded} in Equation (2.96) includes intra-molecular potentials for stretching, bending, torsion and/or improper torsion. $U_{stretching}$, $U_{bending}$ and $U_{imp.dihedrals}$ are modelled with a harmonic potential (k_{ij} and k_{ijk} are the force constants while r and θ the structural parameters for bond lengths and angles respectively with the corresponding equilibrium values labelled as r_{eq} and θ_{eq}). To mimic the periodic oscillation of torsions $U_{dihedral}$ is a cosine Fourier series with ϕ dihedral angle and δ_n the phase. Some force fields use the same periodic expression for improper dihedrals as well. $E_{non-bonded}$ accounts for both intra- and inter-molecular interactions which can either be electrostatic (first term in $E_{non-bonded}$ formula) or Van der Waals interactions (second term in $E_{non-bonded}$ formula). Since no

information about the electronic structure is retained all the chemistry depending on it cannot be properly described via classical MD including reactivity and excited states. Special FFs have been developed to partially recover this missing part. Generally FFs are classified in the following way:

1. Class I FFs: bond stretching and angle bending are described by simple harmonic motion as in $U_{stretching}$ and $U_{bending}$ in Equation (2.96). Some examples are AMBER[191], CHARMM[192], GROMOS[193], OPLS[194]. The expression for the potential energy of these FFs is that reported in Equation (2.96). The CHARMM FF includes an extra term (the Urey-Bradley component) which evaluates angle bending also for 1,3 non-bonded interactions in a similar fashion to $U_{bending}$ in Equation (2.96).

2. Class II FFs: anharmonic cubic and/or quartic terms to the potential energy for bonds and angles are also included. Besides, they contain cross-terms describing the coupling between adjacent bonds, angles and dihedrals. Higher-order terms and cross terms allow for a better description of interactions resulting in a more accurate reproduction of bond and angle vibrations. MMFF94[195] and UFF[196] are two examples.

3. Class III FFs: special effects of organic chemistry are also included. For example polarization, stereoelectronic effects, electronegativity effect, Jahn–Teller effect, *etc.* Examples of class 3 force fields are AMOEBA[197] and DRUDE[198].

4. Reactive FFs: in this category reactive models, which explicitly take into account bond formation and breaking in order to model chemical reactions, are included. ReaxFF[199] procides an example of this class of FF.

A full survey of FFs and their usage is beyond the scope of this thesis, the interested reader can find full detailed information in the referenced material along this thesis. We will proceed with a short list of the most common FFs developed. Most of the attention will be devoted to FFs developed for the

simulation of water and ice and materials since they are those of main interest in this work.

2.4.7 Short description of common FFs for chemical compounds

2.4.7.1 Biomolecules

Biomolecules are not of interest in this thesis but since most of the MD simulations have been developed for the investigation of macromolecules involved in biological process a short mention is mandatory.
AMBER is a class I force field and was originally designed for the simulation of biomolecules. A wide number of parameters for this FF are available for proteins, lipids, carbohydrates, nucleic acids. The General Amber Force Field (GAFF) [200] also provides the required parameters for the simulation of small organics and its applicability has been recently extended to the simulation of polymers.
Also the CHARMM force field in its latest version CHARMM36 [201] is still a Class I force field.

Currently, there are parameters available for the simulation of most of biomolecules and general organic molecules (the General CHARMM Force Field, CGenFF [202]). The development of force fields is an active research field and many more are available with new versions for those above mentioned released.

2.4.7.2 Water

Water is probably the most studied but still mysterious system. Its unique properties with its indisputable relevance in everyday life made essential the development of specific force fields for its simulation. The models developed can be classified according to: (i) the number of points also called sites, (ii) whether the model is rigid or flexible, (iii) whether the model includes polarization effects. The rigid models are considered the simplest ones and rely on non-bonded interactions, while, bonding interactions are implicitly treated by constraints. The electrostatic interaction is described using Coulomb's law, and a Lennard-

Jones potential ($U_{non-bonded}$ in Equation (2.96)) and is used for the dispersion and repulsion forces. The potential in models such as TIP3P [203] (transferable intermolecular potential with 3 points) and TIP4P [204] is represented by:

$$E_{ab} = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{\text{OO}}^{12}} - \frac{B}{r_{\text{OO}}^{6}}, \tag{2.97}$$

where k_C, the electrostatic constant, has a value of 332.1 Å kcal/$\left(\text{mol} \cdot \text{e}^2\right)$, q_i and q_j are the partial charges relative to the charge of the electron; r_{ij} is the distance between two atoms or charged sites, and A and B are the Lennard-Jones parameters. The charged sites may be on the atoms or on virtual sites (which should mimic lone pairs, see Section **2.4.7.2**).

The Figure below (Section **2.4.7.2**) shows water models with 3- to 6-sites. The used geometrical parameters including the OH distance and the HOH angle vary depending on the model.

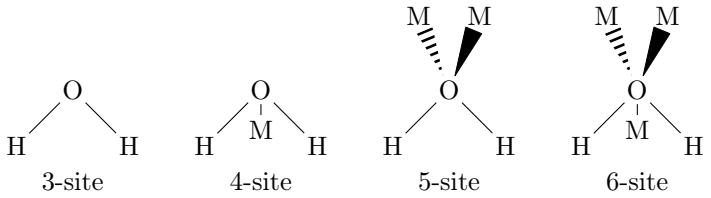

Figure 2.3 Water models for 3 to 6 sites force fields. M is the virtual site designed to account for lone electron pairs.

Rigid force fields show quite bad performances in the prediction of some water properties including IR and Raman spectra. To improve the quality of such analysis some force fields have been reinvented including some degree of flexibily. Most of the effort has been devoted to force fields for the proper simulation of liquid water. Despite this, no universal force field exists and a choice needs to be taken depending on the kind of study we are willing to perform. Definitely less attention has been devoted to crystalline forms of ice and to the optimal reproduction of the ice phase diagram. TIP4P/ICE [205] was one of the first attempts to design a FF made to cope with solid-phase properties.

2.4.7.3 Inorganic materials

ClayFF [206] and INTERFACE (IFF) [207,208] are two force fields specifically derived to simulate clays and minerals respectively using the energy expression derived from Equation (2.96). IFF has been purposely designed for the simulation of silicate interfaces. It comes with a wide database of surface models with a large applicability range for the quantitative prediction of essential properties.

2.4.8 Analysis of trajectory

An essential step in understanding and interpreting MD simulations is the analysis of the trajectories. By extracting meaningful information from the time-evolving coordinates of atoms or particles, trajectory analysis provides insights into the behavior, properties, and interactions within the simulated system. Convergence of system properties is vital for accurate representation of real-world behavior to achieve a stable and representative state of the system by ensuring that these aspects no longer significantly change with time, allowing for accurate and reliable simulation results. It ensures consistent and reliable results, supports parameter tuning, and forms the foundation of meaningful scientific insights. Another common aspect of trajectory analysis is clustering, which groups similar configurations together based on specific properties.

2.4.8.1 Structural Analysis

Structural analysis involves examining the arrangement of atoms or particles in the simulation. Radial distribution functions (RDFs), coordination numbers, bond lengths, angles, and dihedral angles are common analysis. The radial distribution function, in particular, provides information about the probability of finding a particle at a certain distance from a reference point. It reveals important details about the system's structure, such as the presence of ordering, or the formation of local environments (e. g. coordination shells) around specific particles.

2.4.8.2 Dynamical Analysis

Dynamical analysis focuses on the motion and fluctuations of particles over time. Mean square displacement analysis provides insights into diffusive behavior and can help determine diffusion coefficients. Additionally, velocity autocorrelation functions (VACF) and power spectra reveal information about particle velocities and characteristic frequencies within the system.

2.4.8.3 Clustering Analysis

Clustering is a valuable technique for identifying distinct groups or classes of conformations within a trajectory. By partitioning the trajectory into clusters, it becomes possible to study representative structures and their associated dynamics. Clustering algorithms, such as k-means[209], hierarchical clustering, or density-based methods like DBSCAN[210,211], group configurations based on similarity measures like the Root Mean Square Deviation (RMSD). The RMSD calculation measures the structural similarity between different conformations, allowing for the identification of distinct states or conformational transitions. Hydrogen bonds play a vital role in molecular interactions and can be utilized as another criterion for clustering. The number and nature of hydrogen bonds in a configuration provide insight into its stability. By quantifying the hydrogen bonding patterns, one can cluster configurations based on their hydrogen bond networks.

To cluster a trajectory using both RMSD and hydrogen bonds, a combined approach can be employed. First, configurations are grouped according to the number and nature of H-bonds. Within each cluster, further clustering can be performed based on the RMSD using an RMSD-based clustering algorithm. This two-step process allows for a comprehensive analysis of the trajectory, capturing both structural similarity and the variations in hydrogen bonding patterns.

2.4.8.4 Visualization

Visualizing the trajectory and its analysis results is crucial for gaining intuitive insights into the system's behavior. Three-dimensional representations of

molecular structures, animations, and plots of relevant observables aid in understanding complex dynamical processes. Tools such as VMD[212] (Visual Molecular Dynamics), and PyMOL[213] among the others offer interactive visualization capabilities and support the analysis of MD trajectories.

Chapter 3

Reactivity in gas-phase: Ethanimine as test case

3.1 Introduction

Until the latter part of the 20th century, it was believed that the vastness of interstellar space consisted mainly of hydrogen atoms. Apart from molecular hydrogen and a few other simple diatomic species, the harsh conditions of the ISM were thought to be incompatible with polyatomic molecules exhibiting even a small degree of complexity. This idea began to be questioned roughly fifty years ago with the discovery of the first 'complex' molecules (formaldehyde in 1969[15], methanol in 1970[214], and formic acid in 1971[215]). Since these first hints, the pace of molecular detections in space has accelerated. The original paradigm alluded to above has now been completely erased by the discovery of about three hundred molecular species in the ISM and circumstellar shells[24,25]. Among the detected molecules, iCOMs and particularly prebiotic species are of particular relevance because they represent the bridge between simple molecules and biochemical building blocks. Two paradigms are considered for their synthesis: gas-phase chemistry and grain-mediated processes. While the latter will be the subject

of the next chapters, here the focus will be on the investigation of gas-phase chemistry through computational quantum chemistry, taking ethanimine as case study to illustrate this strategy.

Computational chemistry offers an invaluable aid in providing two main pieces of information, namely:
(1) The investigation of reactive PESs from both energetic (thermochemistry) and kinetic points of view. Two possibilities can be actually envisaged: (i) starting from purposely chosen precursors the formation route of the sought product (i.e. a molecule already identified in the ISM) is derived; (ii) starting from small reactive species the possible pathways are elaborated. In both cases, the harsh conditions of the ISM are used as constraints.
(2) Referring to point (1.ii), the accurate prediction of the spectroscopic parameters of those products that can be of interest.
The second piece of information (point (2)) is then used to guide and support laboratory measurements that, in the field of rotational spectroscopy, are mandatory to further proceed toward astronomical searches. Finally, astronomical data can confirm the plausibility of the developed formation pathway. For example, the identification (in astronomical surveys) of the sought product of the reaction investigated can provide an indirect confirmation. Astronomical observations can also be used to support the effectiveness of a gas-phase formation route[216]. Furthermore, as in the case of ethanimine, i.e. in the presence of two different isomers, the computed branching ratios can be compared with the relative astronomical abundance.

As briefly mentioned above, ethanimine has been chosen as a prototypical molecular system for presenting and discussing our strategy. Its astrophysical relevance is related to its prebiotic potential as a possible precursor of amino acids[217], like alanine, by reaction with HCN and H_2O, or with formic acid[218–220]. Both isomers of ethanimine have been detected in Sagittarius B2 North, Sgr-B2 (N), as a result of the GBT-PRIMOS project (this acronym standing for Green Bank Telescope - PRebiotic Interstellar MOlecule Survey)[220]. Furthermore,

this molecule can also be of interest in the study of the nitrogen-based Titan's atmosphere[221], which is considered a good approximation of the primitive Earth[222,223], thus explaining its relevance in the study of prebiotic chemistry in the Universe.

3.2 The thermochemistry-kinetics strategy

As explained above, our strategy can start from a potential iCOM or from a potential gas phase reaction between two molecules already detected in the ISM. The example addressed in this contribution belongs to the first case. In any event, the starting point is the design of a feasible and accessible reactive PES leading to the species of interest in the first case or starting from the chosen reactants in the latter option. This preliminary investigation is performed at a computationally effective level of theory, DFT. Usually, different pathways can be derived and only those that are feasible in the conditions typical of the astronomical environment under consideration are further investigated at a higher level[224]. Therefore, the more promising pathways are re-investigated in order to improve the structural determination of the stationary points as well as to compute the energetics at the state-of-the-art. To conclude, kinetic calculations are carried out, these providing the conclusive information on the effectiveness of the suggested mechanisms and on the branching ratios of the reaction products[225,226]. On the spectroscopic side, as far as rotational spectroscopy is concerned, the target species are better characterized in terms of structural, and spectroscopic properties in order to provide reliable predictions for the laboratory spectroscopy studies[227,228]. Finally, astronomical observations can be performed and might be used to support the outcomes of the theoretical investigation (see e.g. ref.[226,229]).

In the case of ethanimine, several pieces of information and reliable data were already available prior to our investigation. This was indeed the reason why it was chosen as case study to develop and test our strategy. Ethanimine was already detected in the ISM, but the knowledge on its rotational spectrum

was limited. Possible gas-phase mechanisms were also available in the literature, even if not entirely satisfactory from an astronomical point of view. In what follows we will discuss the first part of the work which has been performed by the author of this thesis. Detailed description of the spectroscopy and kinetics results can be found in the provided references[74,230].

3.2.1 Formation pathway

As far as the formation mechanism of ethanimine in the ISM is concerned, among the gas-phase reactions considered by Quan *et al.*[217], the NH + C_2H_5 reaction is certainly the best candidate, because the analogous NH + CH_3 reaction was demonstrated to be the dominant formation route of methanimine in the model proposed by Suzuki *et al.*[231]. Recently, Balucani and coworkers[232] performed an exhaustive search of different possible reaction paths, concluding that only the following three channels are open in the ISM conditions:

$$NH + C_2H_5 \longrightarrow CH_3 + CH_2NH. \quad \text{(R1)}$$

$$NH + C_2H_5 \longrightarrow \text{E-}CH_3CHNH + H. \quad \text{(R2a)}$$

$$NH + C_2H_5 \longrightarrow \text{Z-}CH_3CHNH + H. \quad \text{(R2b)}$$

While the computational level used in this previous work[232] might not be sufficient to obtain quantitative results, the computed energy barriers governing the other considered paths are so high that they can be safely excluded in the present study, with the only exception of channel (indicated as 3 in their paper):

$$NH + C_2H_5 \longrightarrow C_2H_4 + NH_2. \quad \text{(R3)}$$

The reaction of CH_2 with CH_2NH was also proposed[233]. However, the computational level employed to analyze the mechanism was quite low and only the singlet form of CH_2 leads to submerged barriers, thus requiring a spin-orbit coupling mechanism from the more stable triplet form. Furthermore, the first

step of this reaction is the abstraction of one hydrogen atom from CH_2NH by CH_2. Therefore, the effective reaction involves the $CH_3^\bullet$ and $CHNH^\bullet$ radicals:

$$CH_3 + CHNH \longrightarrow \text{E-/Z-}CH_3CHNH \qquad \text{(R4a/4b)}$$

While the $CH_3^\bullet$ radical has been detected in several regions of the ISM, only the $CHNH^+$ cation has been identified so far, the corresponding neutral radical remaining still unobserved. As a consequence we have not investigated in detail this alternative mechanism.

The rate coefficients employed by Quan *et al.*[217] in their model are 8.25 Å$\times 10^{-12}$ cm^3/s for $NH + C_2H_5 \rightarrow$ E-$CH_3CHNH + H$ (R2a) and 2.75 Å$\times 10^{-12}$ cm^3/s for $NH + C_2H_5 \rightarrow$ Z-$CH_3CHNH + H$ (R2b) with no temperature dependence. However, in their model, they did not take into consideration that the channel (R1) consumes a fraction of the reactants, thus reducing the total production rate of ethanimine. The lack of temperature dependence being quite surprising, the computations of ref.[232] indeed pointed out that the rate constants for channels R2a and R2b are very similar to one another and both of them nearly double when increasing the temperature from 10 K to 200 K (i.e. the interval considered in ref.[217]).

In the following section the computational methodology at the basis of the investigation of the formation pathway will be described in detail.

3.3 Computational methodology

Our QC strategy relies on a composite scheme in order to obtain high accuracy at an affordable computational cost. This combines CC techniques with extrapolation to the CBS limit and consideration of core-valence correlation contributions. To account for vibrational effects, also including anharmonicity, we resort to VPT2[165].

The QC calculations described in the following have been performed with

Gaussian 16[168] and CFOUR[234] packages. The former code has been employed for DFT and VPT2 computations, while the latter for CC calculations. Furthermore, the MRCC code[235] interfaced to CFOUR has been used to perform the CC calculations including quadruple excitations.

A preliminary characterization of the reactive PES has been carried out at the B3LYP-D3/aug-cc-pVTZ level[125,236–238] (hereafter shortly denoted as B3), with geometry optimizations combined with harmonic force field calculations in order to characterize the stationary points. More generally, for this introductory analysis of the reactive PES, a double-ζ quality basis set can be safely employed in conjunction with the hybrid B3LYP functional. The results obtained at the B3 level were found in agreement with those reported in literature[232]. Since the double-hybrid B2PLYP functional[141] represents a good compromise between accuracy and computational cost for both structural and spectroscopic properties[239–241] and is able to provide a reliable description of reactive and non-reactive PESs[225,226,239,240,242–245], our strategy relies on it in order to provide the final description of the stationary points from a structural point of view. For this purpose, B2PLYP is usually employed in conjunction with a triple-zeta basis sets incorporating diffuse functions. In the present case, the aug-cc-pVnZ (n=T, Q) basis sets have been employed. In addition, different functionals (DSD-PBEP86[143], revDSD-PBEP86[145]) have been tested. However, these as well as the aug-cc-pVQZ set have not pointed out any substantial modification of either the PES or geometrical parameters with respect to the combination of B2PLYP and aug-cc-pVTZ. Furthermore, to account for dispersion interactions, the DFT-D3 scheme[146] has been used coupled to the Becke-Johnson damping function[147]. Overall, in the following as well as in Figure **3.1**, we only refer to the B2PLYP-D3(BJ)/aug-cc-pVTZ level of theory, hereafter simply denoted as B2.

The interested reader can find a detailed structural analysis in Figure **A.1**, Table **A.1** and Table **A.2** of the appendix to this chapter (Appendix **A**). In passing, it should be noted that the NH species has been considered only in the

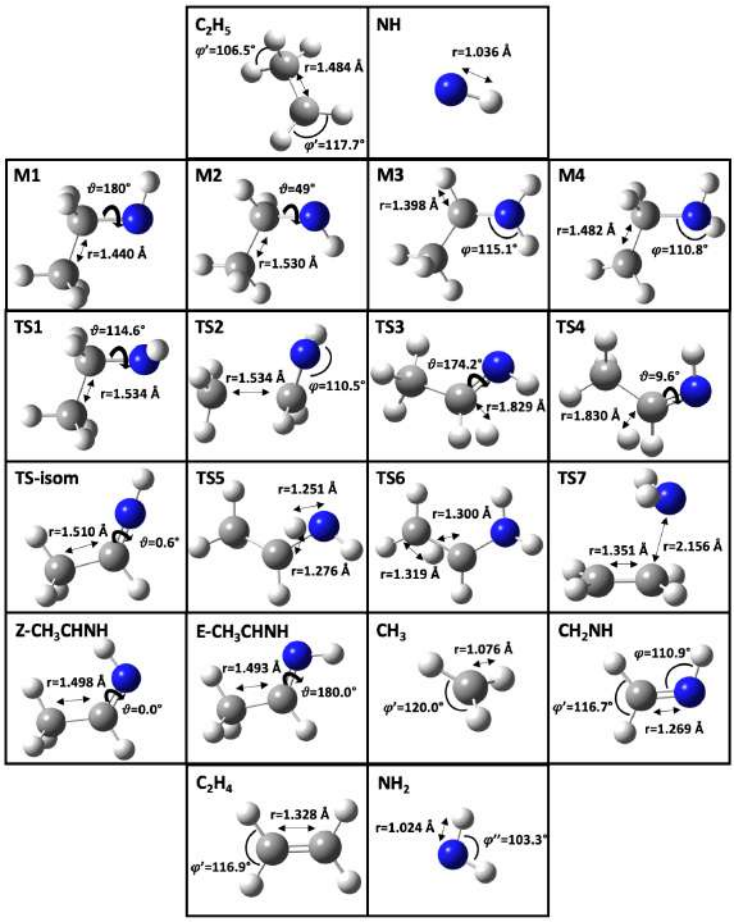

Figure 3.1 Structures and selected geometrical parameters on the $C_2H_5 + NH$ reactive PES optimized at the B2 level. **M** label for minima and **TS** label for transition states. Bond lengths are in Å, whereas HNC (φ), HCH (φ') or HNH (φ") valence angles and CCNH (ϑ) dihedral angles are in degrees.

most stable triplet state $(^3\Sigma^-)$ [246].

The subsequent step of our strategy requires the accurate estimates of reaction energies and barriers. For this purpose, we rely on composite schemes based on CC theory and exploited on top of B2 optimized geometries. In particular, two different types of approach have been employed: (i) schemes entirely based on CC calculations, denotes as "CBS+CV", "CBS+CV+fT+pQ", and "HEAT-like" (see next section), and (ii) the so-called "cheap" model, which involves an important reduction of the computational cost, while keeping high accuracy as discussed in the previous chapter.

Finally, the inclusion of zero-point vibrational energies is required. These can be computed both at harmonic and anharmonic levels. In this respect, as briefly mentioned above, the nature of the stationary points has been checked by the evaluation of analytical Hessians (harmonic force fields), which – in turn – provide harmonic ZPVEs. Anharmonic force fields have also been calculated at the B2 level, thus allowing the evaluation of anharmonic ZPVEs by means of (GVPT2)[165–167,227,247].

3.3.1 Composite schemes

The composite schemes described in the following rely on or start from the CCSD(T) method[248]. The coupled cluster $\mathcal{T}_1$ diagnostic[249] (Section **2.1.2**) has been found to be smaller than 0.02 for closed-shell molecules and smaller than 0.04 for open-shell species (see Table **A.3**), thus confirming that the non-dynamical correlation is negligible for all the systems investigated.

3.3.1.1 The CC-based approaches

The composite scheme denoted as CBS+CV accounts for the extrapolation to the CBS limit and CV corrections at the CCSD(T) level. Within this approach, electronic energies are obtained as described in Equation (2.35).

We evaluated the required terms as follows: (1) the HF-SCF electronic energy extrapolated to the CBS limit using the exponential form proposed by Feller[250] (E_{CBS}^{HF-SCF}) combined with the cc-pVnZ basis sets[238] (n=T,Q,5); (2) the frozen-core (fc) CCSD(T) correlation energy extrapolated to the CBS limit (ΔE_{CBS}^{corr}) employing the two-point n^{-3} formula proposed by Helgaker *et al.*[251] in conjunction with the cc-pVnZ sets, with either the n=T, Q or n=Q, 5 combination of bases; (3) the CV correlation correction (ΔE_{CV}) is calculated as energy difference between all and fc CCSD(T), both with the same basis set. The cc-pCVTZ and cc-pCVQZ[85] bases have been employed depending on whether the extrapolation to the CBS limit has been done using the n=T, Q or n=Q, 5 combinations of basis sets, respectively.

To further improve the energetics, a HEAT-like approach can be (and indeed

has been) exploited, the reference of this scheme being the HEAT protocol[94] (Equation (2.28)). The latter is known to provide sub-kJ mol^{-1} accuracy.

Similarly to the CV contribution evaluated in the CBS+CV appoach, corrections due to the full treatment of triple excitations (fT, see Equation (2.31)), and the perturbative description of quadruples, (pQ, where CCSDT(Q) is employed instead of CCSDTQ in Equation (2.32)), are computed employing the cc-pVTZ and cc-pVDZ basis sets, respectively. Inclusion of fT and pQ contributions leads to the definition of the CBS+CV+fT+pQ scheme. The diagonal Born-Oppenheimer correction[252–255], ΔE_{DBOC}, and the scalar relativistic contribution to the energy, ΔE_{REL},[256,257] are also included in the HEAT-like approach. The former corrections are usually computed at the HF-SCF/aug-cc-pVnZ level[237], with n=D,T, and the relativistic corrections are obtained at the all-CCSD(T)/aug-cc-pCVnZ level (n=D,T) and include the (one-electron) Darwin and mass-velocity terms. In the present study, the last two corrections have been obtained using double-zeta basis sets, even if their convergence with respect to contributions calculated with triple-zeta basis sets has been checked for a few stationary points.

To exploit the high accuracy of the CBS+CV energy approach for molecular structure determinations, the energy-gradient composite scheme introduced in Section **2.1.4.2**[110,111] can be employed. The resulting energy gradient to be employed in geometry optimizations is given by Equation (2.40).

3.3.1.2 The ChS approach

In the present work, electronic energies have also been computed by resorting to a computationally less demanding strategy, the so-called "cheap" scheme (ChS, Equation (2.37)). Our strategy relies on this model when larger systems are involved. In the present context, it has been employed to further demonstrate its high accuracy and reliability, which was already pointed out in previous works[224,228,244].

This approach has been originally developed for the determination of accurate structural parameters[258,259] and then extended to energy evaluations[260]. In this work, the fc-CCSD(T)/cc-pVTZ energy is taken as the starting point and

augmented by contributions accounting for the extrapolation to the CBS limit and CV correlation both calculated at MP2 level.

$\Delta E_{CBS}^{MP2_{corr}}$ is evaluated with in conjunction with the cc-pVTZ and cc-pVQZ basis sets (Equation (2.30)). ΔE_{CV}^{MP2} is evaluated as in Equation (2.36), as 'all-fc' energy difference at the MP2/cc-pCVTZ level.

3.3.2 Kinetic models

Phenomenological rate coefficients have been obtained within the AITSTME approach using the RRKM 1D master equation system solver (MESS) code [155]. For channels possessing a distinct saddle point, rate coefficients have been determined by conventional transition state theory (TST) within the rigid-rotor harmonic-oscillator (RRHO) approximation. Instead, rate constants for barrierless elementary reactions have been computed employing PST [261,262], again within the RRHO assumption.

For what concerns PST, it provides a useful, and easily implemented, reference theory for barrierless reactions, and it largely used in computational kinetics applied to astrochemistry (see, e.g., refs [226,229,263,264]). Its assumption is that the interaction between the two reacting fragments is isotropic and does not affect the internal fragment motions. This approximation is especially valid if the dynamical bottleneck lies at large separations where the interacting fragments have free rotations and unperturbed vibrations. This is generally true for low temperature phenomena, as is the case for the ISM. The isotropic attractive potential is assumed to be described by the functional form $-\frac{C}{R^6}$, with R interatomic distance and the coefficient C obtained by fitting the electronic energies obtained at various long-range distances of fragments. To be more precise, we performed a relaxed scan of the HN-CH_2CH_3 distance and then fitted the corresponding minimum energy path to a $f(x) = f_0 - \frac{C}{R^6}$ function, thus obtaining a C value of 118.86 a_0^6 E_h. Tunnelling has been take into account by using the Eckart model [152].

3.4 Results and Discussion

The thermochemical aspects of the gas-phase reaction between the ethyl and imidogen radicals are first discussed, and then the attention will be focused on the results of RRKM kinetic simulations. Subsequently, the spectroscopic characterization of ethanimine is addressed.

3.4.1 The NH + C_2H_5 reaction: thermochemistry

As described in the Introduction, we have characterized the most favorable reaction channels following the addition of NH to C_2H_5, namely R1, R2a, R2b and R3. The resulting pathways are presented in Figure **3.2**.

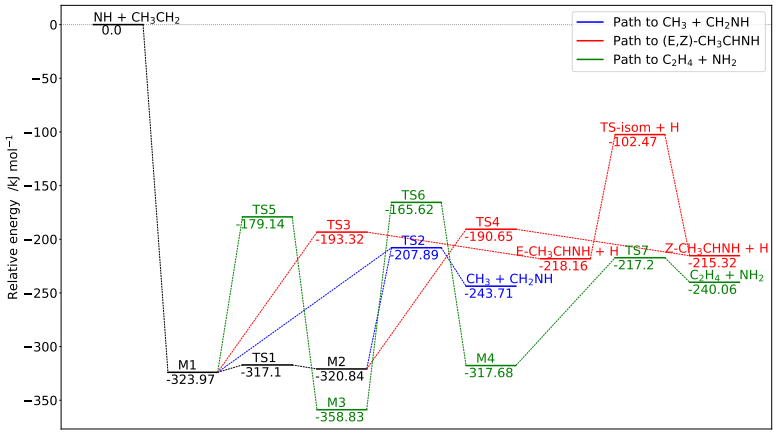

Figure 3.2 Reaction mechanisms for the NH + C_2H_5 reaction: CBS+CV energies augmented by B2 anharmonic ZPVEs. Structures are given in Figure **3.1**.

The initial association of reactants leading to the M1 (see Figure **3.1**) intermediate is highly exothermic, with M1 easily interconverting to M2 through a small energy barrier (TS1). Since the channel leading to $CH_3 + CH_2NH$ requires the smallest number of steps and the lowest barrier (TS2; breaking of the C−C bond), thermodynamic considerations should favor this process. The evolution of M1 to E-CH_2CHNH + H and of M2 to Z-CH_2CHNH + H requires overcoming the transition states TS3 and TS4, respectively, which imply the breaking of one

C−H bond and are less than 10 kJ mol^{-1} higher in energy than TS2. This means that multiple competitive pathways can come into play. Since the E $\longleftrightarrow$ Z isomerization involves a high energy transition state, i.e. TS-isom, once ethanimine is formed, the direct interconversion of the two isomers in the gas phase is very unlikely. As expected, the predicted stability of E-CH_3CHNH by 2.8 kJ mol^{-1} with respect to the Z isomer is in agreement with the value of 2.8 kJ mol^{-1} by Melli *et al.*[230], the level of theory being nearly the same. Analogously, the E $\rightleftharpoons$ Z isomerization barrier of 115.7 kJ mol^{-1} (calculated with respect to E-CH_2CHNH) matches that of ref.[230], while the CCSD(T)/aug-cc-pVTZ value reported in ref.[232] results to be overestimated. Besides leading to methanimine or ethanimine, M1 could potentially undergo multiple hydrogen transfers, according to the path reported in green in Figure **3.2**, to end up with the release of NH_2 and C_2H_4. However, the multiple rearrangements and the relatively high barrier associated to the M3 $\rightleftharpoons$ M4 conversion make this process the less feasible in the harsh conditions of the ISM.

The electronic energies of all the species involved in the paths shown in Figure **3.2**, referred to the isolated reactants, are collected in Table **3.1**, where harmonic and anharmonic ZPVE corrections are also given. In addition to CBS+CV, CBS+CV+fT+pQ, HEAT-like and ChS energies, Table **3.1** also reports B2 results and, for the sake of comparison, the results from ref.[232].

From the inspection of Table **3.1**, it is evident that the CBS+CV energies obtained by extrapolating the correlation energy using either cc-pVTZ/cc-pVQZ or cc-pVQZ/cc-pV5Z are in close agreement, the largest difference being around 0.4 kJ mol^{-1} for TS7. Incorporation of the fT and pQ contributions provides negligible effects, the only exceptions being TS2, TS3 and TS4, for which deviations of about 2-3 kJ mol^{-1} are noted. As expected from our previous experience[224,228,244], the results obtained by using the ChS model are accurate. In fact, they are within 2.5 kJ mol^{-1} from the corresponding CBS+CV values, and within 2 kJ mol^{-1} with respect to CBS+CV+fT+pQ results. Finally, by incorporating the DBOC and relativistic corrections, the HEAT-like values have

Table 3.1 Relative electronic energies together with ZPVE corrections of the stationary points involved in the $NH + C_2H_5$ PES. Values in $kJ\,mol^{-1}$.

	B2	CBS+CV[a]	CBS+CV+fT+pQ[b]	HEAT-like[b]	ChS	harm-ZPE[c]	anharm-ZPE[c]	$\Delta H_0^{0\,d}$	$\Delta H_0^{0\,e}$
$NH + C_2H_5$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
M1	-345.16	-352.10	-351.92	-352.09	-352.82	28.40	28.13	-323.97	-311
		(-352.01)	(-351.83)	(-352.00)					(-323)
M2	-342.24	-349.43	-349.27	-349.44	-350.28	28.94	28.59	-320.84	-308
		(-349.33)	(-349.18)	(-349.35)					(-320)
M3	-382.90	-388.91	-	-	-390.12	31.07	30.08	-358.83	-340
		(-388.89)							(-357)
M4	-336.75	-345.98	-	-	-347.39	28.74	28.30	-317.68	-301
		(-345.71)							-
TS1	-338.10	-345.27	-345.11	-345.28	-346.14	28.51	28.17	-317.10	-304
		(-345.20)	(-345.04)	(-345.21)					(-316)
TS2	-223.85	-226.32	-229.11	-229.22	-227.56	18.47	18.43	-207.89	-197
		(-226.58)	(-229.37)	(-229.48)					(-211)
TS3	-195.51	-202.17	-203.93	-204.09	-202.88	9.31	8.85	-193.32	-180
		(-202.41)	(-204.18)	(-204.34)					(-197)
TS4	-192.92	-199.72	-201.47	-201.62	-200.42	9.48	9.07	-190.65	-177
		(-199.95)	(-201.70)	(-201.85)					(-194)
TS-isom	-94.55	-98.47	-98.17	-98.40	-100.55	-4.13	-4.00	-102.47	-84
		(-98.81)	(-98.51)	(-98.74)					(-101)
TS5	-190.43	-197.33	-	-	-199.01	18.79	18.19	-179.14	-165
		(-197.27)							-
TS6	-176.42	-182.49	-	-	-184.40	17.70	16.87	-165.62	-146
		(-182.51)							-
TS7	-236.01	-237.83	-	-	-238.77	20.67	20.63	-217.20	-207
		(-238.27)							-
$CH_2NH + CH_3$	-245.85	-251.18	-251.36	-251.34	-253.69	7.24	7.47	-243.71	-230
		(-251.38)	(-251.57)	(-251.55)					(-243)
E-$CH_3CHNH + H$	-213.82	-222.30	-221.99	-221.92	-224.10	4.00	4.14	-218.16	-202
		(-222.65)	(-222.34)	(-222.27)					(-217)
Z-$CH_3CHNH + H$	-210.94	-219.52	-219.22	-219.13	-221.31	4.06	4.20	-215.32	-199
		(-219.85)	(-219.55)	(-219.46)					(-214)
$C_2H_4 + NH_2$	-242.67	-248.70	-	-	-250.94	8.26	8.64	-240.06	-228
		(-248.78)							(-241)

[a] CBS+CV scheme: the n=T, Q set of bases for the extrapolation to the CBS limit and cc-pCVTZ for the the CV contribution. Within parentheses, the results for the n=Q, 5 set (CBS) and cc-pCVQZ (CV).

[b] CBS+CV: n=T, Q set (CBS) and cc-pCVTZ (CV). Within parentheses, n=Q, 5 (CBS) and cc-pCVQZ (CV).

[c] Relative ZPVE corrections at the B2 level.

[d] CBS+CV electronic energies augmented by anharmonic ZPVE corrections.

[e] Data from ref.[232]: values at the CCSD(T)/aug-cc-pVTZ level, within parentheses W1 values. ZPVE corrected values.

been obtained. These negligibly differ from the corresponding CBS+CV+fT+pQ results, the sum of the two contributions being on the order of 0.1 —0.2 $kJ\,mol^{-1}$.

In Table **3.1**, the ΔH_0^0 values correspond to the CBS+CV energies augmented by anharmonic B2 ZPVEs. It is noted that they are in agreement with those calculated in ref.[232] at the W1 level, even if available for a lesser number of species. On the other hand, the CCSD(T)/aug-cc-pVTZ level, employed in ref.[232] for all stationary points, shows deviations as large as 19 $kJ\,mol^{-1}$ with respect to our CBS+CV results, with an average overestimation of about 14.5 $kJ\,mol^{-1}$. It is particular interesting that the B2 model chemistry performs better than CCSD(T)/aug-cc-pVTZ computations, indeed the largest deviation

of the former from CBS+CV data is $9\,\mathrm{kJ\,mol^{-1}}$, systematically overestimating the latter results.

3.4.2 The $NH + C_2H_5$ reaction: rate coefficients

As described in Section **3.3.2**, for the reaction paths shown in Figure **3.2**, channel specific rate constants have been computed using AITSTME. For these calculations, the CBS+CV energies, corrected by anharmonic ZPVEs, have been employed for all intermediate and transition states of the PES. The rate coefficients have been evaluated in the 10 —300 K temperature range and at pressure of 1×10^{-12} atm, the results being collected in Table **3.2**. The corresponding temperature dependence plots are shown in Figure **3.3**. The kinetic analysis shows that, in agreement with previous computations, only the channels leading to methanimine (R1), E-ethanimine (R2a), and Z-ethanimine (R2b) are significant. The contribution of the channel leading to $C_2H_4 + NH_2$ is indeed negligible because of the limitations explained above. Focusing on the temperature dependence of the rate constant, an interesting feature can be noted at very low temperature (i.e. 10 K). In fact, at such low temperatures, the trend of k for the channel leading to $C_2H_4 + NH_2$ shows a larger variation with temperature than the other paths. This can be explained as a quantum tunneling effect because the TSs of this path are higher in energy and the rearrangements involve H transfer. As mentioned in Section **3.3**, the two reactants, CH_3CH_2 and NH, are in a spin-doublet and a spin-triplet state, respectively. As a consequence, this results in a total of six spin-microstates, which can be classified in reactive (overall: spin-doublet state) and non-reactive (overall: spin-quartet state). Since the number of the reactive spin-microstates is 2 and that of the non-reactive ones is 4, a statistical factor of 1/3 must be applied to the total reaction rate constant in order to account for this ratio. This has been indeed taken into account in our calculations.

The branching ratios at selected temperatures are shown in Table **3.3**. Is clear that, at all the temperatures considered, the channel leading to CH_3 +

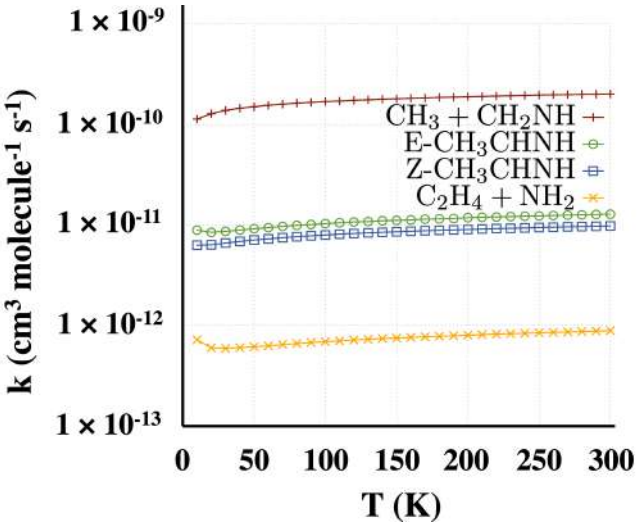

Figure 3.3 Rate coefficients (k) as a function of temperature for the four products.

CH_2NH is by far the dominant one, indeed accounting for more than 87% of the total yield. The ratio between the E- and Z- isomers of ethanimine spans from 1.3 to 1.4 in the entire analyzed range. Even if this value is greater than that obtained by Balucani *et al.*[232], it is still underestimated with respect to the factor of ca. 3 derived from the observation of Loomis *et al.*[220]. While our theoretical estimate is difficult to improve (because already at the state-of-the-art), this discrepancy deserves to be addressed in some details. First of all, the astronomical value is affected by a large uncertainty mostly due to the fact that, for the analysis of E-ethanimine features, the authors had to fix its excitation temperature to that derived for Z-CH_3CHNH, this resulting in an approximate value of the column density of the E species (see ref.[220] for a thorough account). Second, a non-negligible contribution of dust-grain chemistry to the production of ethanimine cannot be excluded.

Finally, kinetic calculations have also been performed using the CBS+CV+fT+pQ energies, corrected by anharmonic B2 ZPVEs. In these calculations, the reaction channel R3 has been excluded, since it has been demonstrated that it gives non relevant contribution to the total reactive flux. The results are almost identical to those obtained exploiting CBS+CV energies. For

Table 3.2 Product-formation rate constants (in cm^3 $molecule^{-1}$ s^{-1}) at 1×10^{-12} atm as a function of the temperature.

T(K)	$CH_3 + CH_2NH$	E-CH_3CHNH	Z-CH_3CHNH	$C_2H_4 + NH_2$
10	1.13×10^{-10}	8.87×10^{-12}	6.29×10^{-12}	7.19×10^{-13}
20	1.28×10^{-10}	8.45×10^{-12}	6.35×10^{-12}	5.98×10^{-13}
30	1.38×10^{-10}	8.64×10^{-12}	6.61×10^{-12}	5.90×10^{-13}
40	1.45×10^{-10}	8.91×10^{-12}	6.85×10^{-12}	6.00×10^{-13}
50	1.50×10^{-10}	9.18×10^{-12}	7.08×10^{-12}	6.14×10^{-13}
60	1.55×10^{-10}	9.41×10^{-12}	7.27×10^{-12}	6.27×10^{-13}
70	1.59×10^{-10}	9.67×10^{-12}	7.47×10^{-12}	6.45×10^{-13}
80	1.63×10^{-10}	9.90×10^{-12}	7.64×10^{-12}	6.61×10^{-13}
90	1.66×10^{-10}	1.01×10^{-11}	7.80×10^{-12}	6.76×10^{-13}
100	1.69×10^{-10}	1.03×10^{-11}	7.95×10^{-12}	6.90×10^{-13}
110	1.71×10^{-10}	1.05×10^{-11}	8.08×10^{-12}	7.03×10^{-13}
120	1.74×10^{-10}	1.07×10^{-11}	8.21×10^{-12}	7.16×10^{-13}
130	1.76×10^{-10}	1.08×10^{-11}	8.33×10^{-12}	7.28×10^{-13}
140	1.78×10^{-10}	1.10×10^{-11}	8.44×10^{-12}	7.39×10^{-13}
150	1.80×10^{-10}	1.11×10^{-11}	8.55×10^{-12}	7.50×10^{-13}
160	1.82×10^{-10}	1.12×10^{-11}	8.65×10^{-12}	7.60×10^{-13}
170	1.84×10^{-10}	1.14×10^{-11}	8.75×10^{-12}	7.70×10^{-13}
180	1.86×10^{-10}	1.15×10^{-11}	8.84×10^{-12}	7.80×10^{-13}
190	1.87×10^{-10}	1.16×10^{-11}	8.93×10^{-12}	7.90×10^{-13}
200	1.89×10^{-10}	1.18×10^{-11}	9.02×10^{-12}	7.99×10^{-13}
210	1.90×10^{-10}	1.19×10^{-11}	9.11×10^{-12}	8.09×10^{-13}
220	1.92×10^{-10}	1.20×10^{-11}	9.19×10^{-12}	8.18×10^{-13}
230	1.93×10^{-10}	1.21×10^{-11}	9.27×10^{-12}	8.26×10^{-13}
240	1.94×10^{-10}	1.22×10^{-11}	9.35×10^{-12}	8.35×10^{-13}
250	1.96×10^{-10}	1.23×10^{-11}	9.42×10^{-12}	8.44×10^{-13}
260	1.97×10^{-10}	1.24×10^{-11}	9.49×10^{-12}	8.52×10^{-13}
270	1.98×10^{-10}	1.25×10^{-11}	9.57×10^{-12}	8.61×10^{-13}
280	1.99×10^{-10}	1.26×10^{-11}	9.64×10^{-12}	8.69×10^{-13}
290	2.00×10^{-10}	1.27×10^{-11}	9.71×10^{-12}	8.78×10^{-13}
300	2.01×10^{-10}	1.28×10^{-11}	9.77×10^{-12}	8.86×10^{-13}

Table 3.3 Product branching ratios at various temperatures.

Branching ratios	$CH_3 + CH_2NH$	E-CH_3CHNH + H	Z-CH_3CHNH + H	$C_2H_4 + NH_2$
10 K	87.7 %	6.9 %	4.9 %	0.6 %
100 K	89.9 %	5.5 %	4.2 %	0.4 %
300 K	89.6 %	5.7 %	4.3 %	0.4 %

the sake of completeness, they are only reported in the Appendix **B**.

3.5 Conclusions

As a result difficulties in experimentally mimicking the extreme conditions that characterize the interstellar medium, accurate state-of-the-art computational approaches play a fundamental role. We have demonstrated how state-of-the-art QC methodologies are able to accurately describe reactive PESs, to be complemented by kinetic calculations. Even if our state-of-the-art investigation on the gas-phase formation of ethanimine is not able to perfectly reproduce the [E-CH_3CHNH]/[Z-CH_3CHNH] ratio of ca. 3 issuing from astronomical observations, the result obtained (1.4) is of the correct order of magnitude and it can be considered as a satisfactory value in view of the complexity of the chemistry in space, involving complex networks of chemical reactions.

Chapter 4

Glinding on ice

4.1 Introduction

In 2018, McGuire published a census of Interstellar, Circumstellar, Extragalactic, Protoplanetary Disks, and Exoplanetary Molecules[24] including more than 200 molecules (containing from 2 to 70 atoms) and this number is steadily increasing thanks to the modern technologies of new observatory telescopes[8]. The identification of many iCOMs defeated the old and general idea that the ISM was an empty vial where chemical reactivity could not operate. Questions about the formation of iCOMs in such extreme conditions and the evolution of molecular complexity fueled the curiosity of astrochemists all over the world[7]. While gas phase reactions seemed the obvious choice to explore the formation pathways of molecular systems in such rarefied environments, the ubiquitous presence of dust and grains and the mismatch between some observations and the molecular abundances predicted by gas phase models have boosted the role of solid state chemistry[26,49]. Since the discovery of the catalytic role of grains for H_2 formation[47,48], astrochemists and physicists have struggled looking for gas-grain models that could provide a comprehensive picture of chemical processes in the ISM. At the low temperatures of molecular clouds (MC), molecules in the gas phase accrete icy mantles freezing out onto grain surfaces[50,54] and leading

to porous and amorphous icy surfaces[265–267], which can host local reactants triggering a molecular reactivity not feasible in the gas phase. The composition and morphological features make the simulation of these icy structures a great challenge in this field[50,268].

The difficulty of performing experimental studies for systems capable of mimicking the harsh conditions of the ISM, calls for computational simulations of periodic surfaces and/or suitable model clusters able to take into the proper account the main structural features responsible for the chemistry at the interface[269,270]. This translates into the necessity of simulating extended systems, thus making the computational burden prohibitive for the accurate state-of-the-art methods developed for isolated molecules[271]. Since water is the main component of polar icy mantles[55,56], a lot of efforts have been devoted to the investigation of the adsorption and formation of iCOMs on water clusters used to mimic interstellar ices. The structures of H_2O clusters containing up to 22 atoms have been worked out from molecular dynamics simulations and made available in online databases[272]. Some years ago, Rimola *et al.* studied iCOMs formation pathways on clusters including up to 33 water molecules obtained by combining two $(H_2O)_{18}$ clusters taken from the (010) surface of ice-XI[273] and removing three molecules to facilitate the construction of the final cluster[274]. Furthermore, attempts to include the structural modifications induced by UV and cosmic rays photo-processing have been made by means of small radical and ionized water clusters[275]. More recently, molecular dynamics has been used to model amorphous water ices[276] and to simulate mixed CO/H_2O ices[277]. Adsorption energies on clusters of larger size have been evaluated by a two-layer our own N-layered integrated molecular orbital molecular mechanics (ONIOM) model, with the higher-level layer treated by means of DFT, and the lower-level one described through molecular mechanics (MM)[278,279] or semiempirical quantum chemical methods[280].

While coupled cluster theory including full treatment of single and double excitations together with perturbative estimation of triple excitations (CCSD(T)),

possibly in conjunction with composite schemes to estimate the CBS limit, is considered the gold-standard for accurate predictions[227], the size of the systems to be dealt with in the case of ice-mediated chemistry makes DFT the only viable route in terms of accuracy to computational cost trade-off. As is well known, the reliability of DFT strongly depends on the specific system and properties at hand and on the choice of the density functional among an ever increasing number of possible formulations. In this respect, benchmark is a fundamental step for ranking the reliability of DFT model chemistries, also in connection with the computational cost, and hence it represents a very active field of research.

Concerning the specific topic of adsorption and reactivity of iCOMs on interstellar ice analogues, to the best of our knowledge, systematic benchmark studies are still lacking. In this connection, Enrique-Romero *et al.*[281] performed a calibration analysis of radical-water interactions and activation energy for NH_2 + HCO and CH_3 + HCO reactions in the presence of one and two water molecules. They tested the accuracy of B3LYP and BHLYP functionals (both with and without dispersion corrections) in conjunction with the 6-311++G(2df,2pd) basis set taking CASPT2/cc-pVTZ and CCSD(T)/aug-cc-pVTZ levels of theory as reference. That analysis was focused on interaction and activation energies, while recent works have highlighted that reliable geometries are fundamental prerequisites for accurate thermochemistry and kinetics[271]. In this respect, the B3LYP functional can be unable to predict correct structures for van der Waals complexes[243] and transition states[282]. Furthermore, the use of CCSD(T)/triple-ζ energies cannot be recommended as a reference in benchmark studies because basis set truncation and lack of core-valence correlation limit the accuracy, thus introducing a bias in the reference values. This issue can be overcome by resorting to composite methods that aim at minimizing the errors relying on well-tested additive approximations[271,283].

In this work, we assess the performances of several DFT model chemistries in evaluating the structural and energetic aspects of ice-mediated interstellar reactions employing the HCN⇌HNC isomerization catalysed by water molecules

as a paradigmatic process. On the one side, this can be considered a model for more complex reactions mediated by ice surfaces and, on the other side, the chosen system is small enough to allow the exploitation of state-of-the-art composite methods to generate accurate reference values for both geometries and reaction energies. The HCN⇌HNC isomerization has been widely studied since the observed HNC/HCN ratio in the ISM can not be predicted on the basis of the proposed gas-phase mechanisms. Moreover, both HCN and HNC can be involved in the formation of amino acid precursors in the Strecker synthesis of glycine[284,285]. Gardebien *et al.* investigated the process for the isolated molecule and with explicit inclusion of two to four water molecules[286] finding that the most favourable mechanism consists of a one-step path involving a proton relay mediated by the water cluster. Koch *et al.* employed a more realistic model including seven additional water molecules to simulate the local environment of the icy surface and employing the polarizable continuum model (PCM) to account for bulk effects[287]. According to the available data, the water cluster acts as a catalyst lowering the energy barrier with respect to the gas-phase, an effect that progressively smooths increasing the number of H_2O molecules. Intermolecular proton transfer drives both the interaction of HCN and HNC with the ice surface and the isomerization process. This represents the most common mechanism through which molecules adsorb and react on ISM polar ices.

On these grounds, we decided to perform a detailed study of the HCN⇌HNC isomerization by state-of-the-art QC methods and realistic cluster models. The work is organized as follows: the computational methods are described in Section **4.2**, while the outcomes of the benchmark are detailed in Section **4.3** concerning both geometries and energies, thus leading to the identification of the best performing DFT model chemistries in terms of the trade-off between accuracy and computational cost. Despite the fact that the benchmark is carried out on a simplified model, the outcomes are expected to be of general validity, especially with respect to the relative performances of the tested methods which

can then be transferred to larger H_2O clusters. With this in mind, at the end of Section **4.3**, the best performing methods are employed to simulate the HCN⇌HNC isomerization catalyzed by a cluster of twenty water molecules, then further embedded in a 172 water slab described through MM. Finally, reaction rates are computed in the framework of the TST including tunneling.

4.2 Computational methodology

For the benchmark study, we selected 10 density functionals belonging to different families: two hybrids (B3LYP, BHLYP)[125,236,288], a long-range corrected DF (ωB97X-D)[139], three meta-hybrids (PW6B95[133], BMK[136] and M06-2X[134]), one meta-NGA (MN15[135]), the B2PLYP[141] and the two spin-component-scaled (DSD-PBEP86 and revDSD-PBEP86)[144,145] double-hybrids. To test the accuracy to computational cost trade-off, for each functional six basis sets have been considered. In particular, we selected the Dunning's aug-cc-pVnZ basis sets (n = D, T)[237,238] as well as the corresponding jun- and jul- modifications from the Truhlar's calendar family.[90] All the DFT calculations include empirical dispersion corrections according to the DFT-D3 scheme proposed by Grimme[146] with the Becke-Johnson damping function[147,289], which are fundamental for the correct prediction of van der Waals complexes[290–292], transition states[293] and surface processes[294,295]. Accurate reference geometries and energies for the benchmark were generated by using the ChS[296,297] and its recent jun-ChS revision[271,283], with the latter appearing the best option because of the increased reliability for non-covalent interactions and the better description of the water dimer structure. Indeed, for $(H_2O)_2$, ChS and jun-ChS geometries were first compared to highly accurate CCSD(T)-F12b/CBS+fT+fQ+CV+REL+DBOC values[298]. The results, reported in Table **B.1** of the Appendix **B**, show that bond lengths and valence angles are reproduced very accurately, with maximum errors of −0.003 Å and −0.2°, while there is a deviation of 3° for the angle defining the orientation of the C_2 axis of the acceptor water molecule with respect to

the O−O axis. On the basis of the reliable geometry delivered by jun-ChS, this method was used as reference for both equilibrium geometries and electronic energies.

Preliminary B3LYP-D3/aug-cc-pVTZ computations of the HCN⇌HNC reactive PES were refined at the jun-ChS level. The nature of the identified stationary points (minima or saddle points) was checked through frequency calculations performed at each level of theory. All calculations have been carried out with the Gaussian software[168], except the geometry optimizations at the ChS and jun-ChS levels, which have been performed using the CFOUR package[106,234]. Since revDSD-PBEP86 is not among the Gaussian built-in functionals, it has been defined by setting proper IOP flags on top of the DSD-PBEP86 functional. Full geometry optimizations were performed for the complexes containing 2 to 4 H_2O molecules, whereas for the 20 water model cut from the ice XI (010) surface, 8 molecules belonging to the cluster edge (see Figure **B.2**) were kept frozen at their positions in the crystal in order to prevent geometrical distortions causing a non-physical breakdown of the crystalline pattern. The best-performing methods were employed within a QM/QM′ strategy for simulating the HCN⇌HNC isomerization on this cluster in order to evaluate the catalytic effect of the ice surface. For the purpose, we employed the ONIOM method[299] treating the reaction center (i.e., the adsorbate and four water molecules) at a higher level of theory (i.e., a double-hybrid DF or even jun-ChS), whereas a less computationally-demanding method (i.e., a meta-hybrid DF) was used for the remaining molecules of the cluster. A much larger cluster containing 192 water molecules was also investigated by means of a three-layers (QM/QM′/MM) ONIOM approach enforcing the so-called mechanical embedding and employing the Amber force field[191]. In this case, the structural degrees of freedom of the adsorbate and the first 20 H_2O molecules were optimized while freezing the coordinates of the remaining 172 waters to those of the regular (010) surface of ice XI. Test computations with the more refined electrostatic embedding showed negligible differences on the relative energies.

Rate constants were computed solving the multi-well one-dimensional master equation using the chemically significant eigenvalues (CSEs) method [155]. Rate coefficients were determined using conventional TST within the RRHO approximation [300], also incorporating tunneling and non-classical reflection effects by means of the Eckart model [152]. The rates evaluated at different temperatures were fitted by a simple Arrhenius Equation or by the three-parameter modified Arrhenius equation proposed by Kooij [301,302]:

$$k(T) = A\left(\frac{T}{300}\right)^{n} \exp\left(-\frac{E_a}{RT}\right), \tag{4.1}$$

where A, n are the fitting parameters, and E_a the activation energy, R is the universal gas constant, and the limiting Arrhenius behaviour is recovered when $n = 0$. All the kinetic computations were performed with the MESS code [155].

4.3 Results and discussion

As widely discussed in the Introduction, the reliable modelling of interstellar ices is an extremely complex task, requiring the assessment of DFT methods for geometry and energy predictions that offer the proper balance between accuracy and computational burden. The lack of systematic studies addressing this issue for solid-state astrochemical processes calls for a dedicated benchmark. While small-size clusters cannot be fully representative of an extended substrate, the interaction of small molecules with water ice surfaces is generally guided by hydrogen bonds between the polar functional groups of the molecule and the exposed H and O atoms of the ice surface, which are already present in the smallest cluster models. Therefore, while the thermochemistry computed by using clusters composed of a small number of H_2O molecules is not representative of real icy-grain chemistry, the outcomes of the benchmark are safely transferable to larger clusters. In the following subsections we report the results of our benchmark study, concerning first geometries and then reaction and activation

energies. Finally, to scale-up to a more realistic water ice model, we report a full characterization of the PES of the HCN⇌HNC isomerization on clusters composed by either 20 or 192 H_2O molecules. We selected such sizes in order to expand the cluster without losing the pattern of the crystalline ice.

4.3.1 The geometry snow-board

The HCN⇌HNC isomerization is an endothermic process involving a high activation energy and the jun-ChS results are close to the current best estimates[303] for both the reaction (61.6 vs. 63.8 kJ/mol) and activation (198.5 vs. 201.1 kJ/mol) energy. Addition of two water molecules leads to the formation of a hydrogen-bonded van der Waals adduct featuring the interactions between the H atom of HCN and the oxygen of one water molecule and between the N atom and one hydrogen of the second water molecule. Then, the reaction proceeds through a transition state for the $(H_2O)_2$-mediated proton transfer reaching, in this way, a post-reactive complex in which carbon is engaged in a weak H-bond with a hydrogen of the first H_2O molecule, while the H atom of HNC interacts with the oxygen of the second water molecule. The structures of all the stationary points ruling the reactive PES are sketched in Figure **4.1** together with selected geometrical parameters obtained at the jun-ChS level.

The accuracy of the considered DFT model chemistries has been evaluated with respect to jun-ChS values and the overall mean absolute errors (MAEs) and mean absolute relative errors (REs) have been evaluated over all the bond lengths, valence and dihedral angles of the species involved in the PES. The full list of data can be found in Table **B.2** and Figure **B.1** of Appendix **B**. As a rule of thumb (with some exceptions for dihedral angles), triple-ζ basis sets show smaller errors than the corresponding double-ζ ones, with the improvement being less pronounced along the jun-, jul- and aug- series. In general, the tested hybrid and meta-hybrid DFs on the one side, and the double-hybrids on the other, give similar trends for the MAEs, with the notable exception of BHLYP-D3 in conjunction with the jul-cc-pVDZ basis set, that strongly overshoots and the

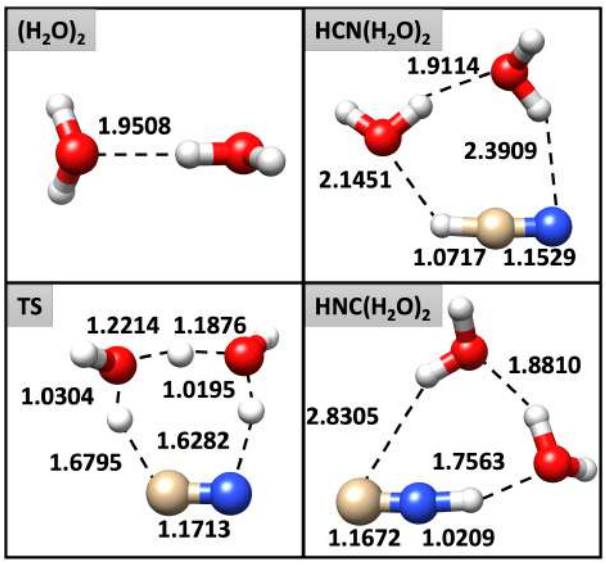

Figure 4.1 Stationary points on the reactive PES of the HCN⇌HNC isomerization catalysed by two water molecules. Representative bond lengths (Å) obtained at jun-ChS level are reported.

ωB97X-D functional that shows larger deviations from the jun-ChS reference values, especially for valence and dihedral angles. In the case of the BHLYP-D3/jul-cc-pVDZ model, MAEs as large as 0.09 Å, 6° and 10° were observed for bond lengths, valence and dihedral angles, respectively. These results are related to the inability of reproducing a tight structure for the post-reactive complex. Specifically, one H-bond in HNC–$(H_2O)_2$ (see Figure **4.1**) is broken and the product collapses into an open structure. All in all, it can be observed that the most promising (meta-)hybrid DFs are PW6B95-D3, BMK-D3, M06-2X and MN15 coupled to triple-ζ basis sets (or, at least, the jul-cc-pVDZ one). Concerning the double-hybrid functionals, the best structural predictions are delivered by DSD-PBEP86-D3 and revDSD-PBEP86-D3 that show comparable accuracy. In order to have a clearer picture of the performance of the different model chemistries in the prediction of the geometries involved in the HCN⇌HNC isomerization assisted by two water molecules, Figure **4.2** reports the overall REs of each method, evaluated by averaging the REs of the geometrical parameters of all the species on the reactive PES. Inspection of this figure reveals that, among the (meta-)hybrid DFs the best results for double-ζ basis sets are delivered by

PW6B95-D3 and BMK-D3. In particular, PW6B95-D3/jul-cc-pVDZ, BMK-D3/aug-cc-pVDZ and PW6B95-D3/aug-cc-pVDZ score REs in the 0.60 % - 0.74 % range. The PW6B95-D3 and BMK-D3 DFs are the best performers also in conjunction with triple-ζ basis sets showing REs around 0.55 %. Concerning the double-hybrid functionals, it is apparent that their use in conjunction with a double-ζ basis set does not justify the computational overload in comparison with hybrid functionals; however, both DSD-PBEP86-D3 and its revision predict improved geometries when employed in conjuncion with triple-ζ basis sets, reaching a RE of only 0.4 % for the jul-cc-pVTZ basis set. In passing, it is interesting to point out that, these functionals have also demonstrated to be excellent performers in predicting structural and spectroscopic properties of gas-phase molecules[304,305].

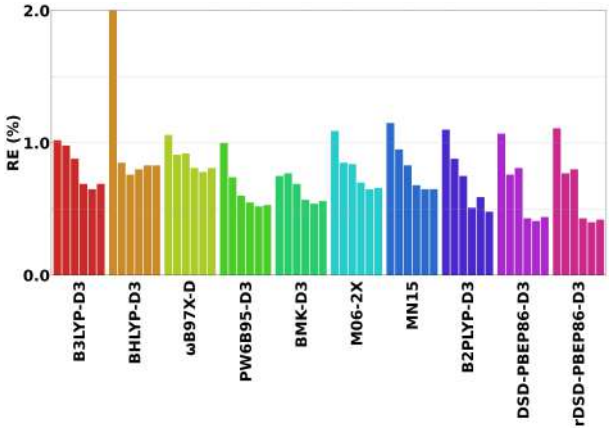

Figure 4.2 Total relative errors (REs) (%) of the geometries of the species on the PES of the HCN⇌HNC isomerization assisted by two water molecules for the investigated DFT methods with respect to jun-ChS reference values. For each functional, the different basis sets are reported in the following order: jun-DZ, jul-DZ, aug-DZ, jun-TZ, jul-TZ and aug-TZ.

4.3.2 Skiing on adsorption, reaction and activation energies

The functional/basis set combinations with the optimal accuracy/cost trade-off for geometry predictions have been identified in the previous section. Reactivity studies require the calculation of accurate formation and activation energies for the subsequent kinetic analysis. For this reason, some of the DFT methods

delivering the best geometrical predictions have been selected and their accuracy for computing adsorption, activation and reaction (electronic) energies explored using again jun-ChS results as references. In a first step, the impact of the geometry on the energetics has been assessed, by evaluating jun-ChS electronic energies for the different DFT structures. In a second step, the formation energies stemming from full DFT computations (for both geometries and energies) have been analysed.

Electronic energies obtained at the jun-ChS level on top of selected DFT geometries are reported in Table **4.1**, while the corresponding error analysis is presented in Figure **4.3**.

Table 4.1 jun-ChS formation energies (kJ/mol) with respect to isolated HCN and $(H_2O)_2$ for each species along the HCN/HNC···$(H_2O)_2$ isomerization PES evaluated on top of DFT geometries.

Level of theory for geometry	HCN···$(H_2O)_2$	TS	CNH···$(H_2O)_2$	HNC + $(H_2O)_2$
PW6B95-D3/jul-DZ	-33.38	99.37	15.23	62.31
BHLYP-D3/aug-DZ	-33.40	99.21	14.87	61.88
PW6B95-D3/aug-DZ	-33.44	99.40	15.17	62.28
BMK-D3/aug-DZ	-33.12	99.13	15.04	62.26
M06-2X/aug-DZ	-33.32	99.67	15.38	62.28
MN15/aug-DZ	-33.39	99.36	15.61	62.63
PW6B95-D3/jul-TZ	-33.49	99.15	14.73	61.89
BMK-D3/jul-TZ	-33.47	99.18	14.59	61.66
M06-2X/jul-TZ	-33.38	99.49	14.91	61.85
MN15/jul-TZ	-33.51	99.33	14.98	62.03
DSD-PBEP86-D3/jul-TZ	-33.52	99.35	15.06	62.22
revDSD-PBEP86-D3/jul-TZ	-33.54	99.31	15.02	62.22
jun-ChS	-33.42	99.26	15.03	62.27

Table 4.2 DFT formation energies (kJ/mol) with respect to isolated HCN and water dimer $(H_2O)_2$ for each species along the HCN···$(H_2O)_2$ isomerization PES.

Level of theory[a]	HCN···$(H_2O)_2$	TS	CNH···$(H_2O)_2$	HNC + $(H_2O)_2$
PW6B95-D3/jul-DZ	-34.12	86.96	8.34	56.74
BHLYP-D3/aug-DZ	-38.42	89.54	-1.35	51.12
PW6B95-D3/aug-DZ	-33.99	88.45	7.70	56.03
BMK-D3/aug-DZ	-36.30	81.76	-6.60	42.58
M06-2X/aug-DZ	-37.46	70.46	1.14	52.78
MN15/aug-DZ	-36.67	76.94	-2.82	47.31
PW6B95-D3/jul-TZ	-32.67	94.10	11.31	56.81
BMK-D3/jul-TZ	-35.97	88.16	0.11	48.23
M06-2X/jul-TZ	-37.16	76.00	4.64	54.06
MN15/jul-TZ	-35.26	85.53	1.92	48.88
DSD-PBEP86-D3/jul-TZ	-35.08	88.86	14.86	64.93
revDSD-PBEP86-D3/jul-TZ	-33.97	93.50	16.21	64.85
jun-ChS	-33.42	99.26	15.03	62.27

[a] For both energy and geometry.

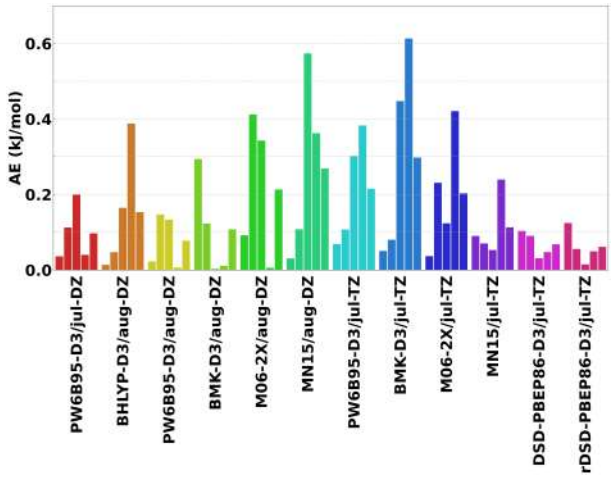

Figure 4.3 Absolute errors (AEs) analysis for jun-ChS formation energy (kJ/mol) obtained on top of DFT geometries in comparison with full (both energies and geometries) jun-ChS results. Each color corresponds to a DFT model chemistry and collects absolute errors for the formation energy of each species along the PES with respect to isolated reactants: 1. pre-reactive complex; 2. transition state; 3. post-reactive complex; 4. products; 5. MAE over all of the steps along the PES.

It is quite apparent that the energetic results obtained employing geometries optimized with all the tested methods are in remarkable agreement with the jun-ChS reference, with deviations smaller than 0.6 kJ/mol, even though some of them provide an unbalanced description of the different elementary processes. For example, the MN15/aug-cc-pVDZ and BMK-D3/jul-cc-pVTZ models yield excellent predictions of both the interaction energy of hydrogen cyanide with $(H_2O)_2$ and the transition state energy, with errors around 0.05 and 0.1 kJ/mol, respectively; however, the computed HNC formation energy (at the BMK-D3 level) and its interaction energy with the water dimer (at the MN15 level) show significantly larger errors. Among the (meta-)hybrid functionals, the best and most consistent energetic description is given by PW6B95-D3 in conjunction with jul- or aug-cc-pVDZ basis sets, which reaches an overall MAE (evaluated by considering the relative electronic energies of all the stationary points ruling the PES) close to 0.1 kJ/mol and a maximum deviation of 0.2 kJ/mol.

Moving to the double-hybrid DFs, the DSD-PBEP86-D3 and revDSD-PBEP86-D3 models in conjunction with the jul-cc-pVTZ basis set yield excellent

performances, scoring a MAE of about 0.06 kJ/mol and reproducing the formation energies of all the elementary steps with a maximum deviation of 0.12 kJ/mol for the pre-reactive complex at the revDSD-PBEP86-D3/jul-cc-pVTZ level. The relative electronic energies of all the stationary points fully evaluated at different DFT levels (i.e. energies and geometries) are collected in Table **4.2** and the MAEs from the jun-ChS computations are shown in Figure **4.4**.

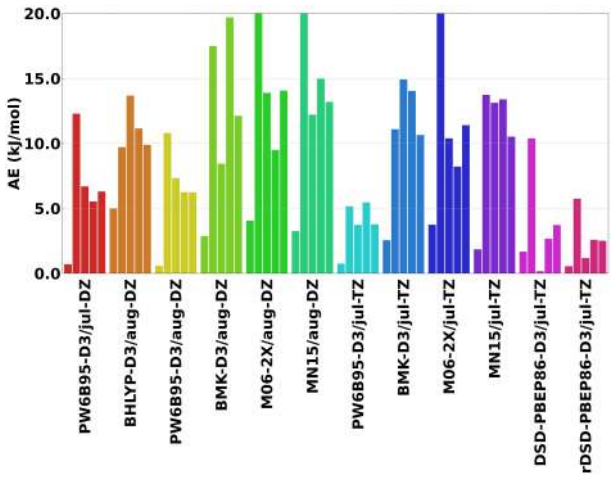

Figure 4.4 Absolute errors (AEs) analysis for DFT formation energies (kJ/mol) in comparison with jun-ChS values. Each color corresponds to a DFT model chemistry (used for both geometry and energy) and collects absolute errors for the formation energy of each species along the PES with respect to isolated reactants: 1. pre-reactive complex; 2. transition state; 3. post-reactive complex; 4. products; 5. MAE over all of the steps along the PES.

In general terms, the results mirror those obtained for jun-ChS energies evaluated on top of DFT geometries, with the only difference being the much larger deviations, which now span the 5 —29 kJ/mol range. Furthermore, the relative stability of CNH$\cdots$$(H_2O)_2$ is always strongly underestimated (becoming even negative with BHLYP-D3, BMK-D3 and MN15 functionals) except at the PW6B95-D3 and, especially, DSD-PBEP86-D3 and revDSD-PBEP86-D3 levels in conjunction with the jul-cc-pVTZ basis set. All the (meta-)hybrid DFs show MAEs larger than 10 kJ/mol, with the exception of PW6B95-D3, which is the only functional that reaches a MAE around 6 kJ/mol in conjunction with the jul- and aug-cc-pVDZ basis set and of 3.8 kJ/mol employing the jul-cc-pVTZ basis. The DSD-PBEP86-D3 and revDSD-PBEP86-D3 functionals confirm their good

performances in conjunction with the jun-cc-pVTZ basis set, with MAE around 3 kJ/mol and maximum deviations of 10.4 kJ/mol. Hence, the model chemistries with the optimal accuracy for structural parameters are also the best choices for thermochemistry. These results confirm the conclusions of recent benchmarks about the quality of PW6B95-D3/jul-cc-pVDZ and DSD-PBEP86-D3/jul-cc-pVTZ models for geometries, vibrational frequencies and other spectroscopic parameters.[304,305] Noted is that core-valence correlation has not been included for double-hybrid functionals because it was not taken into account in their original parametrization and its contribution is anyway within the expected error bar at least for molecular systems containing only hydrogen and second-row atoms (see Table **B.3** in Appendix **B** for CV contributions in jun-ChS results). Furthermore, some test computations performed with quadruple-ζ basis sets showed that CBS extrapolation has a negligible effect on all the trends discussed above. For example, the relative electronic energies of the stationary points obtained by using the DSD-PBEP86-D3 functional in conjunction with the aug-cc-pVQZ basis set (ΔE = -35.04, 89.51, 14.83 and 64.78 kJ/mol) differ from the counterparts obtained employing the jul-cc-pVTZ basis set by 0.65 kJ/mol at most (for the TS). Finally, although triple-ζ basis sets possibly deliver more robust results for hybrid functionals, this computational level will be used in the following only to describe small environmental effects in the framework of QM/QM′ computations where the increased computational cost with respect to double-ζ results is not justified, in our opinion, by the marginally improved robustness.

4.3.3 Scaling-up toward extended systems: best performers at work

The benchmark performed for both geometries and energies permits the identification of the best candidates for setting up a QM/QM′ ONIOM strategy for the study of the HCN⇌HNC isomerization on large clusters capable of providing a more realistic modelling of the icy-grain and of the molecule-surface

interactions.

At first, a cluster composed by 20 water molecules (shown in Figure **4.5** and Figure **B.2** of Appendix **B**) has been used, in which the pattern of exposed water molecules is suitable for a H-relay mechanism mediated by four water molecules. It should be noted that in ref.[287] a proton-relay mechanism mediated by three water molecules, in turn solvated by seven additional waters, was used. In the present work, the four H_2O molecules involved in the hydrogen transfer and the adsorbed species have been considered as the reaction center of the process under study, hence they constitute the higher-level QM portion of the system. Following the outcomes of the benchmark study, the DSD-PBEP86-D3 functional in conjunction with the jul-cc-pVTZ basis set has been used for the purpose, while the remaining part of the cluster, treated at a lower QM′ level, has been described by the PW6B95-D3 DF in conjunction with the jul-cc-pVDZ basis set. The energetic profile of the HCN⇌HNC isomerization occurring on the $(H_2O)_{20}$ cluster is reported in Figure **4.5** where it is also compared with that for the $(H_2O)_2$-mediated process. Going from the process assisted by two waters to that assisted by four water molecules in the $(H_2O)_{20}$ cluster lowers the energy of all the species present in the reactive PES. The most remarkable effect is the reduction of the energy barrier ruling the isomerization when considering the 20 water cluster in place of just two water molecules involved in the simplest possible relay mechanism.

The dependence of the energy profile on the number of water molecules involved in the relay mechanism was already pointed out[286,287]. However, only few water molecules were considered and no attempt to simulate the effect of ice bulk has been reported beyond the PCM level, whose reliability is, however, questionable for hydrogen-bonding solids. For comparison, Table **B.5** lists the relative energies. As it can be seen, the relative energy for HCN interacting with the water cluster is only marginally affected by the cluster size, but there is a huge effect on the activation barrier. While an overall fair agreement between the present results and those obtained in ref.[287] can be noted, there

Table 4.3 Relative ground state energies (kJ/mol) with respect to HNC-$(H_2O)_n$ post reactive complex and comparison with the results of ref.[287]. Both the total number of water molecules (n) and the number of water molecules directly involved in the relay mechanism (n_R) are indicated. All values include ZPVEs.

	total H_2O (n)	**relay H_2O** (n_R)	**TS**[a]	**HCN···$(H_2O)_n$**[a]
ref.[287]	2	2	74.1	-42.3
	3	3	43.9 (-30.2)	-41.4 (-0.9)
	10	3	13.8 (-60.3)	-41.4 (-0.9)
B3LYP[b]	2	2	73.8	-39.8
	3	3	44.1 (-29.7)	-41.7 (-1.9)
PW6B95-D3[c]	2	2	70.9	-43.6
	3	3	52.4 (-18.5)	-42.0 (-1.6)
	4	4	49.3 (-21.6)	-42.7 (-0.9)
	192	4	36.6[d] (-33.4)	-40.0[d] (-3.6)
DSD-PBEP86-D3[e]	2	2	68.1	-51.5
	3	3	48.3 (-19.8)	-49.6 (-1.9).
	4	4	46.1 (-22.0)	-47.8 (-3.7)
	20	4	32.3[f] (-35.8)	-43.7[f] (-7.8)
			32.1[g] (-36.0)	-41.7[g] (-9.8)
	192	4	32.4[h] (-35.7)	-45.7[h] (-6.8)
jun-ChS	2	2	78.3	-50.0
	3	3	58.5 (-19.8)[i]	-48.0 (-2.0)[i]
	20	4	44.1 (-34.2)[j]	-40.8 (-9.2)[j]

[a] In parentheses is the difference with respect to $(H_2O)_2$ results.
[b] 6-31+G(d,p) basis set as in ref[287].
[c] jul-cc-pVDZ basis set.
[d] QM/MM energies and ZPVEs. 20 waters molecules treated at PW6B95-D3 level, the remaining molecules described by the Amber force field.
[e] jul-cc-pVTZ basis set.
[f] ONIOM geometries and ZPVE. DSD-PBEP86/jul-cc-pVTZ for adsorbate and molecules involved in the relay mechanism, PW6B96-D3/jul-cc-pVDZ for the water molecules not involved in the relay mechanism.
[g] DSD-PBEP86/jul-cc-pVTZ energies while geometries and ZPVE at PW6B95-D3/jul-cc-pVDZ level.
[h] DSD-PBEP86:PW6B95-D3:Amber energies on PW6B95D3:Amber geometries. ZPEs at PW6B95-D3:Amber level.
[i] jun-ChS electronic energy, PW6B95-D3/jul-cc-pVDZ geometry and ZPVE.
[j] jun-ChS:PW6B95 electronic energy, PW6B95-D3/jul-cc-pVDZ geometry and ZPVE.

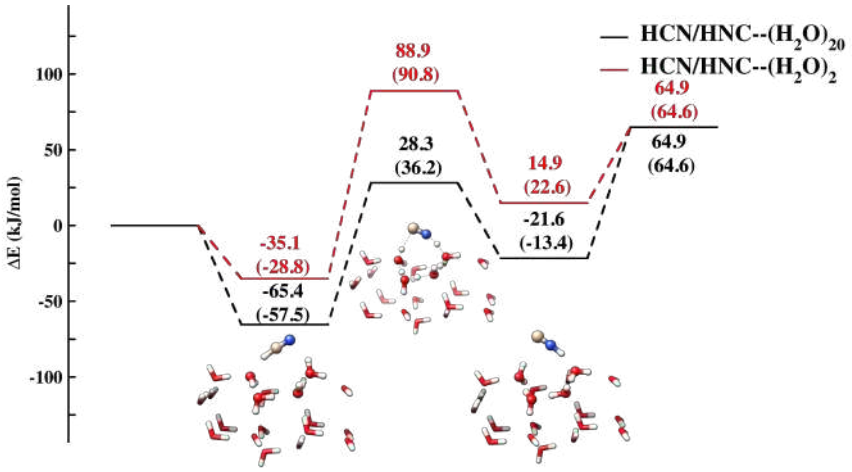

Figure 4.5 Potential energy profile for HCN⇌HNC isomerization mediated by the $(H_2O)_{20}$ cluster and the $(H_2O)_2$ dimer. Red lines refer to the HCN isomerization catalysed by $(H_2O)_2$ and both geometries and ΔE have been computed at the DSD-PBEP86-D3/jul-cc-pVDZ level. Black lines refer to the ONIOM results for the reaction catalysed by $(H_2O)_{20}$. The ball and stick representation is used for atoms of the highest QM level (DSD-PBEP86-D3/jul-cc-pVDZ) while the tube representation is used for the atoms belonging to the QM′ (PW6B95-D3/jul-cc-pVDZ) portion. ΔE corrected for ZPVE are reported in parenthesis with ZPVEs calculated at the same level of theory as the corresponding energies and geometries.

is a difference of about 18 kJ/mol for the energy of the transition state. This can be explained by considering that Koch *et al.*[287] investigated the role of the crystalline environment by optimizing for the different stationary points the positions of seven water molecules around the HNC···$(H_2O)_3$ complex without any constraint related to the arrangement of water molecules in icy structures. The importance of the morphological pattern in ice is highlighted by the present results: indeed, using a $(H_2O)_{20}$ cluster with the same molecular arrangement as in ice XI rules out the possibility of a process catalysed by two or three water molecules. Rather, the molecular arrangement at the surface permits a process assisted by four water molecules (see Figure **4.5**).

Further support to the reliability of the results is provided by the comparable barrier obtained by another ONIOM computation in which the high-level part of the system (HCN···$(H_2O)_4$) is treated at the jun-ChS instead of DSD-PBEP86-D3 level without any additional geometry optimization (last line of Table **4.3**). What

is even more gratifying is that the differences between the results obtained for the smallest HCN···$(H_2O)_2$ model and the larger model clusters (values in parenthesis in Table **4.3**) obtained at the DSD-PBEP86 level are in quantitative agreement with the jun-ChS counterparts. This paves the route toward the computation of very reliable parameters for reactions occurring on icy grains by combining jun-ChS results for small models and ONIOM(DSD-PBEP86:PW6B95-D3) values for large model clusters.

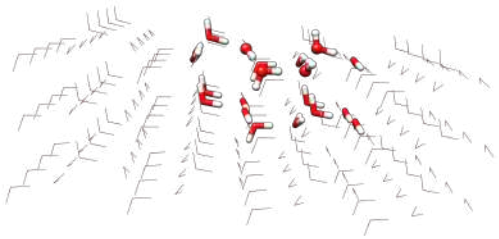

Figure 4.6 Structural model for the $(H_2O)_{192}$ cluster treated by three-layer ONIOM DSD-PBEP86:PW6B95:Amber strategy (geometry at PW6B95:Amber level). Ball and stick and tubular representation for the QM sections treated at DSD-PBEP86/jul-cc-pVTZ and PW6B95-D3/jul-cc-pVDZ level, respectively.

This approach can be further extended to very large models by employing a three-layer QM/QM′/MM ONIOM model. In order to analyze also this aspect, we have embedded the HCN···$(H_2O)_{20}$ cluster in a large model of ice-XI containing 172 water molecules described by the Amber force field (see Figure **4.6**). The results collected in Table **4.3** show that inclusion of the MM layer further stabilizes the HCN isomer with respect to the HNC counterpart by about 4 kJ/mol, but has a negligible effect on the energy barrier (less than 0.4 kJ/mol). Taking into account the estimated error bar of the overall computational approach (about 4 kJ/mol), the results obtained for the HCN···$(H_2O)_{20}$ model can be considered essentially converged with respect to further extension of the ice substrate.

4.3.4 Reaction rates

In an astrochemical context HNC can either isomerize to HCN or diffuse on ice surfaces and then react with another molecule (e.g., CH_2NH to produce acetonitrile) at the low temperatures typical of the ISM.

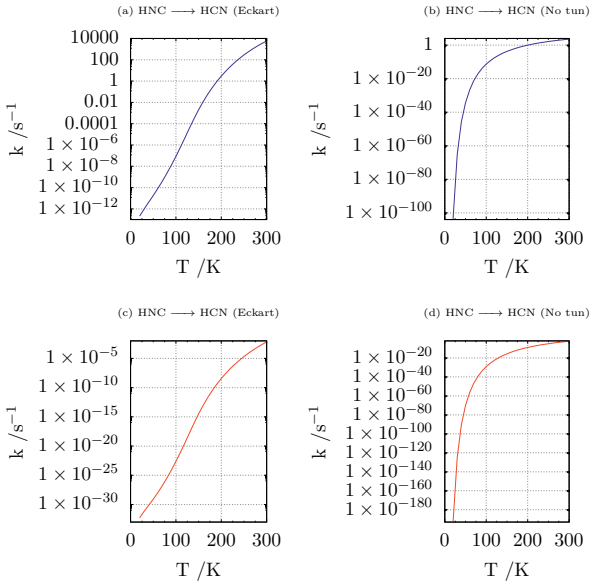

Figure 4.7 Reaction rates for the HNC⇌HCN isomerization including (Eckart) or excluding (No tun) tunneling. Panels a) and b) refer to the HNC···$(H_2O)_{20}$ model, whereas panels c) and d) refer to the HNC···$(H_2O)_2$ model.

The reaction rates computed for the HNC⇌HCN isomerization with the methodology described in Section **4.2** are shown in Figure **4.7**. It is apparent that the rates computed for the HNC···$(H_2O)_2$ system (Figure **4.7**, panels *c* and *d*) are very slow irrespective of the inclusion or not of tunneling. The situation is completely different for the HNC···$(H_2O)_{20}$ model, where the rate not including tunneling (corresponding to the one used by Koch and coworkers [287]) remains very small at low temperatures (Figure **4.7**b), but inclusion of tunneling (Figure **4.7**a) permits an effective reaction even at temperatures characteristic of the ISM. Noted is that the rates computed taking tunneling into account show a clear bimodal shape and cannot be fitted by a simple Arrhenius (or Kooij) function [301,302].

Unfortunately, the diffusion coefficients of HNC (or even HCN) on ice have not yet been reported[306]. According to a recent classification of ice adsorbates[307], HCN (hence probably HNC) is assigned to the intermediate class, which induces some deformation of the surface, but does not form hydrates nor penetrates rapidly into the ice bulk. An upper limit to the surface diffusion coefficient can be estimated with reference to the guess of $4 \times 10^{-11}\,\mathrm{cm^2\,s^{-1}}$ at 130 K provided by Livingston *et al.* for SO_2,[308] which corresponds to a mean distance of 1260 Å in 1 s. Since the computed isomerization rate at 130 K is about $10^{-4}\ \mathrm{s^{-1}}$ (which lowers to $1\times10^{-10}\,\mathrm{s^{-1}}$ at 50 K), the average diffusion of HNC before isomerization can reach 100 Å at 130 K (1 cm at 50 K).

Therefore, if the formation of aminoacetonitrile is faster than the isomerization to HCN when HNC and CH_2NH are nearest neighbors[285], our results suggest that diffusion of HNC along significant distances could permit the formation of aminoacetonitrile on icy grains containing CH_2NH even at low concentrations.

4.3.5 Conclusions and outlook

The main aim of this work was the implementation and validation of a general computational strategy for the study of the thermochemistry and kinetics of chemical processes taking place on interstellar icy-grains. To this end, composite methods rooted in the coupled cluster ansatz have been combined with hybrid and double-hybrid functionals together with molecular mechanics force field to characterize the stationary points ruling the reactive potential energy surfaces on model clusters sufficiently large to minimize spurious boundary effects. Next powerful master equation / RRKM models have been employed to compute reaction rates including tunneling effects. As a demanding test case we have selected the HCN/HNC reactions for which the available computational results are not fully satisfactory.

Ten different (meta-)hybrid and double-hybrid density functionals have been considered in conjunction with the jun-, jul- and aug-cc-pVnZ basis sets of double- and triple-ζ quality and their accuracy in predicting geometries together with

thermochemical and kinetic data (adsorption, activation and reaction energies) has been assessed in comparison to reference values computed using the jun-ChS composite method. This benchmark has led to the conclusion that, among (meta-)hybrid functionals, BMK-D3 and PW6B95-D3 in conjunction with partially augmented double- and triple-ζ basis sets yield the most reliable description of geometries, with the optimal trade-off between accuracy and computational cost being offered by the PW6B95-D3/jul-cc-pVDZ model chemistry. Concerning double-hybrids, DSD-PBEP86-D3 and revDSD-PBEP86-D3 in conjunction with the jul-cc-pVTZ basis set deliver accurate predictions of both geometries and reaction energies. Next, these outcomes have been used to investigate the effect of cluster size and ice surface on the isomerization process of HCN. In particular, a cluster containing 20 water molecules has been cut from the (010) surface of ice XI, and used in a multiscale ONIOM calculation, in which the reaction center has been modeled at the DSD-PBEP86-D3/jul-cc-pVTZ level, while for the remaining portion of the $(H_2O)_{20}$ cluster the PW6B95-D3 functional has been employed in conjunction with the jul-cc-pVDZ basis set. This approach has allowed the proper modelling of the surface with an accurate, yet, cost-effective strategy. The pivotal role of the structural arrangement of surface molecules in driving the evolution of catalytic processes has been pointed out. The accuracy of the results has been further improved by combining jun-ChS results for small models to QM/QM′ (DSD-PBEP86:PW6B95-D3) values for medium size model clusters and/or three-layer QM/QM′/MM computations for very large clusters. On top of these computations, reaction rates have been computed by methods rooted in the transition state theory including tunneling which plays the dominant role at low temperature for processes involving the motion of light atoms. At variance with previous investigations, our results show that the isomerization is ruled by a proton relay mechanism directly involving four water molecules, but tuned by relatively distant waters belonging to the model cluster employed to mimic the ice surface. The resulting activation energy is strongly reduced with respect to that governing the isomerization of the bare HCN molecule, but

only tunneling allows for effective isomerization of HNC in the harsh conditions characterizing astrochemical processes.

Together with the intrinsic interest of the studied system, the results of the present work have allowed to define the best strategy for future modelling of iCOMs-ices interactions in the framework of a QM/QM′/MM approach. This also represents the starting point for hybrid QM/QM′/periodic approaches, in which the outcome of the multiscale (QM/QM′) description of the cluster is corrected for environmental effects obtained by simulating the surface using periodic boundary conditions[309]. However, the crystalline water ice surfaces usually employed to simulate icy dust grains could be inadequate to describe their amorphous structure. Work in this and related directions is under way in our laboratory in order to achieve a more realistic modeling of chemical processes occurring on icy mantles of interstellar grains.

Chapter 5

iCOMs at ice interface

5.1 Introduction

The contribution of iCOMs to the occurance of possible debated prebiotic or biotic life is still an open discussion. Containing six or more atoms[26,310], iCOMs possess unique characteristics and their potential in the evolution of a carbon-based chemistry cannot be overlooked. It is thought that iCOMs formed during the early stages of stellar evolution are inherited by the organic material incorporated in the newly born solar system[2]. Therefore, studying the chemistry of iCOMs in different stages of stellar evolution can enhance our understanding of the molecular inheritance in newly formed planetary systems, including our own Solar System[311]. Indeed, chemical and stellar evolution are strongly connected. Prestellar cores are the nurseries of new stars and they are characterized by very low temperature and density. In such conditions, atoms and molecules stick on the surface of ice and grains and hydrogenation seems the only way for any chemistry to evolve. Active hydrogenation processes lead to the formation of icy mantles covering interstellar solid particles with molecules trapped within the solid matrix. Spectroscopic techniques scanning the IR region are used to detect the spectroscopic fingerprints of these trapped molecules.

As discussed in the Introduction to this thesis (see Chapter **1**), radio-

astronomy is the golden technique for the detection of small molecules in the gas-phase, while the IR window is more suited to study interstellar ices. Laboratory experiments intend to reproduce the extreme space conditions, however, the fine tuning of the operational conditions is challenging and has an high impact on the morphology of the synthesised ice structures.

Computational astrochemistry plays a pivotal role in comprehending the involved chemistry. With this intent, we developed a procedure for the simulation of the vibrational fingerprint of iCOMs at the interstellar ice interface.

In this chapter, we consider three isomers of formula $C_2O_2H_4$: acetic acid (AA), glycolaldehyde (GA) and methyl formate (MF) (see Figure **5.1**).

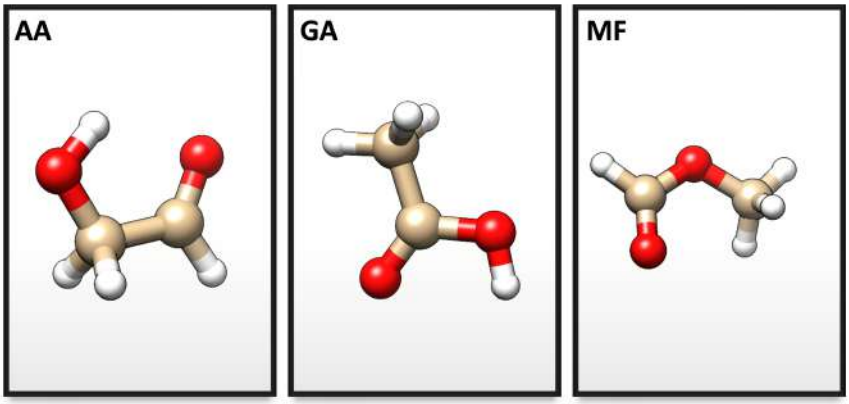

Figure 5.1 Optimized gas-phase structures of acetic acid (AA), glycolaldehyde (GA) and methyl formate (MF) at B2PLYP-D3/jul-cc-pVTZ level.

AA has been detected in 1997 toward Sgr-B2(N-LMH)[312] and has gained attention for its potential role as a precursor toward the synthesis of glycine. The formation mechanisms of acetic acid are still debated and so far, the proposed pathways do not fully explain the observed abundances[313,314]. GA is considered the simplest sugar and has been detected in comets like Lovejoy and 67P/Churyumov-Gerasimenko and other regions like, among the others, the hot molecular core G31.41+0.31 and the Class 0 protostellar binary IRAS

16293-2422[28,315]. The formose reaction, involving formaldehyde dimerization as first step in an auto-catalytic process resulting in sugars of higher molecular weight is the most accredited biochemical mechanism. However, the condition of the ISM would prevent this reaction to evolve in the gas-phase and alternative photo-activated pathways involving interstellar ices have been proposed[34]. MF is the most abundant among the three isomers and it has been detected in several astrophysical environments[316].

The relative abundances of these isomers reflect active chemical processes in both gas-phase and on the surface of interstellar ice. The strength of the interaction between the isomer and the ice surface and the physical properties of the specific environments such as the temperature influence the proportions of isomers trapped in the ice or released into the gas phase. AA exhibits a strong interaction with the watery matrix and has a higher desorption temperature compared to the solid matrix itself[317]. MF, on the other hand, shows a weaker interaction with the ice surface and is released at lower temperatures[317]. Studying the interaction of each isomer with the icy surface is useful to provide insights into chemical abundances recorded in the gas phase.

Since interstellar ices predominantly consist of water, research in the literature has concentrated on the adsorption/desorption mechanisms of AA, GA, and MF both on crystalline (CI) and amorphous ices (ASW).

Typically, the thermal evolution of chemical components is studied using a combination of IR spectroscopy and thermal programmed desorption (TPD) analysis. In 2005, Gao and Leung examined the thermal evolution of AA nanodeposits on nano-crystalline and poly-crystalline ice films within the temperature range of 123 —180 K[318]. Their work highlighted the significant IR features induced by the presence of the icy substrate compared to pure AA ice, providing the first IR data for AA adsorption in that temperature range. At temperatures as low as 123 K, the ice deposit predominantly exhibits an amorphous structure, with a noticeable reordering occurring between 153 and 173 K. The adsorption of AA can be understood at two levels. At the

sub-monolayer/monolayer level, individual AA monomers strongly adsorb onto the dangling hydroxyl (OH) groups of the ice surface, leading to a red-shift in the C=O stretching frequency from 1795 cm^{-1} for the gas phase molecule to around 1695 cm^{-1}. At the multilayer level, the main signal is observed around 1717 cm^{-1}, which closely resembles the spectral features of dimeric and chain-like structures typical of pure AA ice. This assignment is supported by the correlation between the C=O band and OH band of water ice, where the increased intensity of the signal at 1695 cm^{-1} is correlated with a decrease in the intensity of adsorption feature from the dangling OH bands, indicating saturation of these groups. The strong interaction between AA and the water ice surface results in coevaporation of AA with water. However, the exact structure responsible for this signal cannot be determined through spectroscopy alone, and computational studies have suggested the nature of the specific interactions formed between the individual iCOM and the ice surface[319]. Later, Bertin *et al.* investigated the same process on both ASW and CI between 80 K and 200 K, respectively[320]. In their work, AA was deposited in the form of dimers, making it challenging to determine the interaction of individual AA monomers with the ice matrix. The main feature of AA appeared at 1720 cm^{-1}, consistent with the band measured by Gao *et al.* in the multilayer regime and corresponding to the region typically assigned to acetic acid dimers (AA2). As the temperature increased, the vibrational spectra showed a drastic change toward a chain-like polymeric organization, similar to pure AA ice. These strongly bound units desorbed at temperatures higher than the surrounding watery matrix. More recently, Burke *et al.* further expanded the temperature range from 20 K to 180 K and investigated different AA/water ice mixtures with layered and mixed structures[317,321]. By studying the characteristics of both TPD and IR profiles of the pure iCOM ices and when co-deposited with water ice, they spotted clear differences in the desorption behaviour with respect to the pure AA ice as a consequence of the trapping and desorption behaviour in water ice[321]. The analysis provided in this work supports the formation of AA

cyclic dimers as main species formed within the monolayer regime thus being somewhat in constrast with the finding of ref.[318]. They also ascribe a signal at 1760 cm^{-1} to monomers on the surface. Moreover, at higher coverage, it is also observed an effect of AA on the morphology of ice with AA preventing ice crystallization at 140 K. The findings from experimental investigations support the formation of a strong interaction between AA and the H_2O ice surface. However, some results are inconsistent with each other, indicating the complexity of interactions involved in the formation of icy AA layers and the variety of oligomers/polymers that can originate. Additionally, the dependence of experimental results on operational conditions complicates a comprehensive discussion of the chemistry. Differences in experimental procedures, such as variations in crystallinity, may contribute to the discrepancies observed between the results of Burke *et al.*[317] and Bertin *et al.*[320]. In 2009, Bahr *et al.* conducted comprehensive experimental and computational characterizations of AA-water interactions[319,322]. AA is known to form dimers in the gas phase, but in aqueous solution, hydrated monomers and polymers become more prevalent with increasing acid concentration. DFT calculations suggested a preferential formation of a bidentate species when AA is adsorbed at the ice interface. The calculated density of states (DOS) compared with TPD/IR and metastable induced electron spectroscopy (MIES) indicated a preference for the adsorption of AA cyclic dimers. However, the coverage at which spectra were recorded was insufficient to investigate the structures involved at the sub-monolayer or monolayer level, specifically the first layer of AA monomers on the surface. Assigning the monomers to small signals around 1294 and 1695 cm^{-1} was done following the results of a previous work by Gao *et al.*[318]. As the coverage increased, the formed multilayers were mainly characterized by distorted dimers and linear chains, resembling the behavior observed in pure AA ices. These signals are the most prominent in the spectra acquired by Burke *et al.*[317] and Bertin *et al.*[320], where the recorded IR spectra are dominated by signals around 1717 cm^{-1}, compatible with the spectral features of a distorted cyclic dimeric

structure. Due to the strong interaction between AA and the ice surface, the classification according to Collings *et al.* categorizes AA as a complex water-like species[323]. This strong interaction is also observed in the TPD profile, where the desorption temperature of AA at the ice interface is equal to or even higher than that of water. Glycolaldehyde is a complex intermediate species, and methyl formate behaves more like a CO-like species, with the weakest interaction with ice among the reported isomers. This reactivity trend is in line with the chemical intuition that suggests carboxylic acids (AA), aldehydes (GA), and esters (MF) to exhibit decreasing interaction strengths with the ice surface. GA, like AA, is known to undergo polymerization, and the variety of conformers GA can form[269] adds to the complexity of the system. However, GA conformers on ice surfaces have not been extensively studied, and the focus in this study is on the most stable conformer, which is the only one observed in the gas-phase[324]. Among the reported isomers, MF is the easiest system to study due to its stability and reduced tendency to polymerization and interaction with the surface. Consequently, MF desorbs from the ice surface at the lowest temperature (115 K)[317]. To date, few studies have explored the interaction of these isomers on interstellar ice analogues. However, most of them have not thoroughly characterized the amorphous nature of the ice, and low-density amorphous ice (LDA), which is believed to be most representative of the conditions in interstellar environments[325] has not been addressed neither experimentally nor theoretically. Computational vibrational spectroscopy offers valuable microscopic insights into the nature of these interactions. A powerful predictive tool would be beneficial in distinguishing the different isomers when interacting with icy systems of astrochemical interest. For this reason in this chapter a computational strategy is developed and applied to simulate the IR fingerprints of acetic acid (AA), glycolaldehyde (GA), and methyl formate (MF) adsorbed on water ice.

5.2 Computational details

To investigate processes occurring on solid surfaces, two main approaches are commonly employed: the molecular cluster method and the periodic boundary conditions (PBC) approach (Section **2.4**). The cluster approach takes advantage from the use of molecular codes that are more advanced than periodic ones. As such it offers advanced tools (i.e., inclusion of anharmonic effects, intrinsic reaction path coordinates, and reliable transition state search algorithms) for the simulation of both spectroscopic properties and reaction pathways. The PBC (periodic) approach has the main advantage of providing a reliable description of the morphology of crystalline solids, avoiding finite size effects, and of allowing a consistent description of the electronic distribution. In this study, we adopt a combined approach that incorporates both methodologies. By leveraging the strengths of each method, we aim to simulate the adsorption of AA, GA, and MF on the surface of ice. This combined approach enables us to simulate characteristic vibrational fingerprints associated with these adsorbates beyond the double-harmonic approximation. By employing the molecular cluster method, we can capture the intricate details of the localized adsorption process, while the PBC approach enables us to properly model the extended ice surface. The synergy between these approaches allows us to achieve an atomistic understanding of the adsorption phenomena under investigation.

We used CRYSTAL17[179] for the PBC calculations of the CI model. We optimized both the cell constants and the atomic positions of the bulk of ice XI (space group $Cmc2_1$). These calculations were performed at the PBE0-D3 level, which combines the PBE0 functional[326] with Grimme's dispersion D3 and BJ damping[146,147]. The pob-TZVP basis set, optimized for solid-state calculations, was employed[327]. The calculated cell constants at PBE0-D3(BJ)/pob-TZVP are within 10% of the experimental ones which have been determined to be a = 4.502 Å, b = 7.798 Å, c = 6.822 Å[328] (see Figure **5.2** for the present theoretical results). To create the surface model, we sliced a three-layer slab along the

(010) surface, a commonly model surface to study reactivity in processes of astrochemical interest[277]. Further optimization of atomic positions was carried out on this slab (Figure **5.2**).

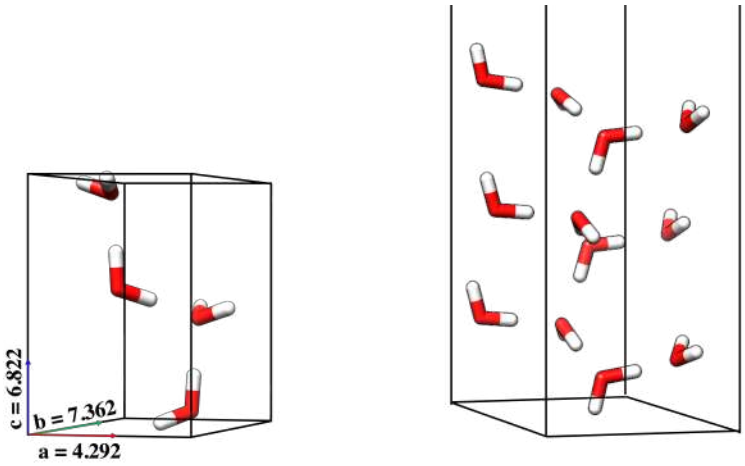

Figure 5.2 Optimized bulk (left) and slab (right) for crystalline ice XI. The slab model includes space along the z axis in order to correctly represent a surface model. Cell constants at PBE0-D3/pob-TZVP reported in Å.

We studied the adsorption of AA monomer, AA dimer (AA2), GA and MF on the optimized CI slab. To prevent lateral interactions we used a 2×2 supercell for AA, GA and MF and a 3×3 supercell for AA2. We used PBE0-D3(BJ) in conjunction with the cc-pVDZ basis set for the iCOM and the pob-DZVP for the surface. To determine the adsorption energy (E_{ads}) for each optimized system (iCOM@CI), we employed the following equation:

$$E_{ads} = E_{iCOM@CI} - E_{iCOM} - nE_{CI} - \Delta E_{BSSE}, \tag{5.1}$$

where, $E_{iCOM@CI}$ represents the energy of the organic molecule adsorbed on the CI surface, E_{iCOM} is the gas-phase energy of the organic molecule, nE_{CI} is the energy of the CI surface with the same number of surface atoms as in $E_{iCOM@CI}$, n is the ratio between the number of atoms in the slab and in the bulk ensuring we are comparing systems with the same number of atoms. In calculations involving Gaussian basis sets, it is necessary to account for the basis set superposition error (BSSE) in the calculation of the interaction energy[329].

The term ΔE_{BSSE} in the adsorption energy equation is computed as follows:

$$\Delta E_{BSSE} = (E_{iCOM}^{full} - E_{iCOM}^{mol}) + (E_{CI}^{full} - E_{CI}^{surf}), \tag{5.2}$$

where the superscript $full$ refers to the entire basis set used for the adsorption (including both surface and molecule) while mol and $surf$ refer to the basis set used for the iCOM and the surface simulations respectively. Single-point calculations were performed separately for the iCOM and iCOM@CI systems of interest. The input files were prepared with appropriate basis sets defined for each atom in the system.

On the optimized structure harmonic frequency calculations were performed. To reduce the computational effort we reduced the evaluation of the Hessian exclusively to the atoms involved in the carbonyl function (carbon and oxygen). We validated the accuracy of such procedure by performing additional frequency calculations that included either all atoms of the organic molecule or the water molecules participating in hydrogen bonding with the organics. At this stage, our main focus was to obtain a reliable structural model rather than accurate vibrational data. While the PBC approach allows a proper description of the CI morphology, the reliable prediction of vibrational frequencies for the adsorbed molecules requires the inclusion of anharmonic effects. For the purpose a cluster composed of three water ice layers and the adsorbed iCOM was cut from the periodic structure. The obtained cluster was re-optimized at QM level. We employed the ONIOM approach, as developed and tested in Chapter **4**[75]. The iCOMs and four water molecules involved in the adsorption process were assigned to the QM region, while the remaining molecules were assigned to the QM′ region and treated at a lower level of theory. Specifically, we used the DSD-PBEP86/jul-cc-pVTZ[90,143] method for the QM region and PW6B95-D3/jul-cc-pVDZ[90,133] for the QM′ region. To maintain the structural representation of the bulk, only the molecules in the QM region were optimized. Finally, anharmonic contributions were obtained by using GVPT2 with reduced dimensionality schemes. We

considered two RD approaches. In RD1 only diagonal contributions to the anharmonic force field (f_{iii} and f_{iiii}) were calculated. In RD2 all the force constants involving the active mode were calculated.

In the second part of our study, aimed at accurately simulating Low-Density Amorphous (LDA) ice by using MD simulations. First, we employed GenIce2 to construct a hydrogen-disorder configuration of the ice model[330,331]. A cubic cell with dimensions of 5.07 nm was utilized. Initially, the system was relaxed using a steepest descent algorithm. Subsequently, equilibration was achieved in the NPT ensemble, with temperature and pressure controlled by coupling to a V-rescale thermostat[190] and a Berendsen barostat[189], respectively. The temperature was set to 77 K and the pressure to 1 bar. Leap-frog integration with a time step of 1 fs was employed for a total simulation time of 5 ns. Equilibration was considered attained when both density and energy remained constant.

Exploring the morphology of ice presents considerable complexity, as evidenced by recent research at the University College of London, where a new form called Medium Density Amorphous (MDA) ice has been identified[332]. However, under ISM conditions, LDA is believed to be the most prevalent form of ice[325], even though many studies on the simulation of amorphous ice models for astrochemical processes have utilized ice forms with higher density[333,334]. Initially, attention was paid in building a proper model of the LDA morphology and in selecting a suitable force field to reproduce its properties. In particular, the TIP4PQ/2005 was used at first, however, as it is well known, rigid models employing Lennard-Jones potentials to describe non-bonded interactions fail to predict IR and Raman spectra. Therefore, the flexible TIP4P/2005f force field was also considered[335]. To analyze the structural properties of the simulated LDA ice, the average density (d), the radial distribution function (RDF) and the orientational tetrahedral order (oto) were calculated.

The RDF, which characterizes the local structure and intermolecular correlations in the amorphous ice system, was computed using the *gmx rdf* tool from the GROMACS package. The RDF provides information about the distribution

of molecules around a central molecule as a function of radial distance. The (RDF) or pair correlation function $g_{(AB)}(r)$ between particles of type A and B is defined in the following way:

$$g_{(AB)}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{\text{local}}} = \frac{1}{\langle \rho_B \rangle_{\text{local}}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta\left(r_{ij} - r\right)}{4\pi r^2}, \tag{5.3}$$

with $\langle \rho_B(r) \rangle$ the particle density of type B at a distance r around particles A, and $\langle \rho_B \rangle_{\text{local}}$ the particle density of type B averaged over all spheres around particles A with radius $r_{\max}$. $r_{\max}$ is usually set as half of the box length. By averaging over the entire simulation trajectory, we obtained a comprehensive representation of the system's intermolecular distances. The RDF analysis reveals characteristic peaks corresponding to specific intermolecular distances, providing information about the coordination shells around each water molecule. Moreover, the integration of the RDF provides the number of particles within a distance r. Specifically, in the case of water, the first inflection point in this curve corresponds to the coordination number in the first shell around a central water molecule.

The *oto* is the most common type of tetrahedral order parameter. This order parameter uses the four closest water oxygen neighbors for the calculation. The orientational tetrahedral parameter q is calculated as:

$$q = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos \psi_{jk} + \frac{1}{3} \right)^2, \tag{5.4}$$

where ψ_{jk} is the angle formed by the oxygen atom under consideration and the two nearest neighbor atoms j and k. The value of q ranges from 0 to 1 , where 0 is for an ideal gas and 1 is for a regular tetrahedron. To evaluate the *oto* in the amorphous ice, we employed the ORDER software[336]. This allowed us to quantify the extent to which water molecules arrange themselves in a tetrahedral-like configuration resembling the structure of ice Ih.

In order to investigate the influence of ice morphology on the adsorption process of iCOMs on LDA ice, we prepared a 1.5 nm thick ice slab from the

LDA bulk. To represent strong and weak interactions with the ice, we selected two iCOMs, AA and MF. The CHARMM General Force Field (CGenFF) was employed to describe the organic molecules[202]. Using the *insert-mol* tool available in GROMACS[187,188], we generated four different starting configurations for each iCOM. Each configuration was equilibrated in the NPT ensemble following the same procedure used for the LDA bulk system. The simulations were conducted for 2 ns per configuration, resulting in a total simulation time of 8 ns. In all cases, the adsorption process occurred nearly instantaneously, and the iCOMs predominantly remained localized on the adsorption sites without surface diffusion.

MD simulations provided a comprehensive sampling of the phase space for the adsorption process. However, accurate spectroscopic predictions of vibrational fingerprints require accounting for anharmonicity using QM methods. To this end, it has to be noted that performing anharmonic QM calculations on the whole system used for MD simulations (consisting of more than 10000 atoms) is not practical. Furthermore, computing the vibrational spectrum for each frame in the MD trajectory is not feasible in terms of computational time, but even not necessary. To overcome this issue, the MD trajectory was clustered. We employed a *tcl* script in VMD to identify the number of hydrogen bonds between the surface and the iCOMs for each frame, along with the water molecules involved in these interactions. This information allowed us to extract smaller clusters containing the iCOMs and the water molecules within a specific distance from the organics. Care was given to include all water molecules that formed hydrogen bonds with the adsorbate during the MD trajectory (see Figure **5.3**). Subsequently, the original trajectory was divided into groups based on the number of hydrogen bonds between the iCOMs and the superficial water molecules. Each of these sub-trajectories was then clustered according to structural similarities to obtain representative geometries for each group. We employed the *gromos* clustering algorithm with a cutoff of 0.3 nm, ensuring that over 60% of the configurations were assigned to the same cluster. Ultimately,

different configurations representative of all possible interactions between the organic molecules and the ice surface were collected and fed into QM calculations.

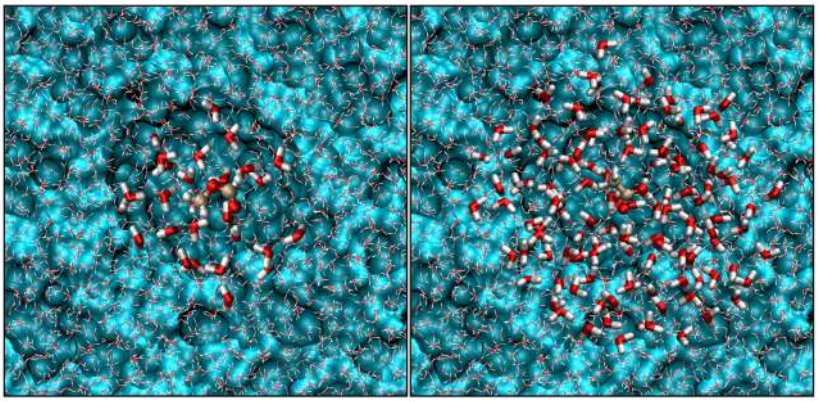

Figure 5.3 Top view of methyl formate and acetic acid adsobed on top of LDA surface labelled as MF@LDA (left panel) and AA@LDA (right panel) respectively. The water molecules highlighted with the licorice representation have been selected as within 6 Å of MF and 12 Å of AA. This group (organic + liquorice water) has been used for the further clustering and subsequent ONIOM calculations.

QM simulations were carried out adopting the same two layers ONIOM approach described above for the absorption on CI. It has to be noted that in the case of AA adsorption on LDA surface, two clusters of different sizes were selected to avoid the molecule sitting at the edge of the cluster in certain configurations. We calculated the adsorption energy of each of the obtained clusters as follows:

$$E_{ads} = E_{iCOM@LDA} - E_{iCOM} - E_{LDA}, \tag{5.5}$$

where $E_{iCOM@LDA}$ is the ONIOM energy of the optimized cluster where the iCOM is adsorbed, E_{iCOM} is the energy of the organics as calculated at the same level of theory as the QM region and E_{LDA} is the energy of the extracted ice cluster using the same QM/QM′ partition used for the evaluation of $E_{iCOM@LDA}$. The vibrational analysis was performed as described above for the case of CI.

5.3 Results and discussion

5.3.1 Adsorption of $C_2O_2H_4$ isomers on crystalline ice

The optimized geometries of the clusters describing the adsorption of AA, AA2, MF and GA on the CI surface extracted from ice XI cut along the (010) plane are reported in Figure **5.4**.

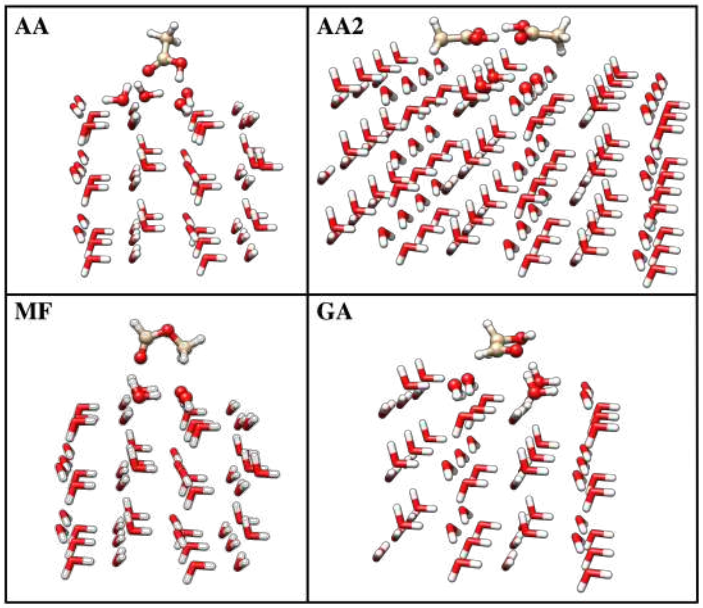

Figure 5.4 DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ optimized clusters for AA, AA2, MF and GA adsorption at the CI (010) surface. Ball and stick representation used for the QM region treated at DSD-PBEP86/jul-cc-pVTZ level. Licorice representation for the QM′ region treated at PW6B95-D3/jul-cc-pVDZ level.

Previous theoretical studies have assessed the adsorption of AA and MF on the CI surface[319,337]. AA and MF adsorb preferentially with a bidentate and a monodentate configuration, respectively. The adsorption configurations here obtained are consistent with those reported in the literature. To the best of our knowledge no previous theoretical studies assessed the adsorption of GA on CI surface. Bahr *et al.* studied the interaction between AA and CI surfaces both theoretically and experimentally[319,322]. They modelled a wide

variety of monomers and dimers interacting with the ice surface and suggested the most stable interactions. For AA, two hydrogen bonds involving both the hydroxyl (as donor) and the carbonyl (as acceptor) functions seem to be key in the physisorption at the CI surface. While in the case of AA2, there is a preference in keeping intact the cyclic dimeric structure instead of generating H-bonds with the surface which is coherent with our geometry for AA2@CI reported in Figure **5.4**. The adsorption energies for the studied systems are collected in Table **5.1** along with the results from the frequency calculations. The calculated adsorption energy for AA is -18.9 kcal/mol and agrees quite well with the previous calculated values of ~ -16 kcal/mol[319,337]. For AA2 a value of -14.3 kcal/mol is obtained.

The more negative the adsorption energy the stronger the interaction with the CI surface is. According to our results the strength of the interaction with the surface follows the order:

$$AA > GA > AA2 > MF, \tag{5.6}$$

which agrees with the classification proposed by Burke *et al.*[317].

Table 5.1 Adsorption energies (kcal/mol) and C=O stretching frequencies (cm^{-1}) obtained for AA, AA2, MF, and GA adsorbed on the (010) surface of ice XI.

	E_{ads} This work	E_{ads} Literature	ω (C=O)[a] This work	ν^{RD1} (C=O)[b] This work	ν^{RD2} (C=O)[c] This work	ν (C=O) Exp.
AA	-18.90	-16.43[319], -16.2[337]	1773	1767	1741	1695[318], 1747[322]
AA2	-14.30	-9.08[319]	1761	1764	1701	1720, 1717[320,322]
GA	-14.73	-	1728	1715	1681	-
MF	-14.17	-9.07[320], -8.9[337]	1760	1751	1721	1726[320]

[a] Harmonic frequency at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.
[b] Anharmonic frequency according to RD1 scheme at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.
[c] Anharmonic frequency according to RD2 scheme at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.

For what concerns the frequency prediction we should consider that experiments used here as reference have been performed at the following temperatures. Data for AA and AA2 as reported by Bertin *et al.*[320] refers to 155 K while results for MF have been reported at 80 K. The band assignment in the case of

AA needs a more in-depth discussion.

In this section, we examine the calculated IR frequencies for the carbonyl stretching vibrations of AA monomers when adsorbed on a crystalline ice surface and compare them to experimental data. We consider several relevant studies that provide insights into the interaction of AA with the CI surface and the associated vibrational frequencies.

Gao *et al.*[318] reported the identification of monomers on the ice surface with a characteristic peak at 1695 cm^{-1}. They correlated the saturation of this peak with the disappearance of the dangling OH stretching frequency of water as a sign of full monolayer coverage. Bertin *et al.*[320] investigated the deposition of AA at sub-monolayer levels and suggested that dimers interact with the ice surface, forming a slightly distorted cyclic structure. They used acetic acid dimers as a depositing agent. This choice of depositing agent may have influenced their ability to observe monomers. This implies the possible involvement of dimers in the adsorption process. Similarly, Bahr *et al.*[319,322] also proposed that dimers could interact with the ice surface. However, one should consider that Bahr *et al.* used MIES as a technique that primarily probes the topmost layer of the surface. Moreover, the coverage employed in their work might be too high to observe phenomena occurring at the monolayer or sub-monolayer level, where the interaction of a single monomer with the surface might be more pronounced. It is worth noting that Burke *et al.*[317] did not provide frequency values for sub-monolayer adsorption at the crystalline ice surface. Unfortunately, this limits our ability to directly compare our calculated frequencies with Burke's experimental observations. Although experimental indications regarding the interaction of single AA molecule with the ice surface are somewhat incomplete and contradictory, some studies mention the existence of a monomeric species interacting with the surface, characterized by a signal at 1790 cm^{-1}. However, caution is necessary when attributing this peak to surface interaction, as its frequency closely resembles that of an isolated AA molecule in the gas phase, suggesting minimal interaction with the surface. On

the other hand, it is interesting to note that the C=O stretching frequencies obtained with the most refined RD2 scheme (1741 cm^{-1}) well correlates with the value measured by Bahr *et al.* (1747 cm^{-1}) that was speculatively assigned to monomers trapped into the bulk of crystalline ice.

For GA we do not have experimental data to use as reference to ascertain the performance of the procedure. Since, GA interacts with the surface with two H-bonds we may expect a similar trend as in the case of AA. Still, GA tends to polymerize and its interaction with the surface can be difficult to disentangle.

For MF, we can accurately predict the target frequency. The improved agreement can be attributed to the simplicity of the chemistry involved in MF adsorption compared to the other isomers. In fact, MF does not strongly interact with other MF molecules. Therefore, the polymerization process, which is typical of AA and GA chemistry and increases the complexity of the recorded spectra and peak assignment, is not a concern.

5.3.2 Adsorption of acetic acid and methyl formate on low density amorphous ice

To enhance our comprehension of adsorption chemistry, we present our findings on the adsorption process of AA and MF on LDA. This choice is motivated by the fact that on the one side interstellar ice is believed to be the LDA form, on the other the majority of iCOMs adsorption at low temperature focus on amorphous ice. In these experimental investigations the amorphous nature of ice is speculated through examination of the OH stretching bandshape around 3300 cm^{-1}. However, specific measurements aimed at determining the structural pattern of ice are rarely performed, thus giving rise to a certain degree of ambiguity. For example, Burke and colleagues[317] have raised concerns regarding the actual crystallinity of the form utilized by Bertin *et al.*[320], thus hindering direct comparison among different measurements and challenging their interpretation.

As pointed out in Section **5.2**, LDA was obtained through MD simulation.

A first issue to be addressed was the selection of a reliable FF. The primary objective of testing force fields is to ensure that the simulated properties and behaviors of water ice align with experimental observations. To evaluate the reliability of our models, we compared calculated structural data with experimental results[338]. For the purpose, simulated and experimental oxygen-oxygen RDF, average number of particles and *oto* are compared in Figure **5.5**. As it can be observed, both TIP4PQ/2005[204] and TIP4P/2005f[335] yield satisfactory agreement with experimental behavior of the oxygen-oxygen radial distribution function ($g_{(O-O)}(r)$).

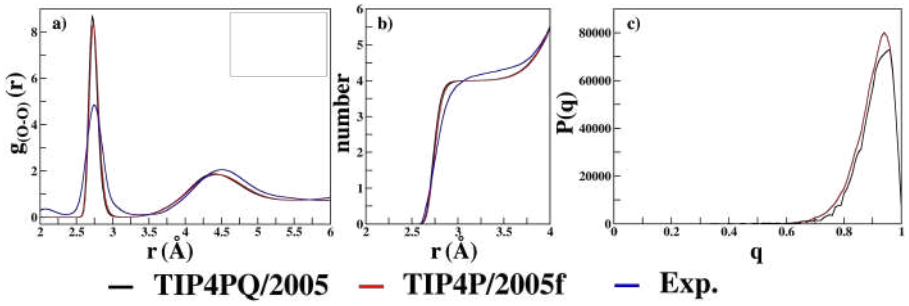

Figure 5.5 Structural analysis of LDA. Panel *a*: oxygen-oxygen radial distribution function. Panel *b*: average number of particles within a distance r. Panel *c*: orientational tetrahedral order. Experimental data from ref.[338].

Our results demonstrate very good agreement with experimental findings, as it can be appreciated from Table **5.2** in which the obtained density is also reported.

Table 5.2 First coordination shell of $g_{(O-O)}(r)$ ($r(O-O)$), density (d) and coordination number (c_n) of LDA. Results for TIP4PQ/2005 and TIP4P/2005f are compared with respect to experiment.

	$r(O-O)$ (Å)	d (g/cm^3)	c_n
TIP4PQ/2005	2.72	0.942	4
TIP4P/2005f	2.74	0.949	4
Exp.[338]	2.750 ± 0.002	0.94(35)	4

LDA has been observed to retain some correlation with the structure of ice *Ih*[339]. Consequently, the tetrahedrality order typically reaches approximately 0.9, which is consistent with our calculated results. Specifically, the *oto* values

are 0.900 and 0.895 for TIP4PQ/2005 and TIP4P/2005f force fields, respectively (panel *c* of Figure **5.5**). The tetrahedral arrangement of molecules within the first coordination shell is further supported by the value of the coordination number, c_n (for the cases studied, $c_n = 4$, indicating that a central oxygen atom is surrounded by four other oxygens). All in all, our simulations provide a successful description of the main structural properties of LDA. Furthermore, the two force fields employed yield essentially the same results and thus they both provide reliable description of LDA structure.

Starting from the obtained structural model, the conformational dynamics of AA and MF interacting with the LDA surface was also sampled by MD simulations. The QM/QM′ optimized geometries of the clusters extracted from the MD trajectories for MF and AA adsorption at the LDA interface are reported in Figure **5.6** and Figure **5.7**, respectively.

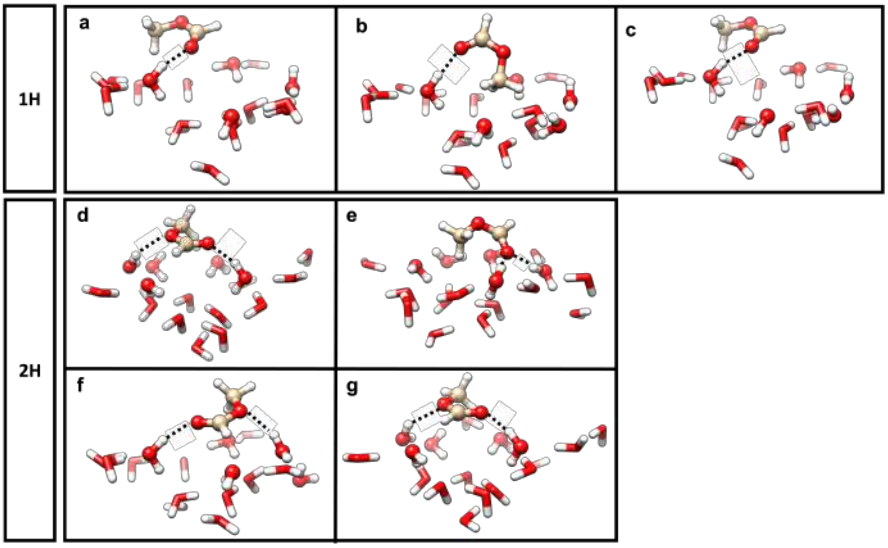

Figure 5.6 Adsorption configurations for MF@LDA obtained at DSDPBEP86/jul-cc-pVTZ:PW6B95D3/jul-cc-pVDZ level. Ball and stick for atoms at QM level while licorice representation for lower level QM′ layer. Top panel (1H in figure) for configurations featuring one H-bond between MF and the water cluster. Bottom panel (2H in figure) for two H-bonds adsorption modes. H-bonds are highlighted with black dashed lines.

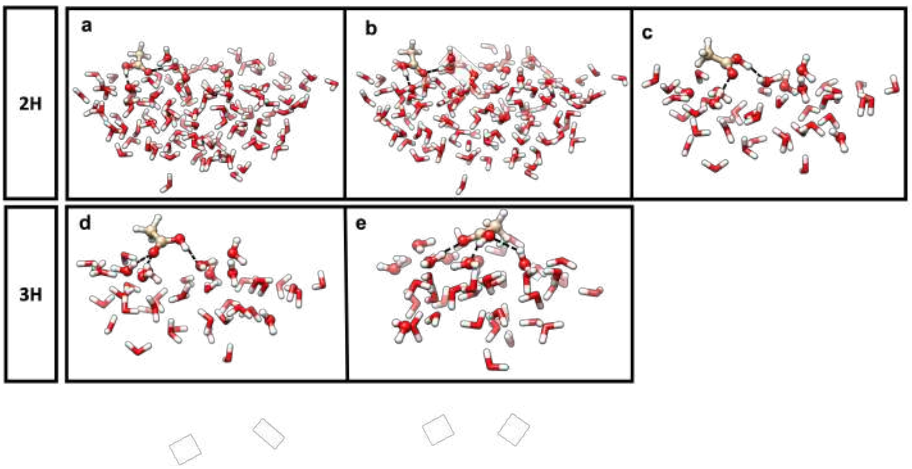

Figure 5.7 Adsorption configurations for AA@LDA obtained at DSDPBEP86/jul-cc-pVTZ:PW6B95D3/jul-cc-pVDZ level. Ball and stick for atoms at QM level while licorice representation for lower level QM′ layer. Top panel (2H in figure) for configurations featuring two H-bonds between AA and the water cluster. Bottom panel (3H in figure) for three H-bonds adsorption modes. H-bonds are highlighted with black dashed lines.

In the context of the reported configurations, the most stable arrangement involves two hydrogen bonds for MF adsorption (panel *g* in Figure **5.6**) and three for AA adsorption (panel *e* in Figure **5.7**). As expected, the adsorption on LDA exhibits slight variations compared to our observations in the CI case (cfr. Figure **5.4**). The relative stability of the two isomers aligns with the observed trend for CI, with AA showing a stronger interaction with the dangling water molecules, as expected based on its functional groups. In addition, it is worth mentioning that the configuration of CI with the adsorbate perpendicular to the surface appears to be the less energetically favorable option (see for example *b* and *e* for the case of MF). Conversely, an adsorption configuration with the molecule settling almost parallel to the plane of the surface seems to be preferred, indicating a lower energy state for this orientation because of the cooperative effects of stronger dispersion interactions in addition to the H-bonds. This structural arrangement can be attributed to the different surface morphology between CI and LDA. The more amorphous structure of LDA causes a less compact packing and a larger number of adsorption sites, which facilitate a parallel orientation of the adsorbate, capable of maximizing H-bonds and

dispersion interactions.

The IR spectra between 1000 and 2000 cm^{-1} simulated for the different representative adsorption configurations of MF and AA on LDA (see Figure **5.6** and Figure **5.7**) are presented in Figure **5.8**, while the corresponding C=O stretching frequencies are collected in Table **5.3**. Additionally, we include the spectra obtained by averaging the spectra of all the clusters (Mean/violet curves in Figure **5.8**). Each transition has been convoluted with a Gaussian function having a full width at half maximum of 22 cm^{-1} to simulate what reported in the experimental works.

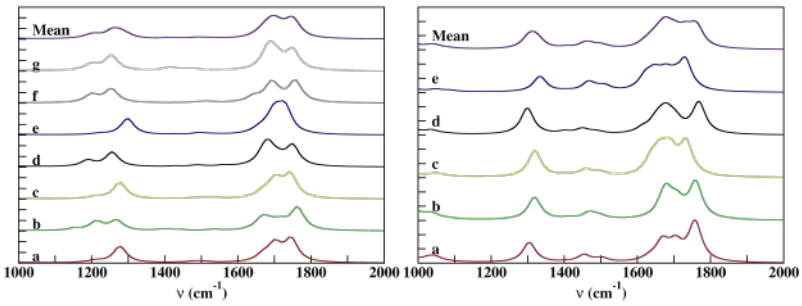

Figure 5.8 Anharmonic IR spectra in the range 1000 and 2000 cm^{-1} for different adsorption models of MF (left panel, for structures see Figure **5.6**) and AA (right panel, for structures see Figure **5.7**). The curve labelled as Mean is the average of the spectra for the different interaction models.

Our focus was specifically on the carbonyl stretching region (C=O) for both MF and AA, as it is the fingerprint adsorption for their interaction with the ice surface. Furthermore, we extended our analysis to include the stretching of the C-O single bond of the ester function in MF, as well as a set of frequencies related to the bending of the methyl group and the C-H stretching of the formyl moiety in MF (Table **5.4**). As discussed extensively in the previous section regarding adsorption on the ice XI (010) surface, interpreting the adsorption of AA on the LDA surface is challenging. The experimental spectra typically exhibit dimerization and polymerization, resulting in the formation of long chains similar to those found in solid AA ices. Experimental studies have reported multiple frequency values spanning a wide wavenumber range, influenced by the AA-water

interactions. In what follows we restrict our discussion to available experimental results for the monolayer/sub-monolayer regime that should be representative of the single AA molecule interacting with the surface. At sub-monolayer exposure, Gao *et al.* reported the C=O stretching mode for noncrystalline ice films as a weak profile consisting of two components at 1717 and 1695 cm^{-1}. The first has been assigned to the formation of dimers and the second to AA monomers interacting with dangling OH groups. The same interpretation was given in subsequent experimental studies[320,322]. Our results though could open up the way to a new interpretation. Indeed, the signal around 1717 cm^{-1} usually attributed to AA2 is very close to the results obtained for the adsorption configurations *b* and *e* (see Figure **5.7**). The peak around 1695 cm^{-1} is consistent with the adsorption configuration of the type observed in models *a*, *c* and *d*. For two isomers we also obtained a good agreement in the prediction of the C-O stretching frequency within 10 cm^{-1} from the reported experimental data.

Table 5.3 Adsorption energies (kcal/mol) and C=O stretching frequencies (cm^{-1}) obtained for the different adsorption configurations of AA and MF on LDA at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ.

	Adsorption model [a]	$\Delta(E_{ads}+ZPE)$	ω(C=O) [b]	ν^{RD1}(C=O) [c]	ν^{RD2}(C=O) [d]
AA	a	-15.13	1765	1758	1705
	b	-15.25	1762	1759	1723
	c	-15.86	1730	1731	1703
	d	-11.91	1738	1733	1686
	e	-20.30	1774	1768	1720
MF	a	-9.94	1753	1746	1711
	b	-7.76	1773	1763	1717
	c	-10.71	1753	1744	1706
	d	-10.26	1759	1750	1713
	e	-8.17	1734	1730	1686
	f	-9.98	1767	1757	1742
	g	-10.99	1759	1750	1711

[a] Adsorption models for AA and MF labelled as in Figure **5.7** and in Figure **5.6** respectively.
[b] Harmonic frequency at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.
[c] Anharmonic frequency according to RD1 scheme at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.
[d] Anharmonic frequency according to RD2 scheme at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.

It is important to note that a more comprehensive sampling of the phase space, particularly for the CI surface, could unveil other configurations responsible for

additional peaks. Likewise, further calculations incorporating proper sampling of AA2 adsorption could provide a comprehensive theoretical interpretation of the phenomenon. The comparison between the IR spectra here simulated with the experimental ones recorded by Bertin *et al.*[320] is presented in Figure **5.9** where the theoretical spectrum has been obtained by averaging the bandshape of the different adsorption models calculated within the RD2 scheme.

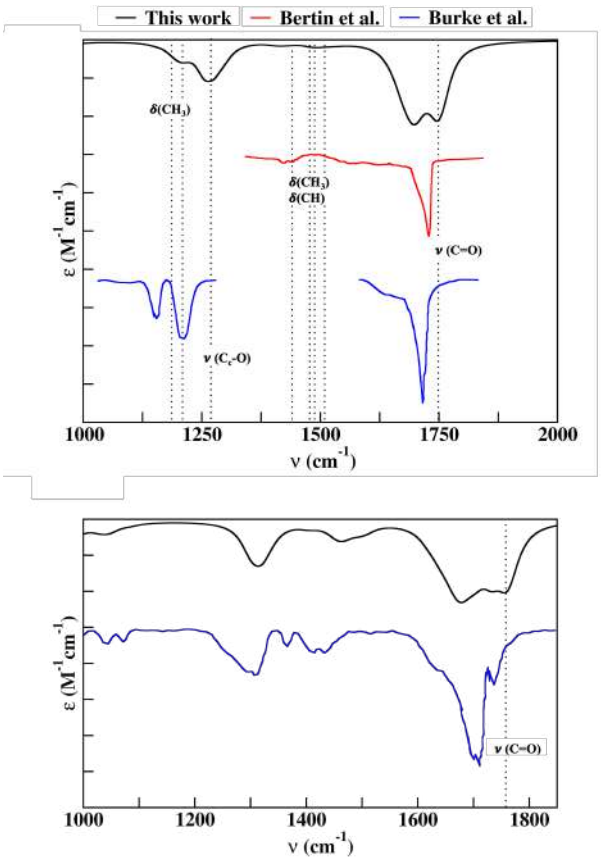

Figure 5.9 Comparison between calculated and experimental IR spectra in the region 1000 —2000 cm^{-1} for MF (top panel) and 1000 —1850 cm^{-1} for AA (bottom panel). The calculated spectra (black lines) corresponds to the violet spectra in Figure **5.8**. The C=O, C–O stretching ($\nu(C = O)$ and $\nu(C_C - O)$) and the CH_3 bending ($\delta(CH_3)$) regions are highlighted in the picture. Top panel for MF and bottom panel for AA. Red line refers to the spectrum at 80 K as reported in Figure 4 of ref.[320]. Burke *et al.* (blue line) did not provide any data between 1350 and 1600 cm^{-1}[317]; the spectra reported are taken from Figure 2 (for MF) and Figure 5 (for AA) of the referenced paper.

Table 5.4 Vibrational frequencies (cm^{-1}) of MF adsorbed on LDA between 1000 and 2000 cm^{-1} obtained at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ and comparison with available experimental values.

	ω^a	ν^{RD1b}	ν^{RD2c}	Exp.[320]	Exp.[317]
ν(C=O)	1757	1749	1712	1733	1737
ν(C-O)	1270	1287	1169	1225	1233
δ(CH)	1418	1441	1401	1383	-
$\delta(CH_3)$	1190	1087	1173	-	1172
$\delta(CH_3)$	1207	1209	1103	-	-
$\delta_s(CH_3)$	1485	1488	1341	1435	-
$\delta_{as}(CH_3)$	1510	1509	1425	1450	-
$\delta_{as}(CH_3)$	1526	1488	1209		-

[a] Harmonic frequency at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.
[b] Anharmonic frequency according to RD1 scheme at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.
[c] Anharmonic frequency according to RD2 scheme at DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ as described in Section **5.2**.

The carbonyl stretching frequency of MF adsorbed on LDA spans the frequency range 1686 —1742 cm^{-1} depending on the adsorption model with the most strongly bound configurations featuring the C=O stretching bond located around 1712 cm^{-1}. This value is in good agreement with experimental observations which reported a value of 1735 ± 2 cm^{-1}.

Bertin *et al.* also provided values for the stretching of the C-H group (δ(CH)) which still results to be overestimated of 56 cm $^{-1}$ in the case of the mean value resulting from all the studied configurations. On the contrary, the mean value predicted for δ(CO) is within less than 20 cm^{-1} from the experimental value of 1383 cm^{-1}. In the case of the frequencies involving the methyl group, the agreement between $\nu^{RD2}(\delta(CH_3))$ and the experimental value is remarkable. Bertin *et al.* reported values for the symmetric and asymmetric bending of the methyl group which however are quite far from our calculated values. Inspection of the values collected in Table **5.4** reveals that the cause of the mismatch is due to the large and nonphysical anharmonic corrections affecting these vibrations. While the causes for this behaviour should be investigated in more detail, possible

sources of inaccuracy can be seeked in a RD treatment involving a too small set of active modes or in the presence of LAMs.

5.4 Conclusions and astrochemical implication

In this chapter, we investigated the adsorption of acetic acid (monomer AA and dimer AA2), methyl formate (MF), and glycolaldehyde (GA) on the (010) surface of crystalline ice XI. Additionally, we extended our investigation to include studies on low-density amorphous ice for both acetic acid and methyl formate. Our research aimed to provide new insights into the interaction and segregation of these molecules on the ice surface and their potential impact on their molecular abundances in the gas phase.

Vibrational spectroscopy has proven to be a powerful tool in astrochemistry, allowing the investigation of the chemical composition of interstellar ices. Computational vibrational spectroscopy, on the other side, offers valuable atomistic into molecular vibrations and interaction with surfaces. Understanding the adsorption processes of molecules on solid surfaces is crucial for elucidating the chemical pathways that lead to the formation of complex organic molecules, which are of great interest in astrochemical environments.

By investigating the adsorption of AA, MF, and GA on crystalline surfaces, we gained important insights into the preferred adsorption configurations and the stability of different isomers. The results obtained shed light on the factors influencing the strength of interactions between the molecules and the surface. Furthermore, our analysis on the acetic acid dimer provided additional understanding of the behavior of this molecule and the role of monomers and dimers.

Extension of our investigations to amorphous ice surfaces allowed us to examine the adsorption behavior of AA and MF in a more realistic and complex environment. Our findings provided a new interpretation of the molecules interaction and segregation on the ice surface.

While our results are promising and contribute to the understanding of adsorption processes on ice surfaces, further analysis is still required. The complexity of these systems necessitates more comprehensive studies to fully capture the intricacies of molecular adsorption and the underlying surface dynamics. Additionally, considering other factors such as temperature and coverage in future investigations would provide a more complete picture of the adsorption processes.

In conclusion, our research on the adsorption of acetic acid, methyl formate, and glycolaldehyde on crystalline and amorphous ice surfaces, along with our examination of acetic acid dimer formation, has provided a microscopic description of the behavior of these molecules in interaction with interstellar ices. Our results lay the foundation for further investigations in this field and contribute to the broader understanding of astrochemical processes.

Overall, these findings demonstrate the effectiveness of our procedure and highlight the potential influence of specific adsorption configurations on the observed IR spectra. While more detailed investigations and sampling are necessary, our results provide valuable insights into the underlying mechanisms of adsorption on both CI and LDA ice surfaces.

Chapter 6

Exploring ice growth on top of amorphous silica

6.1 Introduction

In the previous chapter, we presented a combined MD/QM procedure that effectively simulates the LDA structure of interstellar ices. Specifically, we employed DFT methods to establish a robust ONIOM approach for predicting characteristic IR fingerprints of iCOMs at the interstellar ice interface. However, it is important to acknowledge that icy mantles in the interstellar medium often freeze onto mineral dust grains, and the precise composition of these grains remains uncertain.

Previous studies have shed some light on the mineral composition and main IR fingerprints of these grains[51], but as our observations and sample collection expand, we discover a broader range of possibilities in terms of size, morphology, and composition[340–342]. Consequently, developing a simplistic molecular design for such systems becomes challenging. The analysis of comet 67P/Churyumov–Gerasimenko has demonstrated that its core consists of a mixture of ices, iron sulfides, silicates, and hydrocarbons, suggesting that ice and

dust are physically mixed within comet nuclei rather than organized in layered structures[343,344]. Additionally, it is expected that asteroids retain ices in their interiors.

Experiments, models, and available samples further indicate that dust grains in astrophysical environments form aggregates with high porosity, and it is reasonable to assume that ice fills the pores of these grain aggregates. Consequently, in laboratory experiments, ice mixed with dust is a more appropriate model than the traditional assumption of ice layered on dust. Spectroscopic analysis of astronomical observations provides column density measurements for ice abundances, which are highly variable for interstellar ices, ranging from $0.3 \cdot 10^{18}$ to $12 \cdot 10^{18}$ cm^{-2} [56,345,346]. Recent JWST observations have provided new evidence of the highly variable molecular composition of ice, posing significant challenges for both experimental and theoretical studies[347].

Despite the considerable interest in layered and mixed ices, as well as the investigation of pure and multi-component ices and their photo-processing-induced reactivity, a key question arises: does the composition of the underlying mineral core influence the observed ice morphology?

While various models have been developed to understand the interactions between cosmic grains, they often rely on simplifications of chemical compositions, morphology, and oversimplified interfaces. Describing the structure-property relationships in such complex systems is a daunting task. Therefore, in this study, we propose employing classical MD simulations to investigate the impact of an amorphous silica (aSIL) core on the deposition and growth of water ice. This approach allows us to develop a more realistic model of the interstellar ice grain surface, which is essential for understanding if the structural properties of the icy mantle are affected by the underneath silica core, that in turn could influence the reactivity of the ice surface. To gain insights into the morphology of interstellar ice at different coverage, we conducted simulations of systems with an increased number of water molecules, spanning the sub-monolayer and monolayer regimes to investigate the structural properties and infrared fingerprints that develop at

the subsequent stages of ice accretion.

6.2 Computational Methods

For the simulation of the mineral core we used an amorphous silica model available in the INTERFACE force field (IFF) library[207,208]. Since the ISM is highly rich in hydrogen we selected a fully hydrogenated surface with no charge. A fundamental aspect that needs to be considered is the surface density of OH groups (α_{OH}) which is dependent on the specific synthesis conditions of the mineral. Zhuravlev *et al.* studied the dependence of the density of sylanol groups on the silica surface on the pretreatment temperature of the SiO_2 samples[348]. For a wide range of synthetic strategies, they found an average value of α_{OH} equals to 4.6 OH nm^{-2}. For space formed minerals, to the best of our knowledge, no information is available hence we selected that with an hydrogenation degree closer to the average value found by Zhuravlev in his work. The IFF comes with twenty different surface models that span the full range of surface chemistry and pH. We selected the available model for amorphous silica with no charge and $\alpha_{OH} = 4.6$ nm^{-2}. To model the ice we selected the TIP4P/2005f force field[335] as from our previous testing performs well in the prediction of LDA structural features and includes the flexible description required for the proper simulation of the IR spectrum. We equilibrated the system in the NPT ensemble at 77 K with the same procedure described in the previous chapter for the LDA model. We progressively increased the number of water molecules from 100 up to 500 in order to reach a full monolayer regime. For each system the total coverage of the two surfaces was evaluated as:

$$\theta = \frac{n_{OH\cdots H}}{N_{OH}} \cdot 100, \tag{6.1}$$

with $n_{OH\cdots H}$ number of superficial OH groups involved in H bonds with the water molecules and N_{OH} total number of OH groups. To analyze the vibrational

dynamics of a molecular system using classical molecular dynamics simulations, we employed a procedure to calculate the VACFs and obtain the corresponding vibrational power spectrum. This methodology allowed us to gain insights into the system's vibrational modes and their frequencies.

After the simulation, we extracted the velocities of the atoms or particles of interest from the trajectory. The VACFs were calculated as the time correlation of the velocities for each atom or particle, using the following equation:

$$C(t) = \langle v(t_0) \cdot v(t_0 + t) \rangle, \tag{6.2}$$

where $C(t)$ is the VACF at time lag t, $v(t_0)$ is the velocity of the atom or particle at time t_0, and the angular brackets denote a time average.

To obtain the vibrational power spectrum, we performed a Fourier transform of the VACFs. This transformation provided information about the frequencies present in the system's vibrational modes. The power spectrum $P(\omega)$ is given by:

$$P(\omega) = | \int C(t) \exp^{-i\omega t} \, dt|^2, \tag{6.3}$$

where $P(\omega)$ represents the power at frequency ω, $C(t)$ is the VACF, and, the integral is performed over the appropriate time range.

Leap-frog integration with a time step of 1 fs was employed for a total simulation time of 2 ns. We collected atom velocities and coordinates every 6 fs. Finally, the IR spectra were calculated by means of the *gmx velacc* command. By following this methodology, we were able to extract information about the vibrational dynamics of the molecular system under study.

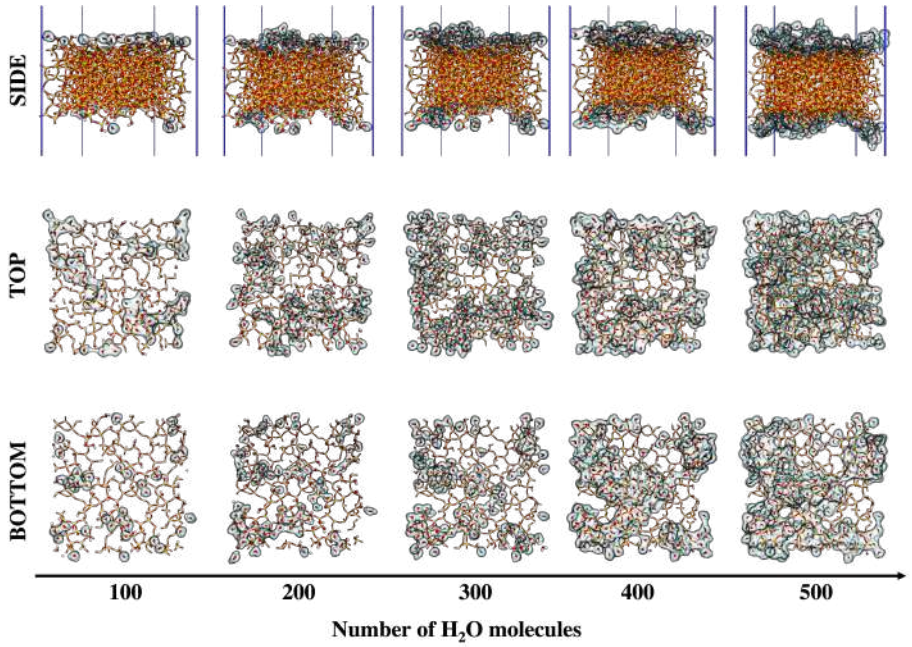

Figure 6.1 Equilibrated systems for water ice at the amorphous silica interface, with the number of H_2O molecules ranging from 100 (left) to 500 (right). Side, Top and Bottom views are reported for each system.

6.3 Results discussion

6.3.1 Structural analysis

A pictorial and numeric estimate of the surface coverage range explored is reported in Figures **6.1** and **6.2**.

The coverage percentage explored ranges from 78% to 96% for the system with 100 and 500 water molecules respectively. This assures that all the analysis we are going to perform will be applicable to the sub-monolayer and monolayer regime. Upon closer examination of the figures, it becomes evident that clusters and islands of water molecules form on the surface. These clusters and islands subsequently coalesce, leading to the formation of a thick monolayer. This observation aligns with previous studies, providing support for the proposed findings. Potapov *et al.* proposed that during the prestellar, protostellar, and protoplanetary phases, ice layers on grains may exist as sub-monolayers or a

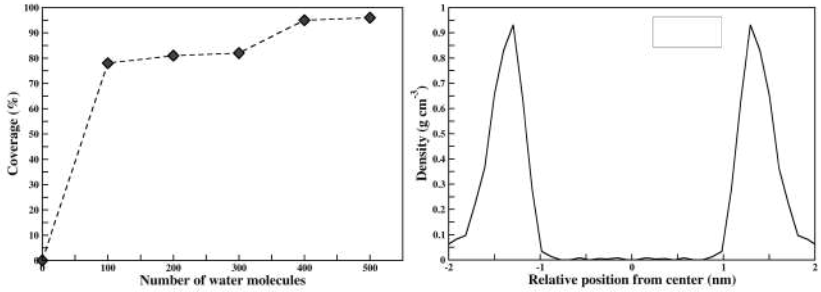

Figure 6.2 Surface coverage as function of the number of water molecules in the equilibrated systems (left panel). Density profile of ice at the silica interface at full monolayer regime as for the system with 500 water molecules (right panel). The zero is set at the center of the mineral slab and the two peaks corresponds to the ice layers growing on the two exposed mineral surfaces (see TOP and BOTTOM in Figure **6.1**).

few monolayers and the presence of clusters, islands, and voids is expected[341]. A low coverage of water molecules has implications for the reactivity of the surface sites that are accessible and available to actively participate in chemical reactions. This increased accessibility of surface sites enhances the potential for reactivity and opens up opportunities for various surface-mediated processes[349]. In the right panel of Figure **6.2** we report the density across the simulation box calculated with respect to its distance from the aSIL slab center which density is set to zero for reference. We examined the 500 water molecules system since the full monolayer regime can be considered achieved. The peak is found at 0.94 g/cm^3 which is close to the expected density of LDA.

To further characterize the water-silica systems, we calculated RDFs. The RDF provides valuable information about the spatial distribution of water molecules with respect to the silica surface. In panel *a* of Figure **6.3** the oxygen-oxygen RDF for water molecules ($r_{(OW-OW)}$) provides structural information about the growing ice layer and can be directly compared to LDA data from our previous calculated profiles as well as available experimental data collected in the previous chapter. The RDFs obtained for all the systems under study exhibited similar profiles. The first peak in the RDF, corresponding to the nearest neighbor water-water interactions, was observed at a distance of 2.57 Å,

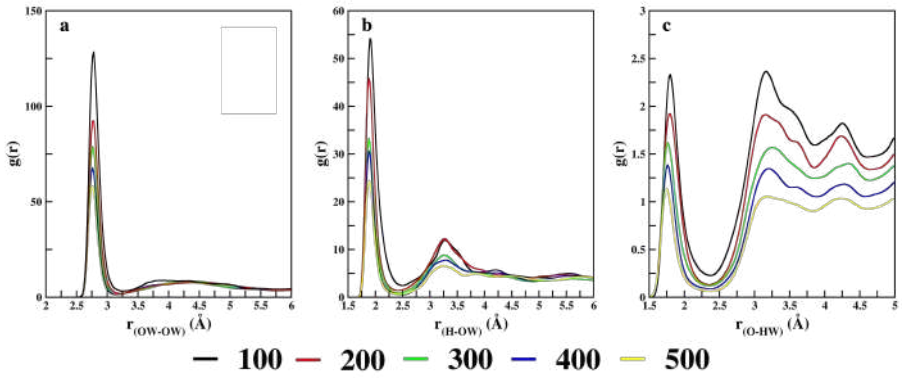

Figure 6.3 RDFs profiles for $(H_2O)_n$ with n = 100, 200, 300, 400 and 500. Panel *a* is the oxygen-oxygen RDF for water molecules. Panel *b* is the hydrogen-oxygen RDF where the hydrogen and the oxygen belong to the superficial hydroxil group of the silica and the water, respectively. Panel *c* is the oxygen-hydrogen RDF where the oxygen and the hydrogen belong to the hydroxil group of the silica and the water, respectively.

consistent with the data reported for the LDA model[338]. However, due to the monolayer coverage level, the visibility of the second shell peak was limited, as the number of water molecules was not sufficient to form distinct density shells beyond the first peak. In order to characterize the water-silica interface, panels *b* and *c* in Figure **6.3** collect RDFs for the interactions involving H and O of superficial OH groups and the O and H of the water molecule, H-OW and O-HW respectively. They provide information about the hydroxil groups acting as donor and acceptor in the formation of the hydrogen bonds with the surrounding water molecules. What emerges is that the peaks at the nearest-neighbour distance are sharp and well-defined suggesting the formation of a structured interface. These findings support the reliability of our computational approach in capturing the key features of water adsorption on the silica surface with the formation of an ice layer with structural similarities to LDA.

Overall, the RDF analysis demonstrated the presence of well-defined water-silica interactions at the nearest neighbor distance, while the density distribution confirmed the formation of a compact water monolayer on the silica surface. The obtained results provide an atomistic picture of the interfacial behavior of water

molecules on amorphous silica at low temperature and provide insights into the structural properties of the system at the monolayer level.

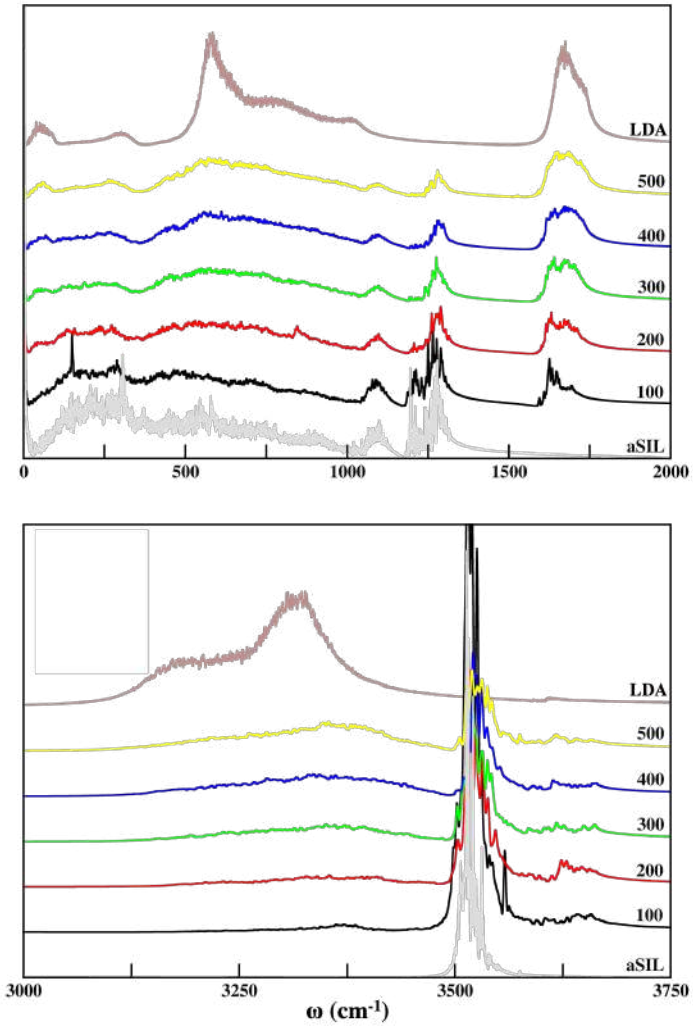

Figure 6.4 Low frequency (top) and high frequency (bottom) power vibrational spectra for the studied systems with $(H_2O)_n$@aSIL with n = 100, 200, 300, 400, 500. Spectra of the mineral core (aSIL) and of the LDA as obtained in the previous chapter (Chapter **5**) are also reported for comparison.

6.3.2 Infrared spectroscopy

The IR spectra for the studied systems are reported in Figure **6.4**. Let us consider first the IR spectra obtained for the LDA and aSIL models. In the case of LDA the attention has to be focused on three regions: below 1000 cm^{-1},

around 1600 cm^{-1} and above 3000 cm^{-1}. The observed bandshapes suggest an amorphous structure as expected on the basis of the structural analysis. The low frequency range is quite intricate and the bandshapes here reported are very similar to those obtained in the previous study by Belosludov *et al.*[350]. To the best of our knowledge, not many studies have focused on this spectral region, and a proper assignment is beyond the scope of the present work. It seems reasonable to assume that collective low-frequency vibrational modes, involving more water molecules, are responsible of these signatures[350]. The H-O-H bending is correctly predicted around 1600 cm^{-1} and this band is observed to show two peaks at low temperature in the case of amorphous structures which may explain the shoulder of the main peak appearing in our calculated spectra. The feature around 3300 cm^{-1} is due to the OH stretchings and the broad bandshape is consistent with an amorphous structure. Even if barely discernible, there is also a small peak around 3610 cm^{-1} which is due to the OH stretching of superficial dangling groups. For the aSIL model the Si-O-Si stretching and bending vibrations are correctly found around 1000 cm^{-1} and 555 cm^{-1}. Bending and stretching vibrations of the hydroxil groups are responsible for the peaks centered at about 1290 cm^{-1} and 3500 cm^{-1} respectively. By inspection of the spectra obtained by increasing the H_2O leading from 100 to 500 water molecules, the progressive appearances of the main signatures of the ice profile can be noticed as the coverage increases. In the low frequency range the structure is quite intricate since ice and silicate fingerprints overlap. However, as the coverage increases the water ice absorption below 500 cm^{-1} looks more structured resembling those reported for the bare ice. The same phenomenon is observed for the peak around 1600 cm^{-1} ascribed to the OH bending. At higher frequencies (bottom panel in Figure **6.4**) the spectra is dominated by the signal of the silica hydroxyl group. As the coverage increases the peak starts to be saturated and the lower frequency broad signals of the OH stretching starts to emerge. Higher frequency peaks can be ascribed to the OH stretching of the superficial water molecules.

6.4 Conclusions

This study has presented a procedure for simulating the LDA structure of interstellar ices at the ice-mineral interface. In the previous chapter, by employing DFT methods and developing a robust ONIOM approach, we successfully predicted characteristic IR fingerprints of iCOMs on interstellar ices. However, it is crucial to acknowledge that the composition and morphology of dust grains in the interstellar medium are still not fully understood, posing challenges in developing simplistic molecular designs for these systems.

Our investigation of the impact of amorphous silica (aSIL) on the deposition and growth of ice using classical MD simulations considered various coverage levels, spanning the sub-monolayer and monolayer regimes, and allowed us to explore the structural properties and infrared fingerprints of water ice.

The results showed a range of surface coverages from 78% to 96% for the studied systems, ensuring the applicability of the analysis to both sub-monolayer and monolayer regimes. Density profiles confirmed the formation of a compact water monolayer on the silica surface, with a density close to the expected value for LDA. The RDF analysis demonstrated well-defined water-silica interactions at the nearest neighbor distance.

Furthermore, the obtained IR spectra of the studied systems revealed the progressive appearance of ice signatures as the coverage increased. The low-frequency range showed structured absorption features resembling those reported for ice, while the high-frequency range was dominated by the signal of the silica hydroxyl group and superficial O-H stretching features.

Overall, this study contributes to our understanding of interstellar ice formation and provides valuable insights into the structural properties and vibrational dynamics of interstellar ice at the ice-mineral interface.

Future research should continue to explore the effects of different mineral compositions, morphologies, and interfaces on the observed ice morphology. Additionally, combining experimental studies with theoretical simulations at

various levels can further enhance our understanding of interstellar ices and their reactivity in the interstellar medium. By advancing our knowledge in these areas, we can deepen our understanding of the fundamental processes underlying the formation and evolution of complex organic molecules in space.

Chapter 7

Conclusions and future perspective

This thesis has been focused on the development of a computational strategy for the accurate simulation of the chemical processes mediated by interstellar ices. The severe conditions of the ISM pose many difficulties to the set up of experiments and well-designed computational strategies are key to disentangle the complex chemistry emerging from both astrochemical observations and experiments performed on interstellar ice analogues. In this work, we have proved that computational chemistry is a powerful tool in the study of interstellar ice and the chemistry evolving at the icy interface.

The main conclusions of each part of this thesis are summarized in the following sections.

7.1 Gas-phase reactivity in the ISM

Gas-phase reactions have long been regarded as the primary pathway for chemical evolution in the ISM. In Chapter **3**, we revisited the gas-phase formation of ethanimine, shedding light on the accuracy of state-of-the-art methods em-

ployed in the field of astrochemistry. The significance of ethanimine arises from its potential prebiotic role in the chemical pathways leading to the formation of life-supporting molecules.

Our investigation focused on the reaction between the amidogen radical ($NH^{\bullet}$) and the ethyl radical ($C_2H_5{}^{\bullet}$), employing state-of-the-art composite schemes followed by kinetics calculations. Our results demonstrated that this particular reaction proceeds at an exceptionally high rate. Although, under typical interstellar medium conditions, the dominant reaction channel leads to the formation of methanimine and the methyl radical, our calculations predict the production of the two E,Z stereoisomers of ethanimine at a level of approximately 10 %. Notably, the predicted [E−CH_3CHNH]/[Z−CH_3CHNH] ratio is around 1.4, which falls short of the value determined from astronomical observations, approximately 3.

These findings highlight the complex nature of the formation of ethanimine in the gas phase and emphasize the limitations of our current understanding. While our calculations capture a portion of the observed ethanimine abundance, the disparity between the predicted and observed [E−CH_3CHNH]/[Z−CH_3CHNH] ratios calls for further investigation. Our work exemplifies the continuous need for refining and improving computational protocols for astrochemical applications. Furthermore, it becomes increasingly apparent that interdisciplinary efforts, combining experimental data, theoretical modeling, and astronomical observations, are necessary to unravel the complexities of astrochemical processes accurately.

7.2 Multiscale strategies for reactivity on interstellar ice surfaces

The accuracy of extrapolation schemes, such as ChS and jun-ChS, in predicting both energy and structures has gained widespread recognition. However, the high computational demand associated with these methods limits their applicabil-

ity to small to medium molecular systems, excluding chemical processes involving solid surfaces. In Chapter **4**, we addressed this challenge by developing a multi-layer ONIOM strategy to be implemented for the accurate simulation of medium to large size systems. We benchmarked ten of the most common density functionals in conjuction with six different basis sets of increasing size and selected the best performing DFT methods. We studied the HCN/HNC isomerization on ice clusters. Among the studied functionals, the optimal trade-off between accuracy and computational cost is offered by the PW6B95-D3/jul-cc-pVDZ and the double-hybrids, DSD-PBEP86-D3 and revDSD-PBEP86-D3 in conjunction with the jul-cc-pVTZ. These methods have been used in an ONIOM framework to investigate the effect of cluster size and ice surface on the isomerization process of HCN. To properly account for the surface environment we employed a QM/QM′ strategy for the 20 water molecules cluster extracted from the (010) surface of ice XI. The superficial arrangement of the water molecules favours a proton relay mechanism over the direct hydrogen transfer. By extending the size of the cluster up to 192 water molecules in the framework of a three-layer QM/QM′/MM model the energy barrier ruling the isomerization is not affected significantly showing that the structural and catalytic properties of the crystalline ice surface are correctly captured by the $(H_2O)_{20}$ cluster. Computation of reaction rates by transition state theory indicates that on icy surfaces the isomerization of HNC to HCN could occur quite easily even at low temperatures thanks to tunneling effects. This strategy has been proposed and then implemented for the modelling of iCOMs-ices interactions ruling reactivity and spectroscopic properties.

7.3 Anharmonic fingerprints of iCOMs at the icy interface

Crystalline water ice surfaces are commonly utilized to simulate icy dust grains; however, under the conditions found in the ISM, the observed infrared

features more closely resemble those of amorphous structures. In Chapter **5**, we employed a combined MD/QM strategy to obtain a more realistic model for the interstellar ice structure and simulating the vibrational fingerprint of iCOMs adsorbed on its surface.

Our investigation focused on examining the effect of ice morphology on the adsorption of iCOMs. Specifically, we studied the adsorption of three isomers, acetic acid, methyl formate, and glycolaldehyde, on both crystalline and amorphous ice surfaces. To model crystalline ice, we employed a combined periodic/cluster approach, which involved simulating the adsorption on the (010) surface of ice XI within periodic boundary conditions. We subsequently extracted clusters from these simulations to perform QM calculations. In contrast, for an accurate representation of the adsorption on amorphous low density amorphous ice, we employed MD simulations and selected representative clusters from the trajectory obtained at a temperature of 77 K. These clusters describing the adsorption on either CI or LDA were then utilized to simulate the primary IR features of the adsorbed iCOMs.

For the spectroscopic calculations, we employed the previously developed QM/QM′ strategy with the DSD-PBEP86/jul-cc-pVTZ:PW6B95-D3/jul-cc-pVDZ model chemistry. This approach enabled us to capture the intricacies of the systems under investigation. Additionally, we employed VPT2 to accounting for anharmonicity. Remarkably, the VPT2 calculations yielded excellent results, predicting the carbonyl features within a tenth of a cm^{-1}, thus providing the accuracy required for interpreting experimental or observational signals. By employing our MD/QM approach and utilizing accurate computational methods, we successfully investigated the influence of ice morphology on the adsorption of iCOMs. Our findings shed light on the importance of considering amorphous ice structures when interpreting observed IR features in the ISM. Moreover, the developed methodologies and computational techniques provide valuable insights for future studies aiming to understand the interaction of iCOMs with icy surfaces in astrophysical environments.

Our results not only agree with existing experimental data but also provide new information and additional details that complement the current understanding of these systems. However, it is important to acknowledge that this discussion is not exhaustive, and further calculations and investigations are required for a complete comprehension of the adsorption process and the diverse interactions that can occur. The complexity of these systems, combined with the dependence of experimental results on various operational conditions, necessitates continued research efforts to refine our understanding of the chemistry involved.

Overall, our computational approach successfully reproduces IR fingerprints, provides insights into the sub-monolayer characteristics of adsorption, and offers complementary information to enhance the current knowledge. By bridging computational predictions with experimental data, this study contributes to advancing our understanding of the interaction of AA, MF and GA with interstellar ice analogues, laying the groundwork for further exploration and comprehension of the relevant astrochemical processes ruling the synthesis and abundances of these iCOMs.

7.4 Effect of mineral grain on the ice morphology

In the ISM, ices accumulate on the surfaces of dust and grain, playing a crucial role in various astrophysical processes. In Chapter **6**, we employed MD simulations to investigate the accretion of ice on amorphous silica surfaces, which serve as mineral models for dust and grain cores. Our study focused on exploring the structural and spectroscopic characteristics of the growing ice layer at different levels of coverage.

By analyzing the density profile across the interface and the radial distribution functions, we observed the formation of a low-density amorphous ice structure dominated by hydrogen bonds between the hydroxyl groups present on the silica surface and the water molecules in the ice layer. This interaction plays a fundamental role in determining the overall structure and stability of the ice.

Furthermore, we examined the vibrational power spectra obtained from the velocity autocorrelation functions, which revealed the emergence of prominent ice adsorption features characterized by broad bandshapes consistent with the amorphous nature of LDA ice.

Through our MD simulations, we gained valuable insights into the accretion process and the resulting structural and spectroscopic properties of the ice layer on amorphous silica surfaces. These findings contribute to a deeper understanding of the interplay between ices and dust/grain cores in the ISM and provide a foundation for future studies examining the role of surface interactions in astrophysical environments.

7.5 Future perspectives

In conclusion, our research underscores the critical role of well-designed computational methods in advancing our comprehension of interstellar ice chemistry. Furthermore, it is evident that in the field of astrochemistry, the interplay between various disciplines and backgrounds is pivotal for meaningful progress. While our work has laid the groundwork for studying ice morphology and the chemistry of interstellar complex organic molecules at the ice interface, further investigations are necessary to achieve a comprehensive understanding.

It is essential to acknowledge that ices in the interstellar medium are not homogeneous in composition. Therefore, the inclusion of contaminants is imperative to explore how the water matrix is influenced by these substances. Additionally, considering the constant photo-processing occurring in space and its impact on the chemical nature of the system is crucial. The incorporation of mixed and layered ices into our studies is essential for obtaining a holistic view.

Alongside the development of robust and reliable computational models capable of capturing the essential physical features, obtaining high-quality experimental data is of utmost importance in deepening our understanding. The integration of experimental results with computational models will provide a more

comprehensive understanding of the fascinating and yet unexplored chemical world in the coldest regions of the Universe.

In summary, our research has laid the foundations for bridging the gap between our current knowledge and an understanding of the chemical processes promoted by interstellar ices. By employing a multidisciplinary approach and emphasizing the significance of computational methods, future studies can build upon our work to unravel the complexities of interstellar ice chemistry and its implications for astrochemical processes.

Appendix A

Appendix for Chapter 3

A.1 List of supporting material

1. Figure **A.1**. Structures and labels of the species in the formation pathway of ethanimine at B2PLYP-D3(BJ)/aug-cc-pVTZ level.

2. Table **A.1**. Structural parameters of the stationary points of the NH + C_2H_5 reaction.

3. Table **A.2**. Absolute energies in Hartree.

4. Table **A.3**. $\mathcal{T}_1$ diagnostic.

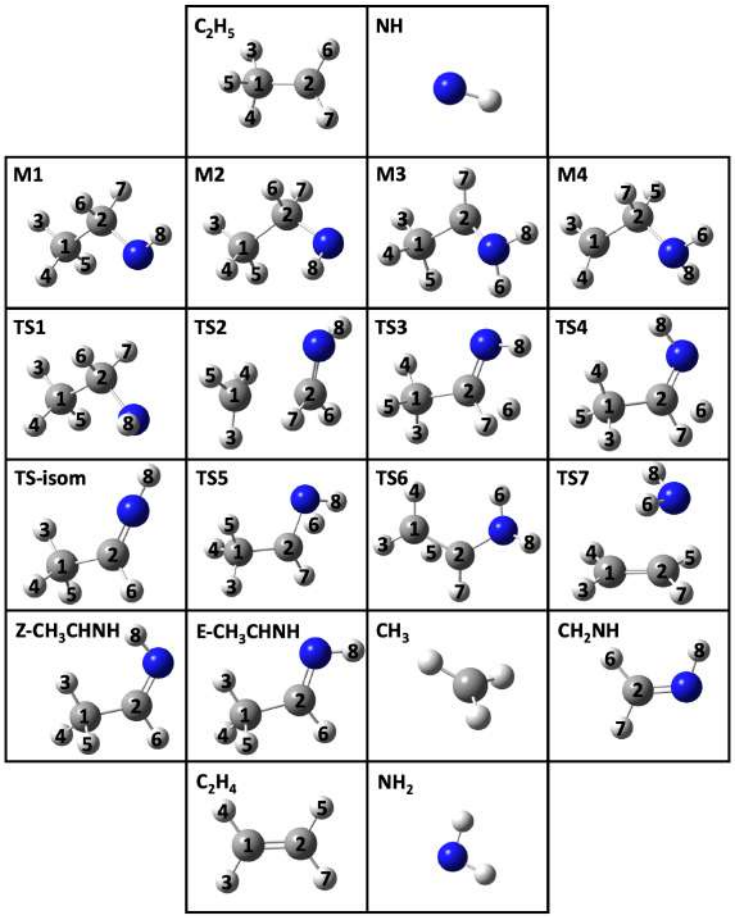

Figure A.1 Structures of the species appearing in the formation pathway of ethanimine as obtained at B2PLYP-D3(BJ)/aug-cc-pVTZ level of theory. Numeric labels are used for structural analysis in Table **A.1**

Table A.1 Structural parameters of the stationary points of the NH + C_2H_5 reaction at different levels of theory. Distances are in Å, angles in °.

		B3LYP-D3	B2PLYP-D3(BJ)	
		aug-cc-pVTZ	aug-cc-pVTZ	aug-cc-pVQZ
NH	r(NH)	1.040	1.036	1.035

Table A.1 continued from previous page

C_2H_5	r(C2H6)	1.080	1.078	1.078
	r(C2H7)	1.080	1.080	1.080
	r(C1C2)	1.484	1.484	1.483
	φ(H6C2C1)	120.9	120.9	120.9
	φ(H3C1C2)	111.9	111.7	111.7
	φ(H3C1H5)	106.3	106.5	106.5
	φ(H5C1H4)	106.3	106.5	106.5
	φ(H6C2H7)	117.7	117.7	117.7
	ϑ(H3C1C2H6)	33.0	33.5	33.4
	ϑ(H5C1C2H7	86.1	85.7	85.8
M1	r(NH8)	1.025	1.023	1.022
	r(C2N)	1.439	1.440	1.439
	r(C1C2)	1.522	1.519	1.518
	r(C1H3)	1.091	1.089	1.089
	r(C1H4)	1.090	1.089	1.088
	r(C1H5)	1.090	1.089	1.088
	r(C2H6)	1.102	1.099	1.099
	r(C2H7)	1.102	1.099	1.099
	φ(C2NH8)	107.6	107.3	107.3
	ϑ(C1C2NH8)	180.0	-180.0	-180.0
	ϑ(H7C2NH8)	57.0	57.2	57.2
	ϑ(H6C2NH8)	-57.0	-57.2	-57.2
M2	r(NH8)	1.026	1.024	1.023
	r(C2N)	1.440	1.441	1.440
	r(C1C2)	1.534	1.530	1.529
	r(C1H3)	1.091	1.089	1.088
	r(C1H4)	1.092	1.090	1.089
	r(C1H5)	1.090	1.089	1.088

Table A.1 continued from previous page

	r(C2H6)	1.103	1.100	1.099
	r(C2H7)	1.092	1.090	1.089
	φ(C2NH8)	107.3	106.8	106.9
	ϑ(C1C2NH8)	-46.3	-49.0	-49.1
	ϑ(H7C2NH8)	-170.4	-172.6	-172.7
	ϑ(H6C2NH8)	74.9	71.9	71.8
M3	r(NH8)	1.008	1.008	1.007
	r(NH6)	1.010	1.009	1.008
	r(C2N)	1.398	1.398	1.396
	r(C1C2)	1.487	1.487	1.486
	r(C1H3)	1.089	1.088	1.087
	r(C1H4)	1.101	1.098	1.098
	r(C1H5)	1.096	1.093	1.093
	r(C2H7)	1.082	1.080	1.080
	φ(C2NH6)	115.2	114.7	114.8
	φ(C2NH8)	115.5	115.1	115.3
	ϑ(C1C2NH6)	40.3	41.7	41.4
	ϑ(C1C2NH8)	172.1	172.3	172.4
	ϑ(H7C2NH6)	-170	-170.2	-170.3
	ϑ(H7C2NH8)	-38.1	-39.5	-39.2
M4	r(NH8)	1.013	1.012	1.011
	r(NH6)	1.012	1.011	1.010
	r(C2N)	1.467	1.466	1.464
	r(C1C2)	1.482	1.482	1.481
	r(C1H3)	1.080	1.078	1.078
	r(C1H4)	1.081	1.079	1.078
	r(C2H7)	1.101	1.098	1.098
	r(C2H5)	1.101	1.098	1.097

Table A.1 continued from previous page

	φ(C2NH6)	111.1	110.8	110.9
	φ(C2NH8)	110.6	110.3	110.3
	ϑ(C1C2NH6)	-175.8	-176.4	-176.2
	ϑ(C1C2NH8)	64.8	65.0	65.0
	ϑ(H5C2NH6)	59.8	59.6	59.8
	ϑ(H7C2NH6)	-55.9	-56.6	-56.4
	ϑ(H5C2NH8)	-59.5	-59.0	-59.0
	ϑ(H7C2NH8)	-175.3	-175.2	-175.2
TS1	r(NH8)	1.023	1.021	1.021
	r(C2N)	1.445	1.446	1.446
	r(C1C2)	1.539	1.534	1.534
	r(C1H3)	1.091	1.089	1.089
	r(C1H4)	1.090	1.089	1.089
	r(C1H5)	1.089	1.088	1.088
	r(C2H6)	1.095	1.093	1.093
	r(C2H7)	1.097	1.094	1.094
	φ(C2NH8)	109.0	108.9	108.9
	ϑ(C1C2NH8)	-114.2	-114.6	-114.6
	ϑ(H7C2NH8)	129.0	128.6	128.6
	ϑ(H6C2NH8)	9.7	9.1	9.1
TS2	r(NH8)	1.020	1.019	1.019
	r(C2N)	1.293	1.293	1.293
	r(C1C2)	2.271	2.224	2.224
	r(C1H3)	1.081	1.080	1.080
	r(C1H4)	1.080	1.078	1.078
	r(C1H5)	1.079	1.078	1.078
	r(C2H6)	1.091	1.090	1.090
	r(C2H7)	1.087	1.085	1.085

Table A.1 continued from previous page

	φ(C2NH8)	110.8	110.5	110.5
	ϑ(C1C2NH8)	-89.8	-89.5	-89.5
	ϑ(H7C2NH8)	172.7	172.7	172.7
	ϑ(H6C2NH8)	10.8	11.0	11.0
TS3	r(NH8)	1.019	1.018	1.018
	r(C2N)	1.284	1.286	1.286
	r(C1C2)	1.501	1.500	1.500
	r(C1H3)	1.090	1.088	1.088
	r(C1H4)	1.088	1.086	1.086
	r(C1H5)	1.094	1.092	1.092
	r(C2H6)	1.926	1.829	1.829
	r(C2H7)	1.096	1.094	1.094
	φ(C2NH8)	111.1	110.7	110.7
	ϑ(C1C2NH8)	175.0	174.2	174.2
	ϑ(H7C2NH8)	7.6	8.3	8.3
	ϑ(H6C2NH8)	-79.5	-80.8	-80.8
TS4	r(NH8)	1.022	1.021	1.021
	r(C2N)	1.283	1.285	1.285
	r(C1C2)	1.508	1.506	1.506
	r(C1H3)	1.090	1.088	1.088
	r(C1H4)	1.093	1.092	1.092
	r(C1H5)	1.090	1.089	1.089
	r(C2H6)	1.091	1.090	1.090
	r(C2H7)	1.927	1.830	1.830
	φ(C2NH8)	110.9	110.4	110.4
	ϑ(C1C2NH8)	8.7	9.6	9.6
	ϑ(H7C2NH8)	-97.2	-96.4	-96.4
	ϑ(H6C2NH8)	175.7	175.2	175.2

Table A.1 continued from previous page

TS-isom	r(NH8)	0.987	0.987	0.987
	r(C2N)	1.234	1.237	1.237
	r(C1C2)	1.512	1.510	1.510
	r(C1H3)	1.088	1.087	1.087
	r(C1H4)	1.094	1.092	1.092
	r(C1H5)	1.094	1.092	1.092
	r(C2H6)	1.110	1.107	1.107
	φ(C2NH8)	179.4	179.4	179.4
	ϑ(C1C2NH8)	-0.1	0.6	0.6
	ϑ(H6C2NH8)	179.9	-179.4	-179.4
TS5	r(NH8)	1.020	1.020	1.019
	r(NH6)	1.252	1.251	1.251
	r(C2N)	1.452	1.453	1.451
	r(C1C2)	1.494	1.492	1.491
	r(C1H3)	1.090	1.089	1.088
	r(C1H4)	1.097	1.095	1.094
	r(C1H5)	1.091	1.089	1.088
	r(C2H6)	1.284	1.276	1.275
	r(C2H7)	1.087	1.085	1.084
	φ(C2NH8)	107.6	107.2	107.3
	ϑ(C1C2NH8)	-165.1	-165.6	-165.6
	ϑ(H7C2NH8)	-14.0	-14.4	-14.4
	ϑ(H7C2NH6)	-106.2	-106.3	-106.3
TS6	r(NH8)	1.009	1.009	1.008
	r(NH6)	1.009	1.009	1.008
	r(C2N)	1.413	1.413	1.411
	r(C1C2)	1.483	1.481	1.480
	r(C1H3)	1.080	1.078	1.077

Table A.1 continued from previous page

	r(C1H4)	1.080	1.079	1.078
	r(C1H5)	1.326	1.319	1.318
	r(C2H5)	1.306	1.300	1.299
	r(C2H7)	1.083	1.081	1.080
	φ(C2NH8)	114.7	114.4	114.5
	φ(C2NH6)	113.9	113.3	113.4
	ϑ(C1C2NH8)	145.5	144.1	144.1
	ϑ(C1C2NH6)	16.8	16.6	16.3
	ϑ(H7C2NH8)	-66.2	-67.8	-67.6
	ϑ(H7C2NH6)	165.1	164.8	164.6
TS7	r(NH8)	1.024	1.023	1.022
	r(NH6)	1.024	1.023	1.022
	r(C2N)	2.184	2.156	2.154
	r(C1C2)	1.353	1.351	1.350
	r(C1H3)	1.082	1.080	1.080
	r(C1H4)	1.082	1.080	1.080
	r(C2H5)	1.080	1.079	1.079
	r(C2H7)	1.080	1.079	1-079
	φ(C2NH8)	98.6	98.6	98.6
	φ(C2NH6)	98.6	98.6	98.6
	ϑ(C1C2NH8)	-52.5	-52.4	-52.4
	ϑ(C1C2NH6)	52.5	52.4	52.4
	ϑ(H7C2NH8)	-174.6	-174.5	-174.5
	ϑ(H7C2NH6)	-69.7	-69.6	-69.6
CH_3	r(CH)	1.078x3	1.076x3	1.075x3
	φ(HCH)	120.0x3	120.0x3	120.0x3
CH_2NH	r(C2N)	1.264	1.269	1.268
	φ(C2NH8)	111.4	110.9	110.9

Table A.1 continued from previous page

	φ(H6C2H7)	116.1	116.7	116.7
	ϑ(H6C2NH8)	0.0	0.0	0.0
	ϑ(H7C2NH8)	180.0	180.0	180.0
E-CH_3CHNH	r(C2N)	1.267	1.272	1.270
	r(C1C2)	1.495	1.493	1.492
	r(C2H6)	1.097	1.094	1.094
	φ(C2NH8)	111.5	110.8	110.9
	ϑ(C1C2NH8)	180.0	180.0	180.0
	ϑ(H6C2NH8)	0.0	0.0	0.0
Z-CH_3CHNH	r(C2N)	1.266	1.271	1.269
	r(C1C2)	1.500	1.498	1.498
	r(C2H6)	1.093	1.090	1.090
	φ(C2NH8)	111.3	110.6	110.6
	ϑ(C1C2NH8)	0.0	0.0	0.0
	ϑ(H6C2NH8)	180.0	180.0	180.0
NH_2	r(NH)	1.027	1.024	1.023
	φ(HNH)	103.3	103.3	103.4
C_2H_4	r(C1C2)	1.325	1.328	1.327
	φ(HCC)	121.7x3	121.6x3	121.6x3
	φ(HCH)	116.5x2	116.9x2	116.9x2
	ϑ(H4C1C2H5)	0.0	0.0	0.0
	ϑ(H4C1C2H7)	180.0	180.0	180.0

A.2 Absolute energies

Table A.2 Energies are in Hartree. Reference geometries at B2PLYP-D3(BJ)/aug-cc-pVTZ level of theory.

	B2PLYP-D3(BJ)/aug-cc-pVTZ	CCSD(T)/CBS+CV		CCSD(T)/CBS+CV+fT+pQ		ChS
		T, Q	Q, 5	T, Q	Q, 5	
NH	-55.2025220	-55.2143895	-55.2190340	-55.2149417	-55.2195863	-55.2141825
C_2H_5	-79.1066887	-79.1401015	-79.1476187	-79.1411415	-79.1486587	-79.1405321
M1	-134.4406761	-134.4886002	-134.5007278	-134.4901229	-134.5022506	-134.4890983
M2	-134.4395633	-134.4875802	-134.4997073	-134.4891120	-134.5012390	-134.4881290
M3	-134.4550505	-134.5026174	-134.5147721	-	-	-134.5033022
M4	-134.4374715	-134.4862685	-134.4983261	-	-	-134.4870289
TS1	-134.4379875	-134.4859987	-134.4981322	-134.4875296	-134.4996631	-134.4865507
TS2	-134.3944707	-134.4406919	-134.4529517	-134.4433467	-134.4556065	-134.4413882
TS3	-134.3836749	-134.4314920	-134.4437482	-134.4337551	-134.4460113	-134.4319856
TS4	-134.3826904	-134.4305599	-134.4428110	-134.4328182	-134.4450693	-134.4310511
TS-isom	-133.8465558	-133.8919738	-133.9042667	-133.8934507	-133.9057435	-133.8929897
TS5	-134.3817405	-134.4296502	-134.4417896	-	-	-134.4305127
TS6	-134.3764047	-134.4239981	-134.4361661	-	-	-134.4249503
TS7	-134.3991008	-134.4450773	-134.4574038	-	-	-134.4456568
CH_2NH	-94.5902902	-94.6208211	-94.6293610	-94.6219034	-94.6304432	-94.6216334
CH_3	-39.8125589	-39.8293382	-39.8330384	-39.8299194	-39.8336195	-39.8297072
E-CH_3CHNH	-133.8919820	-133.9391388	-133.9514338	-133.9406117	-133.9529067	-133.9400468
Z-CH_3CHNH	-133.8908845	-133.9380799	-133.9503673	-133.9395581	-133.9518454	-133.9389855
H	-0.4986682	-0.5000222	-0.5000222	-0.5000222	-0.5000222	-0.5000222
C_2H_4	-78.5427309	-78.5759224	-78.5836340	-	-	-78.5765986
NH_2	-55.8589088	-55.8732924	-55.8777758	-	-	-55.8736928

A.3 $\mathcal{T}_1$ diagnostic

Table A.3 Coupled-cluster $\mathcal{T}_1$ diagnostic for all the species considered.

	$\mathcal{T}_1$	
	cc-pVTZ	cc-pVQZ
NH	0.010	0.011
C_2H_5	0.010	0.010
M1	0.012	0.013
M2	0.012	0.013
M3	0.014	0.014
M4	0.010	0.011
TS1	0.012	0.013
TS2	0.031	0.031
TS3	0.031	0.031
TS4	0.031	0.031
TS-isom	0.011	0.012
TS5	0.016	0.016
TS6	0.016	0.016
TS7	0.031	0.031
CH_2NH	0.012	0.012
CH_3	0.008	0.009
E-CH_3CHNH	0.012	0.012
Z-CH_3CHNH	0.012	0.012
C_2H_4	0.011	0.011
NH_2	0.008	0.009

Appendix B

Appendix for Chapter 4

B.1 List of supporting material

1. Table **B.1**. Main structural parameters for the water dimer obtained at the ChS and jun-ChS levels and comparison to values at CCSD(T)-F12b/CBS+fT+fQ+CV+REL+DBOC level.

2. Table **B.2**. Mean Absolute Errors and Relative Errors for bond lengths, angles and dihedrals of the HCN$\cdots$$(H_2O)_2$ system.

3. Figure **B.1**. MAE and RE for bond lengths, valence and dihedral angles.

4. Table **B.3**. Contributions to jun-ChS electronic energies.

5. Table **B.4**. Absolute Errors and Mean Absolute Errors of formation energies computed at the jun-ChS level on top of different optimized geometries with respect to full jun-ChS results.

6. Figure **B.2**. Structure of the HCN@(H_2O) system showing molecules frozen during the optimization.

7. Table **B.5**. Relative ground state electronic energies for the stationary points on the HCN$\rightleftharpoons$HNC isomerization PES with respect to isolated HCN and $(H_2O)_n$ for $n = 2, 20$.

Table B.1 Main structural parameters for the water dimer obtained at the ChS and jun-ChS levels and comparison to values at CCSD(T)-F12b/CBS+fT+fQ+CV+REL+DBOC level of ref.[298]. The H-O-H valence angles for the acceptor and the donor respectively are reported in the first column (θ_a and θ_d). α gives a measure of the deviation from a linear hydrogen bond and β gives the orientation of the C_2 axis of the acceptor with respect to the O-O axis. Bond lengths in Å and angles in degrees.

	ChS	**jun-ChS**	**Ref. value**[298]
r(O-O)	2.9049	2.9058	2.9092
$\boldsymbol{\theta_a}$	104.78	104.77	104.95
$\boldsymbol{\theta_d}$	104.87	104.84	104.85
$\boldsymbol{\alpha}$	4.91	5.91	5.69
$\boldsymbol{\beta}$	126	127	123.46

Table B.2 MAE and RE for bond lengths (**r**), angles($\boldsymbol{\alpha}$) and dihedrals ($\boldsymbol{\phi}$) of the HCN–$(H_2O)_2$ system. Averaged values over all the species along the PES are collected. Last column collect total RE obtained averaging over all the structural parameters. MAE for bond lengths in Å and in degrees for angles and dihedrals. RE are in %.

		MAE			**RE**			
		r	$\boldsymbol{\alpha}$	$\boldsymbol{\phi}$	**r**	$\boldsymbol{\alpha}$	$\boldsymbol{\phi}$	**Total**
B3LYP-D3	**Jun-DZ**	0.028	1.72	2.21	1.78	1.31	2.43	1.02
	Jul-DZ	0.023	1.45	3.56	1.47	1.23	3.65	0.98
	Aug-DZ	0.021	1.24	2.53	1.36	1.08	2.61	0.88
	Jun-TZ	0.013	1.07	2.69	0.87	0.89	3.08	0.69
	Jul-TZ	0.013	1.03	2.44	0.86	0.86	2.69	0.65
	Aug-TZ	0.013	1.09	2.51	0.89	0.92	2.82	0.69
BHLYP-D3	**Jun-DZ**	0.092	7.34	9.21	3.50	6.17	10.47	4.82
	Jul-DZ	0.016	1.64	2.52	0.87	1.39	2.38	0.85
	Aug-DZ	0.014	1.46	1.64	0.79	1.25	1.81	0.76
	Jun-TZ	0.014	1.28	1.59	0.96	1.07	1.65	0.80
	Jul-TZ	0.014	1.34	1.54	0.95	1.10	1.82	0.83
	Aug-TZ	0.015	1.34	1.48	0.97	1.11	1.72	0.83

Table B.2 continued from previous page

ωB97X-D	**Jun-DZ**	0.030	2.36	4.01	1.65	1.90	3.28	1.06
	Jul-DZ	0.023	1.82	5.01	1.25	1.57	4.86	0.91
	Aug-DZ	0.021	1.70	5.15	1.14	1.48	4.89	0.92
	Jun-TZ	0.017	1.63	5.06	0.88	1.37	5.23	0.81
	Jul-TZ	0.017	1.59	4.87	0.88	1.34	4.95	0.78
	Aug-TZ	0.018	1.62	5.00	0.93	1.38	5.17	0.81
PW6B95-D3	**Jun-DZ**	0.022	2.03	2.47	1.23	1.50	2.48	1.00
	Jul-DZ	0.015	1.34	3.11	0.80	1.09	3.05	0.74
	Aug-DZ	0.013	1.05	1.96	0.73	0.90	1.65	0.60
	Jun-TZ	0.011	0.96	2.56	0.55	0.79	2.59	0.55
	Jul-TZ	0.010	0.94	2.31	0.53	0.76	2.29	0.52
	Aug-TZ	0.010	0.95	2.34	0.52	0.78	2.36	0.53
BMK-D3	**Jun-DZ**	0.015	1.87	3.32	1.00	1.36	2.71	0.75
	Jul-DZ	0.013	1.67	4.09	0.84	1.38	3.84	0.77
	Aug-DZ	0.012	1.60	2.84	0.77	1.34	2.47	0.69
	Jun-TZ	0.008	1.03	2.32	0.52	0.84	2.04	0.57
	Jul-TZ	0.008	1.00	2.00	0.52	0.81	1.61	0.54
	Aug-TZ	0.008	1.03	2.02	0.51	0.84	1.68	0.56
M06-2X	**Jun-DZ**	0.014	1.95	2.65	1.05	1.37	2.99	1.09
	Jul-DZ	0.013	1.16	2.96	0.91	0.96	2.72	0.85
	Aug-DZ	0.014	1.18	2.62	0.90	0.96	2.74	0.84
	Jun-TZ	0.012	1.05	2.36	0.72	0.82	2.06	0.70
	Jul-TZ	0.012	1.00	1.90	0.74	0.79	1.39	0.65
	Aug-TZ	0.012	1.04	1.86	0.75	0.83	1.41	0.66
MN15	**Jun-DZ**	0.017	2.15	3.08	1.18	1.59	2.62	1.15
	Jul-DZ	0.013	1.49	3.70	0.94	1.17	3.23	0.95
	Aug-DZ	0.010	1.20	2.80	0.82	0.99	2.87	0.83
	Jun-TZ	0.009	1.03	3.13	0.63	0.82	2.48	0.68

Table B.2 continued from previous page

	Jul-TZ	0.008	1.04	2.69	0.60	0.82	1.99	0.65
	Aug-TZ	0.008	1.01	2.60	0.58	0.80	1.93	0.65
B2PLYP-D3	**Jun-DZ**	0.025	1.88	2.36	1.58	1.39	2.89	1.10
	Jul-DZ	0.020	1.23	2.71	1.29	1.03	2.62	0.88
	Aug-DZ	0.018	1.00	1.74	1.21	0.88	1.65	0.75
	Jun-TZ	0.010	0.77	1.65	0.63	0.63	1.77	0.51
	Jul-TZ	0.009	0.69	2.13	0.58	0.57	2.72	0.59
	Aug-TZ	0.009	0.74	1.50	0.60	0.63	1.60	0.48
DSD-PBEP86-D3	**Jun-DZ**	0.021	1.78	2.48	1.44	1.28	3.02	1.07
	Jul-DZ	0.017	0.88	2.63	1.18	0.74	2.60	0.76
	Aug-DZ	0.016	0.96	2.39	1.15	0.84	2.64	0.81
	Jun-TZ	0.008	0.54	1.50	0.60	0.44	1.69	0.43
	Jul-TZ	0.007	0.46	1.39	0.57	0.37	1.55	0.41
	Aug-TZ	0.007	0.59	1.53	0.58	0.49	1.73	0.44
rDSD-PBEP86-D3	**Jun-DZ**	0.023	1.80	2.53	1.49	1.31	2.98	1.11
	Jul-DZ	0.017	0.89	2.64	1.15	0.75	2.58	0.77
	Aug-DZ	0.016	0.96	2.38	1.12	0.83	2.64	0.80
	Jun-TZ	0.008	0.52	1.51	0.58	0.42	1.66	0.43
	Jul-TZ	0.007	0.43	1.36	0.54	0.35	1.50	0.40
	Aug-TZ	0.007	0.54	1.53	0.55	0.45	1.73	0.42

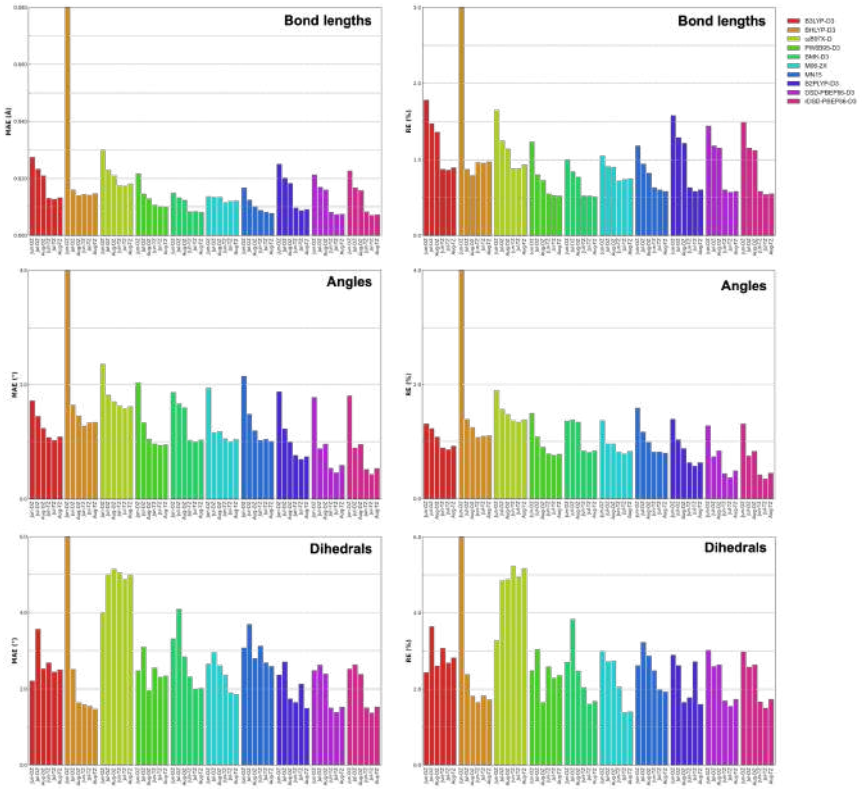

Figure B.1 MAE and RE for bond lengths, valence and dihedral angles. The values are obtained by averaging absolute and relative errors of structural parameters over all the species involved in the reactive PES for the HCN⇌HNC isomerization assisted by two water molecules.

Table B.3 jun-ChS contributions (in Hartree) to energy evaluated on jun-ChS geometries. E(CC) is the CCSD(T) energy computed with the jun-cc-pVTZ basis set; ΔE^{∞}_{MP2} is the difference between the fc-MP2/jun-cc-pVTZ energy and the corresponding extrapolated value estimated by the jun-cc-pVnZ basis sets with n = T and Q. ΔE^{CV}_{MP2} accounts for core-valence correlation and is obtained as difference between ae- and fc- MP2 calculations with cc-pwCVTZ basis set. The final jun-ChS energy is reported in the last row.

	HCN	**$(H_2O)_2$**	**HCN–$(H_2O)_2$**	**TS**	**HNC–$(H_2O)_2$**	**HNC**
E(CC)	-93.2774	-152.6856	-245.9753	-245.9254	-245.9574	-93.2540
ΔE^{∞}_{MP2}	-0.0446	-0.0818	-0.1268	-0.1262	-0.1264	-0.0444
ΔE^{CV}_{MP2}	-0.0944	-0.1029	-0.1974	-0.1974	-0.1973	-0.0943
E(jun-ChS)	-93.4165	-152.8703	-246.2995	-246.2490	-246.2810	-93.3927

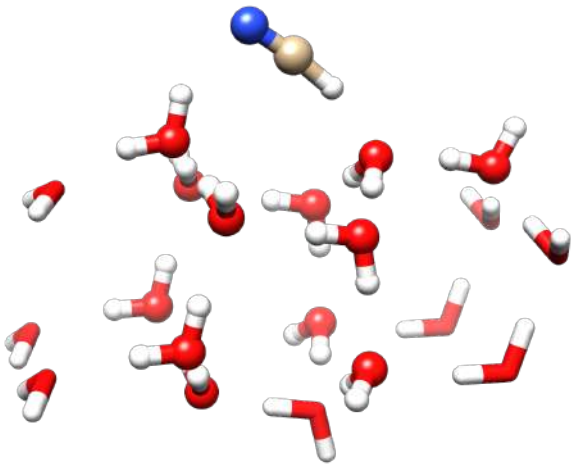

Figure B.2 Structure of the HCN@(H_2O) system showing molecules frozen during the optimization. Ball and stick representation used for atoms free to move while tubular representation for molecules kept frozen in order to prevent structural distortion of the cluster.

Table B.5 Relative ground state electronic energies for the stationary points on the HCN⇌HNC isomerization PES with respect to isolated HCN and $(H_2O)_n$. The number of water molecules involved in the relay mechanism is explicitly indicated together with the total number of water molecules included in the model system (in parenthesis). ZPVEs are in bold while energy corrected for ZPVEs are in parenthesis. Values in kJ/mol. jun-ChS energy and geometries and ZPVEs at DSD-PBEP86-D3/jul-cc-pVTZ for n=2. jun-ChS energies for the adsorbate and the water molecules involved in the proton relay and PW6B95D3/jul-cc-pVDZ for the remaining water molecules in the cluster for n=4 [20]. PW6B95-D3/jul-cc-pVDZ for PW6B95 results. DSD-PBEP86 results refers to ONIOM(DSD-PBEP86/jul-cc-pVTZ:PW6B95D3/jul-cc-pVDZ), both geometries and ZPVEs at this level.

		HCN–$(H_2O)_n$	**TS**	**HNC–$(H_2O)_n$**	**HNC + $(H_2O)_n$**
jun-ChS	n=2 [2]	-33.4	99.3	15.0	62.3
		6.3	**1.9**	**7.8**	**-1.3**
		(-27.2)	(101.2)	(22.8)	(61.0)
	n = 4 [20]	-61.4	43.4	-18.5	62.3
		10.6	**-9.3**	**8.6**	**-2.0**
		(-50.7)	(34.2)	(-9.9)	(60.2)
PW6B95	n=4 [20]	-64.1	27.6	-27.8	56.7
		10.6	**-9.3**	**8.6**	**-2.0**
		(-53.5)	(18.4)	(-19.2)	(54.7)
DSD-PBEP86	n=4 [20]	-65.4	28.3	-21.6	64.9
		7.9	**-8.9**	**8.1**	**-1.3**
		(-57.5)	(19.4)	(-13.4)	(63.6)

Table B.4 Absolute error and MAE of formation energies computed at the jun-ChS level on top of different optimized geometries with respect to full jun-ChS results. The formation energy of each species (in kJ/mol) is calculated with respect to isolated HCN + $(H_2O)_2$. RC stands for reactant complex, i. e. NCH–$(H_2O)_2$, TS for transition state and PC for product complex, i. e. CNH–$(H_2O)_2$.

	RC	**TS**	**PC**	**HCN + $(H_2O)_2$**	**MAE**
PW6B95-D3/jul-DZ	0.04	0.11	0.20	0.04	0.10
BHLYP-D3/aug-DZ	0.01	0.05	0.16	0.39	0.15
PW6B95-D3/aug-DZ	0.02	0.15	0.13	0.01	0.08
BMK-D3/aug-DZ	0.29	0.12	0.00	0.01	0.11
M06-2X/aug-DZ	0.09	0.41	0.34	0.01	0.21
MN15/aug-DZ	0.03	0.11	0.57	0.36	0.27
PW6B95-D3/jul-TZ	0.07	0.11	0.30	0.38	0.22
BMK-D3/jul-TZ	0.05	0.08	0.45	0.61	0.30
M06-2X/jul-TZ	0.04	0.23	0.12	0.42	0.20
MN15/jul-TZ	0.09	0.07	0.05	0.24	0.11

DSD-PBEP86-D3/jul-TZ	0.10	0.09	0.03	0.05	0.07
rDSD-PBEP86-D3/jul-TZ	0.12	0.06	0.02	0.05	0.06

List of Figures

Bibliography

[1] Collaboration, T. E. H. T. et al. First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *The Astrophysical Journal Letters* **2019**, *875*, L1–17.

[2] Caselli, P.; Ceccarelli, C. Our astrochemical heritage. *The Astronomy and Astrophysics Review* **2012**, *20*, 56–123.

[3] Kitadai, N.; Maruyama, S. Origins of building blocks of life: A review. *Geoscience Frontiers* **2018**, *9*, 1117–1153.

[4] Camprubí, E.; de Leeuw, J. W.; House, C. H.; Raulin, F.; Russell, M. J.; Spang, A.; Tirumalai, M. R.; Westall, F. The Emergence of Life. *Space Science Reviews* **2019**, *215*, 1–53.

[5] Hartmann, J. Investigations on the spectrum and orbit of delta Orionis. *Astrophysical Journal* **1904**, *19*, 268–286.

[6] Trumpler, R. J. Absorption of Light in the Galactic System. *Publications of the Astronomical Society of the Pacific* **1930**, *42*, 214–227.

[7] Puzzarini, C. Grand Challenges in Astrochemistry. *Frontiers in Astronomy and Space Sciences* **2020**, *7*, 1–4.

[8] van Dishoeck, E. F. Astrochemistry: overview and challenges. *Proceedings of the International Astronomical Union* **2017**, *13*, 3–22.

[9] Herbst, E.; John T. Yates, J. Introduction: Astrochemistry. *Chemical Reviews* **2013**, *113*, 8707–8709.

[10] Swings, P.; Rosenfeld, L. Considerations Regarding Interstellar Molecules. *Astrophysical Journal* **1937**, *86*, 483–486.

[11] Barrett, A. H. Radio Observations of Interstellar Hydroxyl Radicals. *Science* **1967**, *157*, 881–889.

[12] Herzberg, G. Historical Remarks on the Discovery of Interstellar Molecules. *Journal of the Royal Astronomical Society of Canada* **1988**, *82*, 115–127.

[13] Cheung, A. C.; Rank, D. M.; Townes, C. H.; Thornton, D. D.; Welch, W. J. Detection of NH_3 Molecules in the Interstellar Medium by Their Microwave Emission. *Physical Review Letters* **1968**, *21*, 1701–1705.

[14] Cheung, A. C.; Rank, D. M.; Townes, C. H.; Thornton, D. D.; Welch, W. J. Detection of Water in Interstellar Regions by its Microwave Radiation. *Nature* **1969**, *221*, 626–628.

[15] Snyder, L. E.; Buhl, D.; Zuckerman, B.; Palmer, P. Microwave Detection of Interstellar Formaldehyde. *Physical Review Letters* **1969**, *22*, 679–681.

[16] Puget, J. L.; Leger, A.; Boulanger, F. Contribution of large polycyclic aromatic molecules to the infrared emission of the interstellar medium. *Astronomy and Astrophysics* **1985**, *142*, L19–L22.

[17] Duley, W. W.; Williams, D. A. The infrared spectrum of interstellar dust: Surface functional groups on carbon. *Monthly Notices of the Royal Astronomical Society* **1981**, *196*, 269–274.

[18] Allamandola, L. J.; Tielens, A. G. G. M.; Barker, J. R. Polycyclic aromatic hydrocarbons and the unidentified infrared emission bands: auto exhaust along the milky way. *Astrophysical Journal Letters* **1985**, *290*, L25–L28.

[19] Peeters, E.,; Hony, S.,; Van Kerckhoven, C.,; Tielens, A. G. G. M.,; Allamandola, L. J.,; Hudgins, D. M.,; Bauschlicher, C. W., The rich 6 to

9 μm spectrum of interstellar PAHs*. *Astronomy and Astrophysics* **2002**, *390*, 1089–1113.

[20] Onaka, T. Unidentified Infrared Bands in the Diffuse Interstellar Medium. *Astronomical Society of the Pacific Conference Series* **2004**, *309*, 163–178.

[21] contributors, W. List of interstellar and circumstellar molecules — Wikipedia, The Free Encyclopedia. 2023; `https://en.wikipedia.org/w/index.php?title=List_of_interstellar_and_circumstellar_molecules$&$oldid=1132257236`, [Online; accessed 11-January-2023].

[22] Müller, H. S. P.; Thorwirth, S.; Roth, D. A.; Winnewisser, G. The Cologne Database for Molecular Spectroscopy, CDMS. *Astronomy and Astrophysics* **2001**, *370*, L49–L52.

[23] Müller, H. S.; Schlöder, F.; Stutzki, J.; Winnewisser, G. The Cologne Database for Molecular Spectroscopy, CDMS: a useful tool for astronomers and spectroscopists. *Journal of Molecular Structure* **2005**, *742*, 215–227.

[24] McGuire, B. A. Census of Interstellar, Circumstellar, Extragalactic, Protoplanetary Disk, and Exoplanetary Molecules. *Astrophysical Journal Supplement Series* **2018**, *239*, 17–65.

[25] McGuire, B. A. 2021 Census of Interstellar, Circumstellar, Extragalactic, Protoplanetary Disk, and Exoplanetary Molecules. *Astrophysical Journal Supplement Series* **2022**, *259*, 30–80.

[26] Herbst, E.; van Dishoeck, E. F. Complex Organic Interstellar Molecules. *Annual Review of Astronomy & Astrophysics* **2009**, *47*, 427–480.

[27] Kuan, Y.-J.; Charnley, S. B.; Huang, H.-C.; Kisiel, Z.; Ehrenfreund, P.; Tseng, W.-L.; Yan, C.-H. Searches for interstellar molecules of potential prebiotic importance. *Advances in Space Research* **2004**, *33*, 31–39.

[28] Hollis, J. M.; Lovas, F. J.; Jewell, P. R. Interstellar Glycolaldehyde: The First Sugar. *Astrophysical Journal* **2000**, *540*, L107–L110.

[29] Jiménez-Serra, I.; Martín-Pintado, J.; Rivilla, V. M.; Rodríguez-Almeida, L.; Alonso Alonso, E. R.; Zeng, S.; Cocinero, E. J.; Martín, S.; Requena-Torres, M.; Martín-Domenech, R.; Testi, L. Toward the RNA-World in the Interstellar Medium—Detection of Urea and Search of 2-Amino-oxazole and Simple Sugars. *Astrobiology* **2020**, *20*, 1048–1066.

[30] Cooper, G.; Kimmich, N.; Belisle, W.; Sarinana, J.; Brabham, K.; Garrel, L. Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature* **2001**, *414*, 879–883.

[31] Pizzarello, S.; Schrader, D. L.; Monroe, A. A.; Lauretta, D. S. Large enantiomeric excesses in primitive meteorites and the diverse effects of water in cosmochemical evolution. *Proceedings of the National Academy of Sciences* **2012**, *109*, 11949–11954.

[32] Furukawa, Y.; Chikaraishi, Y.; Ohkouchi, N.; Ogawa, N. O.; Glavin, D. P.; Dworkin, J. P.; Abe, C.; Nakamura, T. Extraterrestrial ribose and other sugars in primitive meteorites. *Proceedings of the National Academy of Sciences* **2019**, *116*, 24440–24445.

[33] de Marcellus, P.; Meinert, C.; Myrgorodska, I.; Nahon, L.; Buhse, T.; d'Hendecourt, L. L. S.; Meierhenrich, U. J. Aldehydes and sugars from evolved precometary ice analogs: Importance of ices in astrochemical and prebiotic evolution. *Proceedings of the National Academy of Sciences* **2015**, *112*, 965–970.

[34] Meinert, C.; Myrgorodska, I.; de Marcellus, P.; Buhse, T.; Nahon, L.; Hoffmann, S. V.; d'Hendecourt, L. L. S.; Meierhenrich, U. J. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **2016**, *352*, 208–212.

[35] Fedoseev, G.; Chuang, K.-J.; Ioppolo, S.; Qasim, D.; van Dishoeck, E. F.; Linnartz, H. Formation of Glycerol through Hydrogenation of CO Ice under Prestellar Core Conditions. *Astrophysical Journal* **2017**, *842*, 52–60.

[36] Hollis, J. M.; Lovas, F. J.; Remijan, A. J.; Jewell, P. R.; Ilyushin, V. V.; Kleiner, I. Detection of Acetamide (CH_3CONH_2): The Largest Interstellar Molecule with a Peptide Bond. *Astrophysical Journal Letters* **2006**, *643*, L25–L28.

[37] Belloche, A.,; Menten, K. M.,; Comito, C.,; Müller, H. S. P.,; Schilke, P.,; Ott, J.,; Thorwirth, S.,; Hieret, C., Detection of amino acetonitrile in Sgr B2(N). *Astronomy and Astrophysics* **2008**, *492*, 769–773.

[38] contributors, W. List of space telescopes — Wikipedia, The Free Encyclopedia. 2023; `https://en.wikipedia.org/w/index.php?title=List_of_space_telescopes&oldid=1131428340`, [Online; accessed 12-January-2023].

[39] Baudry, A. The ALMA radio telescope. *Proceedings of Science* **2009**, *2nd MCCT-SKADS*, 002.

[40] Wootten, A.; Thompson, A. The Atacama Large Millimeter/Submillimeter Array. *Proceedings of the IEEE* **2009**, *97*, 1463–1471.

[41] Gardner, J. P. et al. The James Webb Space Telescope. *Space Science Reviews* **2006**, *123*, 485–606.

[42] van Dishoeck, E. F.; Merín, B.; Brandl, B.; Böker, T.; Greene, T.; Meixner, M.; Ressler, M.; Rieke, G.; Waelkens, C.; Wright, G.; Miri Team, In *Astrochemistry: Recent Successes and Current Challenges*; Lis, D. C., Blake, G. A., Herbst, E., Eds.; 2005; Vol. 231; p 328.

[43] Barone, V.; Biczysko, M.; Puzzarini, C. Quantum Chemistry Meets Spectroscopy for Astrochemistry: Increasing Complexity toward Prebiotic Molecules. *Accounts of Chemical Research* **2015**, *48*, 1413–1422.

[44] Biczysko, M.; Bloino, J.; Puzzarini, C. Computational challenges in Astrochemistry. *WIREs Computational Molecular Science* **2018**, *8*, e1349–1386.

[45] Herbst, E. Unusual Chemical Processes in Interstellar Chemistry: Past and Present. *Frontiers in Astronomy and Space Sciences* **2021**, *8*, 1–18.

[46] Yamamoto, S. *Introduction to Astrochemistry: Chemical Evolution from Interstellar Clouds to Star and Planet Formation*; Astronomy and Astrophysics Library; Springer Japan, 2017; p 293.

[47] Hollenbach, D. J.; Werner, M. W.; Salpeter, E. E. Molecular Hydrogen in H I Regions. *Astrophysical Journal* **1971**, *163*, 165–180.

[48] Wakelam, V.; Bron, E.; Cazaux, S.; Dulieu, F.; Gry, C.; Guillard, P.; Habart, E.; Hornekær, L.; Morisset, S.; Nyman, G.; Pirronello, V.; Price, S. D.; Valdivia, V.; Vidali, G.; Watanabe, N. H_2 formation on interstellar dust grains: The viewpoints of theory, experiments, models and observations. *Molecular Astrophysics* **2017**, *9*, 1–36.

[49] Garrod, R. T.; Weaver, S. L. W.; Herbst, E. Complex Chemistry in Star-forming Regions: An Expanded Gas-Grain Warm-up Chemical Model. *Astrophysical Journal* **2008**, *682*, 283–302.

[50] Burke, D. J.; Brown, W. A. Ice in space: surface science investigations of the thermal desorption of model interstellar ices on dust grain analogue surfaces. *Physical Chemistry Chemical Physics* **2010**, *12*, 5947–5969.

[51] Draine, B. Interstellar Dust Grains. *Annual Review of Astronomy and Astrophysics* **2003**, *41*, 241–289.

[52] Bowey, J. E.; Lee, C.; Tucker, C.; Hofmeister, A. M.; Ade, P. A. R.; Barlow, M. J. Temperature effects on the 15-85 μm spectra of olivines and pyroxenes. *Montly Notices of the Royal Astronomical Society* **2001**, *325*, 886–896.

[53] Koike, C.; Noguchi, R.; Chihara, H.; Suto, H.; Ohtaka, O.; Imai, Y.; Matsumoto, T.; Tsuchiyama, A. INFRARED SPECTRA OF SILICA

POLYMORPHS AND THE CONDITIONS OF THEIR FORMATION. *The Astrophysical Journal* **2013**, *778*, 1–12.

[54] Allamandola, L. J.; Bernstein, M. P.; Sandford, S. A.; Walker, R. L. Evolution of Interstellar Ices. *Space Science Reviews* **1999**, *90*, 219–232.

[55] Ehrenfreund, P.; Schutte, W. ISO observations of interstellar ices: Implications for the pristinity of comets. *Advances in Space Research* **2000**, *25*, 2177–2188.

[56] Gibb, E. L.; Whittet, D. C. B.; Boogert, A. C. A.; Tielens, A. G. G. M. Interstellar Ice: The Infrared Space Observatory Legacy. *Astrophysical Journal Supplement Series* **2004**, *151*, 35–73.

[57] Palumbo, M. E. The morphology of interstellar water ice. *Journal of Physics: Conference Series* **2005**, *6*, 211–216.

[58] Li, H.; Karina, A.; Ladd-Parada, M.; Späh, A.; Perakis, F.; Benmore, C.; Amann-Winkel, K. Long-Range Structures of Amorphous Solid Water. *Journal of Physical Chemistry B* **2021**, *125*, 13320–13328.

[59] Fraser, H. J.; Collings, M. P.; McCoustra, M. R. S.; Williams, D. A. Thermal desorption of water ice in the interstellar medium. *Monthly Notices of the Royal Astronomical Society* **2001**, *327*, 1165–1172.

[60] Allamandola, L.; Sandford, S.; Valero, G. Photochemical and thermal evolution of interstellar/precometary ice analogs. *Icarus* **1988**, *76*, 225–252.

[61] Muñoz Caro, G. M.,; Meierhenrich, U.,; Schutte, W. A.,; W. H.-P. Thiemann,; Greenberg, J. M., UV-photoprocessing of interstellar ice analogs: Detection of hexamethylenetetramine-based species. *Astronomy and Astrophysics* **2004**, *413*, 209–216.

[62] Martín-Doménech, R.; Öberg, K. I.; Rajappan, M. Formation of NH_2CHO

and CH_3CHO upon UV Photoprocessing of Interstellar Ice Analogs. *Astrophysical Journal* **2020**, *894*, 98–113.

[63] Deshapriya, J. D. P., et al. Exposed bright features on the comet 67P/Churyumov-Gerasimenko: distribution and evolution. *Astronomy and Astrophysics* **2018**, *613*, A36, 1–14.

[64] Drozdovskaya, M. N.; van Dishoeck, E. F.; Rubin, M.; Jørgensen, J. K.; Altwegg, K. Ingredients for solar-like systems: protostar IRAS 16293-2422 B versus comet 67P/Churyumov–Gerasimenko. *Monthly Notices of the Royal Astronomical Society* **2019**, *490*, 50–79.

[65] Burbidge, E. M.; Burbidge, G. R.; Fowler, W. A.; Hoyle, F. Synthesis of the Elements in Stars. *Reviews of Modern Physics* **1957**, *29*, 547–650.

[66] Morse, A. D.; Chan, Q. H. S. Observations of Cometary Organics: A Post Rosetta Review. *ACS Earth and Space Chemistry* **2019**, *3*, 1773–1791.

[67] Hänni, N.; Altwegg, K.; Combi, M.; Fuselier, S. A.; De Keyser, J.; Rubin, M.; Wampfler, S. F. Identification and characterization of a new ensemble of cometary organic molecules. *Nature Communications* **2022**, *13*, 3639–3650.

[68] Koga, T.; Naraoka, H. A new family of extraterrestrial amino acids in the Murchison meteorite. *Scientific Reports* **2017**, *7*, 636–643.

[69] Pizzarello, S.; Shock, E. Carbonaceous chondrite meteorites: The chronicle of a potential evolutionary path between stars and life. *Origins of Life and Evolution of Biospheres* **2017**, *47*, 249–260.

[70] Hoover, r. b. Comets, carbonaceous meteorites, and the origin of the biosphere. *Biogeosciences Discussions* **2006**, *3*, 23–70.

[71] Oparin, A. I. In *The Origin of Life*; MacMillan,, Ed.; 1938.

[72] Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529.

[73] Chyba, C. F.; Thomas, P. J.; Brookshaw, L.; Sagan, C. Cometary Delivery of Organic Molecules to the Early Earth. *Science* **1990**, *249*, 366–373.

[74] Baiano, C.; Lupi, J.; Tasinato, N.; Puzzarini, C.; Barone, V. The Role of State-of-the-Art Quantum-Chemical Calculations in Astrochemistry: Formation Route and Spectroscopy of Ethanimine as a Paradigmatic Case. *Molecules* **2020**, *25*.

[75] Baiano, C.; Lupi, J.; Barone, V.; Tasinato, N. Gliding on Ice in Search of Accurate and Cost-Effective Computational Methods for Astrochemistry on Grains: The Puzzling Case of the HCN Isomerization. *Journal of Chemical Theory and Computation* **2022**, *18*, 3111–3121.

[76] Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review* **1926**, *28*, 1049–1070.

[77] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **1927**, *389*, 457–484.

[78] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; McGraw-Hill, 1982.

[79] Jensen, F. *Introduction to computational chemistry*, 3rd ed.; John Wiley & Sons, 2017.

[80] Cramer, C. J. *Essentials of computational chemistry : theories and models*; West Sussex, England ; New York : J. Wiley, [2002] ©2002, 2002.

[81] Sherrill, C. D.; Schaefer III, H. F. In *The Configuration Interaction Method: Advances in Highly Correlated Approaches*; Lowdin, M. C. Z. P.-O., Sabin, E. B. J. R., Eds.; Advances in Quantum Chemistry; Academic Press: London, 1999; Vol. 34; pp 143 – 269.

[82] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618–622.

[83] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989**, *157*, 479–483.

[84] Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *Journal of Chemical Physics* **1969**, *51*, 2657–2664.

[85] Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *Journal of Chemical Physics* **1995**, *103*, 4572–4585.

[86] Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *Journal of the American Chemical Society* **1980**, *102*, 939–947.

[87] Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *Journal of the American Chemical Society* **1982**, *104*, 2797–2803.

[88] Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta* **1973**, *28*, 213–222.

[89] Raffenetti, R. C. General contraction of Gaussian atomic orbitals: Core, valence, polarization, and diffuse basis sets; Molecular integral evaluation. *Journal of Chemical Physics* **1973**, *58*, 4452–4458.

[90] Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *Journal of Chemical Theory and Computation* **2011**, *7*, 3027–3034.

[91] Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A.

Gaussian-1 theory: A general procedure for prediction of molecular energies. *Journal of Chemical Physics* **1989**, *90*, 5622–5629.

[92] Martin, J. M. L.; de Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry—W1 and W2 theory. *Journal of Chemical Physics* **1999**, *111*, 1843–1856.

[93] Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theoretical Chemistry Accounts* **2012**, *131*, 1079.

[94] Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. *Journal of Chemical Physics* **2004**, *121*, 11599–11613.

[95] Bomble, Y. J.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P. G.; Császár, A. G.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. II. Minor improvements to the protocol and a vital simplification. *Journal of Chemical Physics* **2006**, *125*, 064108.

[96] Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *Journal of Chemical Physics* **2008**, *128*, 114111.

[97] Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Multi-coefficient extrapolated density functional theory for thermochemistry and thermochemical kinetics. *Physical Chemistry Chemical Physics* **2005**, *7*, 43–52.

[98] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *Journal of Chemical Physics* **2006**, *124*, 114104.

[99] Feller, D. Application of systematic sequences of wave functions to the water dimer. *Journal of Chemical Physics* **1992**, *96*, 6104–6114.

[100] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. *Chemical Physics Letters* **1998**, *286*, 243–252.

[101] Varandas, A. J. C. Basis-set extrapolation of the correlation energy. *Journal of Chemical Physics* **2000**, *113*, 8880–8887.

[102] Lesiuk, M.; Jeziorski, B. Complete Basis Set Extrapolation of Electronic Correlation Energies Using the Riemann Zeta Function. *Journal of Chemical Theory and Computation* **2019**, *15*, 5398–5403.

[103] Martin, J. M. Ab initio total atomization energies of small molecules — towards the basis set limit. *Chemical Physics Letters* **1996**, *259*, 669–678.

[104] Bakowies, D. Accurate extrapolation of electron correlation energies from small basis sets. *Journal of Chemical Physics* **2007**, *127*, 164109.

[105] Varandas, A. J. C. Extrapolating to the one-electron basis-set limit in electronic structure calculations. *Journal of Chemical Physics* **2007**, *126*, 244105.

[106] Matthews, D. A.; Cheng, L.; Harding, M. E.; Lipparini, F.; Stopkowicz, S.; Jagau, T.-C.; Szalay, P. G.; Gauss, J.; Stanton, J. F. Coupled-cluster techniques for computational chemistry: The CFOUR program package. *Journal of Chemical Physics* **2020**, *152*, 214108.

[107] Truhlar, D. G. Basis-set extrapolation. *Chemical Physics Letters* **1998**, *294*, 45–48.

[108] Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *Journal of Chemical Theory and Computation* **2010**, *6*, 2365–2376.

[109] Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics* **2006**, *8*, 1985–1993.

[110] Heckert, M.; KÁllay, M.; Gauss, J. Molecular equilibrium geometries based on coupled-cluster calculations including quadruple excitations. *Molecular Physics* **2005**, *103*, 2109–2115.

[111] Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory. *Journal of Chemical Physics* **2006**, *125*, 044108.

[112] Puzzarini, C. Extrapolation to the Complete Basis Set Limit of Structural Parameters: Comparison of Different Approaches. *Journal of Physical Chemistry A* **2009**, *113*, 14530–14535.

[113] Lupi, J.; Alessandrini, S.; Puzzarini, C.; Barone, V. junChS and junChS-F12 Models: Parameter-free Efficient yet Accurate Composite Schemes for Energies and Structures of Noncovalent Complexes. *Journal of Chemical Theory and Computation* **2021**, *17*, 6974–6992.

[114] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864–B871.

[115] Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics* **2017**, *115*, 2315–2372.

[116] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chemical Reviews* **2012**, *112*, 289–320.

[117] Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Review of Modern Physics* **2015**, *87*, 897–923.

[118] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133–A1138.

[119] Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *Journal of Chemical Physics* **2005**, *123*, 062201.

[120] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.

[121] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, *38*, 3098–3100.

[122] Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple". *Physical Review Letters* **1998**, *80*, 890–890.

[123] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **1992**, *46*, 6671–6687.

[124] Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B* **1986**, *33*, 8822–8824.

[125] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, *37*, 785–789.

[126] Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and assessment of new exchange-correlation functionals. *Journal of Chemical Physics* **1998**, *109*, 6264–6271.

[127] Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik, M. New generalized gradient approximation functionals. *Journal of Chemical Physics* **2000**, *112*, 1670–1678.

[128] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Physical Review Letters* **2003**, *91*, 146401.

[129] Becke, A. D. A new inhomogeneity parameter in density-functional theory. *Journal of Chemical Physics* **1998**, *109*, 2092–2098.

[130] Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *Journal of Chemical Physics* **1993**, *98*, 1372–1377.

[131] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics* **1993**, *98*, 5648–5652.

[132] Hertwig, R. H.; Koch, W. On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chemical Physics Letters* **1997**, *268*, 345–351.

[133] Zhao, Y.; Truhlar, D. G. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *Journal of Physical Chemistry A* **2005**, *109*, 5656–5667.

[134] Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06 functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

[135] Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn–Sham global-hybrid exchange–correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chemical Science* **2016**, *7*, 5032–5051.

[136] Boese, A. D.; Martin, J. M. L. Development of density functionals for

thermochemical kinetics. *Journal of Chemical Physics* **2004**, *121*, 3405–3416.

[137] Gill, P. M. W.; Adamson, R. D.; Pople, J. A. Coulomb-attenuated exchange energy density functionals. *Molecular Physics* **1996**, *88*, 1005–1009.

[138] Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chemical Physics Letters* **1997**, *275*, 151–160.

[139] Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615–6620.

[140] Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *Journal of Chemical Physics* **2008**, *128*, 084106.

[141] Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation,. *Journal of Chemical Physics* **2006**, *124*, 034108.

[142] Graham, D. C.; Menon, A. S.; Goerigk, L.; Grimme, S.; Radom, L. Optimization and Basis-Set Dependence of a Restricted-Open-Shell Form of B2-PLYP Double-Hybrid Density Functional Theory. *Journal Physical Chemistry A* **2009**, *113*, 9861–9873.

[143] Kozuch, S.; Gruzman, D.; Martin, J. M. L. DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction. *Journal of Physical Chemistry C* **2010**, *114*, 20801–20808.

[144] Kozuch, S.; Martin, J. M. L. DSD-PBEP86: in search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections. *Physical Chemistry Chemical Physics* **2011**, *13*, 20104–20107.

[145] Kozuch, S.; Martin, J. M. L. Spin-component-scaled double hybrids: An extensive search for the best fifth-rung functionals blending DFT and perturbation theory. *Journal of Computational Chemistry* **2013**, *34*, 2327–2344.

[146] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010**, *132*, 154104.

[147] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

[148] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *Journal of Chemical Physics* **2019**, *150*, 154122.

[149] Eyring, H. The Activated Complex in Chemical Reactions. *Journal of Chemical Physics* **2004**, *3*, 107–115.

[150] Arnaut, L.; Formosinho, S.; Burrows, H. In *Chemical Kinetics From Molecular Structure to Chemical Reactivity*; Elsevier,, Ed.; 2006; p 543.

[151] Truhlar, D. G.; Garrett, B. C.; Hipes, P. G.; Kuppermann, A. Test of variational transition state theory against accurate quantal results for a reaction with very large reaction-path curvature and a low barrier. *Journal of Chemical Physics* **1984**, *81*, 3542–3545.

[152] Eckart, C. The Penetration of a Potential Barrier by Electrons. *Physical Review* **1930**, *35*, 1303–1309.

[153] Rice, O. K.; Ramsperger, H. C. Theories of Unimolecular Gas Reactions at Low Pressures. *Journal of the American Chemical Society* **1927**, *49*, 1617–1629.

[154] Di Giacomo, F. A Short Account of RRKM Theory of Unimolecular Reactions and of Marcus Theory of Electron Transfer in a Historical Perspective. *Journal of Chemical Education* **2015**, *92*, 476–481.

[155] Georgievskii, Y.; Miller, J. A.; Burke, M. P.; Klippenstein, S. J. Reformulation and solution of the master equation for multiple-well chemical reactions. *Journal of Physical Chemistry A* **2013**, *117*, 12146–12154.

[156] Light, J. C. Phase-Space Theory of Chemical Kinetics. *Journal of Chemical Physics* **2004**, *40*, 3221–3229.

[157] Eckart, C. Some Studies Concerning Rotating Axes and Polyatomic Molecules. *Physical Review* **1935**, *47*, 552–558.

[158] Eckart, C.; Young, G. M. A principal axis transformation for non-hermitian matrices. *Bulletin of the American Mathematical Society* **1939**, *45*, 118–121.

[159] Nielsen, H. H. The Vibration-Rotation Energies of Molecules. *Reviews of Modern Physics* **1951**, *23*, 90–136.

[160] Watson, J. K. Simplification of the molecular vibration-rotation hamiltonian. *Molecular Physics* **1968**, *15*, 479–490.

[161] Christiansen, O. Selected new developments in vibrational structure theory: potential construction and vibrational wave function calculations. *Physical Chemistry Chemical Physics* **2012**, *14*, 6672–6687.

[162] Roy, T. K.; Gerber, R. B. Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications. *Physical Chemistry Chemical Physics* **2013**, *15*, 9468–9492.

[163] Amat, G.; Nielsen, H. H. Vibrational l-type doubling and l-type resonance in linear polyatomic molecules. *Journal of Molecular Spectroscopy* **1958**, *2*, 152–162.

[164] Sibaev, M.; Crittenden, D. L. PyVCI: A flexible open-source code for calculating accurate molecular infrared spectra. *Computer Physics Communications* **2016**, *203*, 290–297.

[165] Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *Journal of Chemical Physics* **2004**, *122*.

[166] Bloino, J.; Biczysko, M.; Barone, V. General Perturbative Approach for Spectroscopy, Thermodynamics, and Kinetics: Methodological Background and Benchmark Studies. *Journal of Chemical Theory and Computation* **2012**, *8*, 1015–1036.

[167] Martin, J. M. L.; Lee, T. J.; Taylor, P. R.; François, J. The anharmonic force field of ethylene, C2H4, by means of accurate ab initio calculations. *Journal of Chemical Physics* **1995**, *103*, 2589–2602.

[168] Frisch, M. J. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

[169] Piccardo, M.; Bloino, J.; Barone, V. Generalized vibrational perturbation theory for rotovibrational energies of linear, symmetric and asymmetric tops: Theory, approximations, and automated approaches to deal with medium-to-large molecular systems. *International Journal of Quantum Chemistry* **2015**, *115*, 948–982.

[170] Bloino, J.; Barone, V. A second-order perturbation theory route to vibrational averages and transition properties of molecules: General formulation and application to infrared and vibrational circular dichroism spectroscopies. *Journal of Chemical Physics* **2012**, *136*, 124108.

[171] Darling, B. T.; Dennison, D. M. The Water Vapor Molecule. *Physical Review* **1940**, *57*, 128–139.

[172] Puzzarini, C.; Bloino, J.; Tasinato, N.; Barone, V. Accuracy and Interpretability: The Devil and the Holy Grail. New Routes across Old

Boundaries in Computational Spectroscopy. *Chemical Reviews* **2019**, *119*, 8131–8191.

[173] Begue, D.; Carbonniere, P.; Pouchan, C. Calculations of Vibrational Energy Levels by Using a Hybrid ab Initio and DFT Quartic Force Field: Application to Acetonitrile. *Journal of Physical Chemistry A* **2005**, *109*, 4611–4616.

[174] Barone, V.; Biczysko, M.; Bloino, J.; Borkowska-Panek, M.; Carnimeo, I.; Panek, P. Toward anharmonic computations of vibrational spectra for large molecular systems. *International Journal of Quantum Chemistry* **2012**, *112*, 2185–2200.

[175] Puzzarini, C.; Tasinato, N.; Bloino, J.; Spada, L.; Barone, V. State-of-the-art computation of the rotational and IR spectra of the methyl-cyclopropyl cation: Hints on its detection in space. *Physical Chemistry Chemical Physics* **2019**, *21*, 3431 – 3439.

[176] Ceselin, G.; Salta, Z.; Bloino, J.; Tasinato, N.; Barone, V. Accurate Quantum Chemical Spectroscopic Characterization of Glycolic Acid: A Route Toward its Astrophysical Detection. *Journal of Physical Chemistry A* **2022**, *126*, 2373–2387.

[177] Dovesi, R.; Causà, M.; Angonoa, G. Exact-exchange Hartree-Fock calculations for periodic systems. V. Ground-state properties of silicon. *Physical Review B* **1981**, *24*, 4177–4183.

[178] Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. CRYSTAL14: A program for the ab initio investigation of crystalline solids. *International Journal of Quantum Chemistry* **2014**, *114*, 1287–1317.

[179] Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirt-

man, B. Quantum-mechanical condensed matter simulations with CRYSTAL. *WIREs Computational Molecular Science* **2018**, *8*, e1360.

[180] Erba, A.; Desmarais, J. K.; Casassa, S.; Civalleri, B.; Donà, L.; Bush, I. J.; Searle, B.; Maschio, L.; Edith-Daga, L.; Cossard, A.; Ribaldone, C.; Ascrizzi, E.; Marana, N. L.; Flament, J.-P.; Kirtman, B. CRYSTAL23: A Program for Computational Solid State Physics and Chemistry. *Journal of Chemical Theory and Computation* **2022**,

[181] Pacchioni, G.; Bagus, P. S.; Parmigiani, F. In *Cluster Models for Surface and Bulk Phenomena*; Pacchioni, G., Bagus, P. S., Parmigiani, F., Eds.; NATO Science Series B:; 1992; p 712.

[182] Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chemical Reviews* **2015**, *115*, 5678–5796.

[183] Hahn, T.; of Crystallography, I. U. *International tables for crystallography. Volume A, Space-group symmetry*; D. Reidel Dordrecht, Holland: Dordrecht, Holland, 1983.

[184] Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* **2020**, *71*, 361–390.

[185] Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 225–248.

[186] Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chemical Reviews* **2016**, *116*, 7898–7936.

[187] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.

[188] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

[189] Berendsen, H.; Postma, J.; van Gunsteren, W.; Hermans, J. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984**, *81*, 3684–3690.

[190] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **2007**, *126*, 014101.

[191] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Jr., K. M. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.

[192] III, C. L. B.; Brooks, B. R.; Bruccoleri, R. E.; Cerutti, L. F.; Doniach, S. R.; Swaminathan, D. C.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1615.

[193] Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *Journal of Computational Chemistry* **2004**, *25*, 1656–1676.

[194] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.

[195] Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *Journal of Computational Chemistry* **1996**, *17*, 520–552.

[196] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; III, W. A. G.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.

[197] Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.

[198] Huang, J.; Ponder, J. W. Do Drude Polarizable Force Fields Reproduce Both the Thermodynamics and Kinetics of Protein Folding? *Journal of Chemical Theory and Computation* **2011**, *7*, 3872–3882.

[199] van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; III, W. A. G. ReaxFF: A Reactive Force Field for Hydrocarbons. *Journal of Physical Chemistry A* **2001**, *105*, 9396–9409.

[200] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

[201] Huang, J.; Shen, R.; Lee, J.; He, L.; Wan, S.; Sundd, A.; Yu, H.; Lemberg, M. K.; III, C. L. B.; Swope, W. C.; Levitt, M.; Tajkhorshid, E. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71–73.

[202] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31*, 671–690.

[203] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983**, *79*, 926–935.

[204] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water: 2. *Journal of Chemical Physics* **1984**, *81*, 3684–3690.

[205] Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A potential model for the study of ices and amorphous water: TIP4P/Ice. *Journal of Chemical Physics* **2005**, *122*, 234511.

[206] Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G. Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *Journal of Physical Chemistry B* **2004**, *108*, 1255–1266.

[207] Heinz, H.; Lin, T.-J.; Kishore Mishra, R.; Emami, F. S. Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir* **2013**, *29*, 1754–1765.

[208] Stewart, J. J. P.; Munday, M. R.; Todorov, I. T. The INTERFACE Force Field: Derivation and Application to the Simulation of Liquid/Aluminum Interfaces. *Journal of Physical Chemistry C* **2013**, *117*, 16759–16770.

[209] Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **1979**, *28*, 100–108.

[210] Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD Proceedings* **1996**, 226–231.

[211] Schubert, E.; Sander, J.; Ester, M.; Kriegel, H.-P.; Xu, X. DBSCAN

Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems (TODS)* **2017**, *42*, 19:1–19:21.

[212] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

[213] Schrodinger, L. The PyMOL Molecular Graphics System. **2010**, Version 1.3r1.

[214] Ball, J. A.; Gottlieb, C. A.; Lilley, A. E.; Radford, H. E. Detection of Methyl Alcohol in Sagittarius. *Astrophysical Journal Lett.* **1970**, *162*, L203–L210.

[215] Zuckerman, B.; Ball, J. A.; A., G. C. Microwave detection of interstellar formic acid. *Astrophysical Journal* **1971**, *163*, L41–L54.

[216] Codella, C, et al. Seeds of Life in Space (SOLIS): II. Formamide in protostellar shocks: Evidence for gas-phase formation. *Astronomy & Astrophysics* **2017**, *605*, L3.

[217] Quan, D.; Herbst, E.; Corby, J. F.; Durr, A.; Hassel, G. Chemical simulations of prebiotic molecules: Interstellar ethanimine isomers. *Astrophysical Journal* **2016**, *824*, 129.

[218] Woon, D. E. Pathways to Glycine and Other Amino Acids in Ultraviolet-irradiated Astrophysical Ices Determined via Quantum Chemical Modeling. *Astrophysical Journal* **2002**, *571*, L177–L180.

[219] Elsila, J. E.; Dworkin, J. P.; Bernstein, M. P.; Martin, M. P.; Sandford, S. A. Mechanisms of amino acid formation in interstellar ice analogs. *Astrophysical Journal* **2007**, *660*, 911–918.

[220] Loomis, R. A.; Zaleski, D. P.; Steber, A. L.; Neill, J. L.; Muckle, M. T.; Harris, B. J.; Hollis, J. M.; Jewell, P. R.; Lattanzi, V.; Lovas, F. J.; Martinez, O.; McCarthy, M. C.; Remijan, A. J.; Pate, B. H.; Corby, J. F. The detection of interstellar ethanimine (CH_3CHNH) from observations

taken during the GBT PRIMOS survey. *Astrophysical Journal Letters* **2013**, *765*, L9.

[221] Balucani, N.; Leonori, F.; Petrucci, R.; Stazi, M.; Skouteris, D.; Rosi, M.; Casavecchia, P. Formation of nitriles and imines in the atmosphere of Titan: combined crossed-beam and theoretical studies on the reaction dynamics of excited nitrogen atoms N(^{2}D) with ethane. *Faraday Discussions* **2010**, *147*, 189–216.

[222] Clarke, D.; Ferris, J. P. Chemical evolution on Titan: Comparisons to the prebiotic Earth. *Origins of Life and Evolution of Biospheres* **1997**, *27*, 225–248.

[223] Raulin, F., McKay, C., Lunine, J., Owen, T., Eds. *Titan from Cassini-Huygens*; Dordrecht, Springer, 2009.

[224] Salta, Z.; Tasinato, N.; Lupi, J.; Boussessi, R.; Balbi, A.; Puzzarini, C.; Barone, V. Exploring the Maze of $C_2N_2H_5$ Radicals and Their Fragments in the Interstellar Medium with the Help of Quantum-Chemical Computations. *ACS Earth Space Chemistry* **2020**, *4*, 774–782.

[225] Barone, V.; Latouche, C.; Skouteris, D.; Vazart, F.; Balucani, N.; Ceccarelli, C.; Lefloch, B. Gas-phase formation of the prebiotic molecule formamide: Insights from new quantum computations. *Monthly Notices of the Royal Astronomical Society* **2015**, *453*, L31–L35.

[226] Skouteris, D.; Balucani, N.; Ceccarelli, C.; Vazart, F.; Puzzarini, C.; Barone, V.; Codella, C.; Lefloch, B. The Genealogical Tree of Ethanol: Gas-phase Formation of Glycolaldehyde, Acetic Acid, and Formic Acid. *Astrophysical Journal* **2018**, *854*, 135.

[227] Puzzarini, C.; Bloino, J.; Tasinato, N.; Barone, V. Accuracy and Interpretability: The Devil and the Holy Grail. New Routes across Old Boundaries in Computational Spectroscopy. *Chemical Reviews* **2019**, *119*, 8131–8191.

[228] Puzzarini, C.; Barone, V. The challenging playground of astrochemistry: an integrated rotational spectroscopy – quantum chemistry strategy. *Physical Chemistry Chemical Physics* **2020**, *22*, 6507–6523.

[229] Skouteris, D.; Vazart, F.; Ceccarelli, C.; Balucani, N.; Puzzarini, C.; Barone, V. New quantum chemical computations of formamide deuteration support gas-phase formation of this prebiotic molecule. *Monthly Notices of the Royal Astronomical Society* **2017**, *468*, L1–L5.

[230] Melli, A.; Melosso, M.; Tasinato, N.; Bosi, G.; Spada, L.; Bloino, J.; Mendolicchio, M.; Dore, L.; Barone, V.; Puzzarini, C. Rotational and Infrared Spectroscopy of Ethanimine: A Route toward Its Astrophysical and Planetary Detection. *Astrophysical Journal* **2018**, *855*, 13.

[231] Suzuki, T.; Ohishi, M.; Hirota, T.; Saito, M.; Majumdar, L.; Wakelam, V. Survey observations of a possible glycine precursor, methanimine (CH_2NH). *Astrophysical Journal* **2016**, *825*, 79.

[232] Balucani, N.; Skouteris, D.; Ceccarelli, C.; Codella, C.; Falcinelli, S.; Rosi, M. A theoretical investigation of the reaction between the amidogen, NH, and the ethyl, C2H5, radicals: a possible gas-phase formation route of interstellar and planetary ethanimine. *Molecular Astrophysics* **2018**, *13*, 30 – 37.

[233] Singh, K. K.; Shivani,; Tandan, P.; Misra, A. A quantum chemical study on the formation of ethanimine (CH3CHNH) in the interstellar ice. *Astrophysics and Space Science* **2018**, *363*, 213.

[234] Stanton, J. F.; Gauss, J.; Cheng, L.; Harding, M. E.; Matthews, D. A.; Szalay, P. G. CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package. With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch,

K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

[235] Kállay, M. et al. MRCC, a quantum chemical program suite. 2018; For the current version, see http://www.mrcc.hu.

[236] Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *Journal of Chemical Physics* **2008**, *128*, 114111.

[237] Kendall, R. A.; Dunning, T. H. J.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *Journal of Chemical Physics* **1992**, *96*, 6796.

[238] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *Journal of Chemical Physics* **1989**, *90*, 1007–1023.

[239] Chen, J.; Zheng, Y.; Melli, A.; Spada, L.; Lu, T.; Feng, G.; Gou, Q.; Barone, V.; Puzzarini, C. Theory meets experiment for elucidating the structure and stability of non-covalent complexes: water–amine interaction as a proof of concept. *Physical Chemistry Chemical Physics* **2020**, *22*, 5024–5032.

[240] Wang, J.; Spada, L.; Chen, J.; Gao, S.; Alessandrini, S.; Feng, G.; Puzzarini, C.; Gou, Q.; Grabow, J.-U.; Barone, V. The Unexplored World of Cycloalkene–Water Complexes: Primary and Assisting Interactions

Unraveled by Experimental and Computational Spectroscopy. *Angewandte Chemie International Edition* **2019**, *58*, 13935–13941.

[241] Boussessi, R.; Geselin, G.; Tasinato, N.; Barone, V. DFT meets the segmented polarization consistent basis sets: Performances in the computation of molecular structures, rotational and vibrational spectroscopic properties. *Journal of Molecular Structure* **2020**, *1208*, 127886.

[242] Spada, L.; Tasinato, N.; Vazart, F.; Barone, V.; Caminati, W.; Puzzarini, C. Noncovalent Interactions and Internal Dynamics in Pyridine–Ammonia: A Combined Quantum-Chemical and Microwave Spectroscopy Study. *Chemistry A European Journal* **2017**, *23*, 4876–4883.

[243] Spada, L.; Tasinato, N.; Bosi, G.; Vazart, F.; Barone, V.; Puzzarini, C. On the competition between weak OH$\cdots$F and CH$\cdots$F hydrogen bonds, in cooperation with CH$\cdots$O contacts, in the difluoromethane – tert-butyl alcohol cluster. *Journal of Molecular Spectroscopy* **2017**, *337*, 90–95.

[244] Lupi, J.; Puzzarini, C.; Cavallotti, C.; Barone, V. State-of-the-art quantum chemistry meets variable reaction coordinate transition state theory to solve the puzzling case of the $H_2S + Cl$ system. *Journal of Chemical Theory and Computation* **2020**, *16*, 5090–5104.

[245] Puzzarini, C.; Salta, Z.; Tasinato, N.; Lupi, J.; Cavallotti, C.; Barone, V. A twist on the reaction of the CN radical with methylamine in the interstellar medium: new hints from a state-of-the-art quantum-chemical study. *Monthly Notices of the Royal Astronomical Society* **2020**, *496*, 4298–4310.

[246] Rajvanshi, J. S.; Baluja, K. L. Electron-impact study of the NH radical using the *R*-matrix method. *Physical Review A* **2010**, *82*, 062710.

[247] Barone, V. Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation. *Journal of Chemical Physics* **2004**, *120*, 3059–3065.

[248] Purvis, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *Journal of Chemical Physics* **1982**, *76*, 1910–1918.

[249] Lee, T. J.; Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *International Journal of Quantum Chemistry* **1989**, *36*, 199–207.

[250] Feller, D. The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water. *Journal of Chemical Physics* **1993**, *98*, 7059–7071.

[251] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *Journal of Chemical Physics 106*, 9639–9646.

[252] Sellers, H.; Pulay, P. The adiabatic correction to molecular potential surfaces in the SCF approximation. *Chemical Physics Letters* **1984**, *103*, 463–465.

[253] Handy, N. C.; Yamaguchi, Y.; Schaefer, H. F. The diagonal correction to the Born–Oppenheimer approximation: Its effect on the singlet–triplet splitting of CH_2 and other molecular effects. *Journal of Chemical Physics* **1986**, *84*, 4481.

[254] Handy, N. C.; Lee, A. M. The adiabatic approximation. *Chemical Physics Letters* **1996**, *252*, 425–430.

[255] Kutzelnigg, W. The adiabatic approximation I. The physical background of the Born-Handy ansatz. *Molecular Physics* **1997**, *90*, 909–916.

[256] Cowan, R. D.; Griffin, M. Approximate relativistic corrections to atomic radial wave functions. *Journal of the Optical Society of America* **1976**, *66*, 1010–1014.

[257] Martin, R. L. All-electron relativistic calculations on silver hydride. An investigation of the Cowan-Griffin operator in a molecular species. *Journal of Physical Chemistry* **1983**, *87*, 750–754.

[258] Puzzarini, C.; Barone, V. Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil. *Physical Chemistry Chemical Physics* **2011**, *13*, 7189–7197.

[259] Puzzarini, C.; Biczysko, M.; Barone, V.; Peña, I.; Cabezas, C.; Alonso, J. L. Accurate molecular structure and spectroscopic properties of nucleobases: a combined computational–microwave investigation of 2-thiouracil as a case study. *Physical Chemistry Chemical Physics* **2013**, *15*, 16965–16975.

[260] Puzzarini, C.; Biczysko, M.; Barone, V.; Largo, L.; Peña, I.; Cabezas, C.; Alonso, J. L. Accurate Characterization of the Peptide Linkage in the Gas Phase: A Joint Quantum-Chemical and Rotational Spectroscopy Study of the Glycine Dipeptide Analogue. *Journal of Physical Chemistry Letters* **2014**, *5*, 534–540.

[261] Pechukas, P.; Light, J. C. On detailed balancing and statistical theories of chemical kinetics. *Journal of Chemical Physics* **1965**, *42*, 3281–3291.

[262] Chesnavich, W. J. Multiple transition states in unimolecular reactions. *Journal of Chemical Physics* **1986**, *84*, 2615–2619.

[263] Vazart, F.; Latouche, C.; Skouteris, D.; Balucani, N.; Barone, V. Cyanomethanimine Isomers in Cold Interstellar Clouds: Insights from Electronic Structure and Kinetic Calculations. *Astrophysical Journal* **2015**, *810*, 111.

[264] Skouteris, D.; Balucani, N.; Ceccarelli, C.; Faginas Lago, N.; Codella, C.; Falcinelli, S.; Rosi, M. Interstellar dimethyl ether gas-phase formation: a quantum chemistry and kinetics study. *Monthly Notices of the Royal Astronomical Society: Letters* **2018**, *482*, 3567–3575.

[265] Watanabe, N.; Kouchi, A. Ice surface reactions: A key to chemical evolution in space. *Progress in Surface Science* **2008**, *83*, 439–489.

[266] Hama, T.; Watanabe, N. Surface Processes on Interstellar Amorphous Solid Water: Adsorption, Diffusion, Tunneling Reactions, and Nuclear-Spin Conversion. *Chemical Reviews* **2013**, *113*, 8783–8839.

[267] Boogert, A. A.; Gerakines, P. A.; Whittet, D. C. Observations of the Icy Universe. *Annual Review of Astronomy and Astrophysics* **2015**, *53*, 541–581.

[268] Cuppen, H. M.; Walsh, C.; Lamberts, T.; Semenov, D.; Garrod, R. T.; Penteado, E. M.; Ioppolo, S. Grain Surface Models and Data for Astrochemistry. *Space Science Reviews* **2017**, *212*, 1–58.

[269] Tasinato, N.; Ceselin, G.; Stoppa, P.; Pietropolli Charmet, A.; Giorgianni, S. A Bit of Sugar on TiO_2: Quantum Chemical Insights on the Interfacial Interaction of Glycolaldehyde over Titanium Dioxide. *Journal of Physical Chemistry C* **2018**, *122*, 6041–6051.

[270] Rimola, A.; Ferrero, S.; Germain, A.; Corno, M.; Ugliengo, P. Computational Surface Modelling of Ices and Minerals of Interstellar Interest—Insights and Perspectives. *Minerals* **2021**, *11*.

[271] Barone, V.; Lupi, J.; Salta, Z.; Tasinato, N. Development and Validation of a Parameter-Free Model Chemistry for the Computation of Reliable Reaction Rates. *Journal of Chemical Theory and Computation* **2021**, *17*, 4913–4928.

[272] Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. Structure and Stability of Water Clusters $(H_2O)_n$, n = 8—20: An Ab Initio Investigation. *Journal of Physical Chemistry A* **2001**, *105*, 10525–10537.

[273] Rimola, A.; Sodupe, M.; Ugliengo, P. Deep-space glycine formation via Strecker-type reactions activated by ice water dust mantles. A computa-

tional approach. *Physical Chemistry Chemical Physics* **2010**, *12*, 5285–5294.

[274] Rimola, A.; Skouteris, D.; Balucani, N.; Ceccarelli, C.; Enrique-Romero, J.; Taquet, V.; Ugliengo, P. Can Formamide Be Formed on Interstellar Ice? An Atomistic Perspective. *ACS Earth Space Chemistry* **2018**, *2*, 720–734.

[275] Rimola, A.; Sodupe, M.; Ugliengo, P. Computational study of interstellar glycine formation occuring at radical surfaces of water-ice dust particles. *Astrophysical Journal* **2012**, *754*, 24.

[276] Shimonishi, T.; Nakatani, N.; Furuya, K.; Hama, T. Adsorption Energies of Carbon, Nitrogen, and Oxygen Atoms on the Low-temperature Amorphous Water Ice: A Systematic Estimation from Quantum Chemistry Calculations. *Astrophysical Journal* **2018**, *855*, 27.

[277] Zamirri, L.; Casassa, S.; Rimola, A.; Segado-Centellas, M.; Ceccarelli, C.; Ugliengo, P. IR spectral fingerprint of carbon monoxide in interstellar water–ice models. *Monthly Notices of the Royal Astronomical Society* **2018**, *480*, 1427–1444.

[278] Sameera, W. M. C.; Senevirathne, B.; Andersson, S.; Maseras, F.; Nyman, G. ONIOM(QM:AMOEBA09) Study on Binding Energies and Binding Preference of OH, HCO, and CH3 Radicals on Hexagonal Water Ice (Ih). *Journal of Physical Chemistry C* **2017**, *121*, 15223–15232.

[279] Sameera, W. M. C.; Senevirathne, B.; Andersson, S.; Al-lbadi, M.; Hidaka, H.; Kouchi, A.; Nyman, G.; Watanabe, N. CH_3O Radical Binding on Hexagonal Water Ice and Amorphous Solid Water. *Journal of Physical Chemistry A* **2021**, *125*, 387–393.

[280] Duflot, D.; Toubin, C.; Monnerville, M. Theoretical Determination of Binding Energies of Small Molecules on Interstellar Ice Surfaces. *Frontiers in Astronomy and Space Sciences* **2021**, *8*, 24.

[281] Enrique-Romero, J.; Rimola, A.; Ceccarelli, C.; Ugliengo, P.; Balucani, N.; Skouteris, D. Reactivity of HCO with CH_3 and NH_2 on Water Ice Surfaces. A Comprehensive Accurate Quantum Chemistry Study. *ACS Earth Space Chemistry* **2019**, *3*, 2158–2170.

[282] Puzzarini, C.; Barone, V. The challenging playground of astrochemistry: an integrated rotational spectroscopy – quantum chemistry strategy. *Physical Chemistry Chemical Physics* **2020**, *22*, 6507–6523.

[283] Alessandrini, S.; Barone, V.; Puzzarini, C. Extension of the "Cheap" Composite Approach to Noncovalent Interactions: The jun-ChS Scheme. *Journal of Chemical Theory and Computation* **2020**, *16*, 988–1006.

[284] Woon, D. E. Ab Initio Quantum Chemical Studies of Reactions in Astrophysical Ices 3. Reactions of $HOCH_2NH_2$ Formed in $H_2CO/NH_3/H_2O$ Ices. *Journal of Physical Chemistry A* **2001**, *105*, 9478–9481.

[285] Koch, D. M.; Toubin, C.; Peslherbe, G. H.; Hynes, J. T. A Theoretical Study of the Formation of the Aminoacetonitrile Precursor of Glycine on Icy Grain Mantles in the Interstellar Medium. *Journal of Physical Chemistry C* **2008**, *112*, 2972–2980.

[286] Gardebien, F.; Sevin, A. Catalytic Model Reactions for the HCN Isomerization. I. Theoretical Characterization of Some Water-Catalyzed Mechanisms. *Journal of Physical Chemistry A* **2003**, *107*, 3925–3934.

[287] Koch, D. M.; Toubin, C.; Xu, S.; Peslherbe, G. H.; Hynes, J. T. Concerted Proton-Transfer Mechanism and Solvation Effects in the HNC/HCN Isomerization on the Surface of Icy Grain Mantles in the Interstellar Medium. *Journal of Physical Chemistry C* **2007**, *111*, 15026–15033.

[288] Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **1980**, *58*, 1200–1211.

[289] Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *Journal of Chemical Physics Letters* **2016**, *7*, 2197–2203.

[290] Burns, L. A.; Mayagoitia, Á. V.; Sumpter, B. G.; Sherrill, C. D. Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *Journal of Chemical Physics* **2011**, *134*, 084107.

[291] Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *Journal of Chemical Physics* **2012**, *137*, 120901.

[292] Tasinato, N.; Grimme, S. Unveiling the non-covalent interactions of molecular homodimers by dispersion-corrected DFT calculations and collision-induced broadening of ro-vibrational transitions: Application to $(CH_2F_2)_2$ and $(SO_2)_2$. *Physical Chemistry Chemical Physics* **2015**, *17*, 5659–5669.

[293] Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking Density Functional Methods against the S66 and S66x8 Datasets for Non-Covalent Interactions. *Chem. Phys. Chem.* **2011**, *12*, 3421–3433.

[294] Delle Piane, M.; Corno, M.; Ugliengo, P. Does Dispersion Dominate over H-Bonds in Drug–Surface Interactions? The Case of Silica-Based Materials As Excipients and Drug-Delivery Agents. *Journal of Chemical Theory and Computation* **2013**, *9*, 2404—2415.

[295] Tasinato, N.; Moro, D.; Stoppa, P.; Pietropolli Charmet, A.; Toninello, P.; Giorgianni, S. Adsorption of F_2CCFCl on TiO_2 nano-powder: Structures, energetics and vibrational properties from DRIFT spectroscopy and periodic quantum chemical calculations. *Applied Surface Science* **2015**, *353*, 986–994.

[296] Puzzarini, C.; Barone, V. Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil. *Physical Chemistry Chemical Physics* **2011**, *13*, 7189–7197.

[297] Puzzarini, C.; Biczysko, M.; Barone, V.; Peña, I.; Cabezas, C.; Alonso, J. L. Accurate molecular structure and spectroscopic properties of nucleobases: a combined computational–microwave investigation of 2-thiouracil as a case study. *Physical Chemistry Chemical Physics* **2013**, *15*, 16965.

[298] Lane, J. R. CCSDTQ Optimized Geometry of Water Dimer. *Journal of Chemical Theory and Computation* **2013**, *9*, 316–323.

[299] Vreven, T.; Morokuma, K. Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM. *Annual Reports in Computational Chemistry* **2006**, *2*, 35–51.

[300] Fernández-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. Modeling the Kinetics of Bimolecular Reactions. *Chemical Reviews* **2006**, *106*, 4518–4584.

[301] Kooij, D. M. Über die Zersetzung des gasförmigen Phosphorwasserstoffs. *Zeitschrift für Physikalische Chemie* **1893**, *12*, 155–161.

[302] Laidler, K. A. A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996). *Pure and Applied Chemistry* **1996**, *68*, 149–192.

[303] Zheng, J.; Zhao, Y.; Truhlar, D. G. The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *Journal of Chemical Theory and Computation* **2009**, *5*, 808–821.

[304] Barone, V.; Ceselin, G.; Fusé, M.; Tasinato, N. Accuracy Meets Inter-

pretability for Computational Spectroscopy by Means of Hybrid and Double-Hybrid Functionals. *Frontiers Chemistry* **2020**, *8*, 584203–1.

[305] Ceselin, G.; Barone, V.; Tasinato, N. Accurate Biomolecular Structures by the Nano-LEGO Approach: Pick the Bricks and Build Your Geometry. *Journal of Chemical Theory and Computation* **2021**, *17*, 7290–7311.

[306] Huthwelker, T.; Ammam, M.; Peter, T. The uptake of acidic gases on ice. *Chemical Reviews* **2006**, *106*, 1375–1444.

[307] Devlin, J. P.; Buch, V. Vibrational Spectroscopy and Modeling of the Surface and Subsurface of Ice and of Ice Adsorbate Interactions. *Journal of Physical ChemistryB* **1997**, *101*, 6095–6098.

[308] Livingston, F. E.; Smith, J. A.; George, S. M. General Trends for Bulk Diffusion in Ice and Surface Diffusion on Ice. *Journal of Physical Chemistry A* **2002**, *106*, 6309–6318.

[309] Sauer, J. Ab Initio Calculations for Molecule–Surface Interactions with Chemical Accuracy. *Accounts of Chemical Research* **2019**, *52*, 3502–3510.

[310] Ceccarelli, C. et al. Seeds Of Life In Space (SOLIS): The Organic Composition Diversity at 300–1000 au Scale in Solar-type Star-forming Regions*. *Astrophysical Journal* **2017**, *850*, 176.

[311] Bianchi, E.; Ceccarelli, C.; Codella, C.; Enrique-Romero, J.; Favre, C.; Lefloch, B. Astrochemistry as a Tool To Follow Protostellar Evolution: The Class I Stage. *ACS Earth and Space Chemistry* **2019**, *3*, 2659–2674.

[312] Mehringer, D. M.; Snyder, L. E.; Miao, Y.; Lovas, F. J. Detection and Confirmation of Interstellar Acetic Acid. *Astrophysical Journal Letters* **1997**, *480*, L71–L74.

[313] Wlodarczak, G.; Demaison, J. Acetic acid, CH3COOH - Astrophysical predictions. *Astronomy and Astrophysics* **1988**, *192*, 313–315.

[314] Kleimeier, N. F.; Eckhardt, A. K.; Kaiser, R. I. A Mechanistic Study on the Formation of Acetic Acid (CH3COOH) in Polar Interstellar Analog Ices Exploiting Photoionization Reflectron Time-of-flight Mass Spectrometry. *Astrophysical Journal* **2020**, *901*, 84.

[315] Sorrell, W. H. Origin of Amino Acids and Organic Sugars in Interstellar Clouds. *Astrophysical Journal* **2001**, *555*, L129–L132.

[316] Cazaux, S.; Tielens, A. G. G. M.; Ceccarelli, C.; Castets, A.; Wakelam, V.; Caux, E.; Parise, B.; Teyssier, D. The Hot Core around the Low-Mass Protostar IRAS 16293–2422: Scoundrels Rule! *Astrophysical Journal* **2003**, *593*, L51–L55.

[317] Burke, D. J.; Puletti, F.; Woods, P. M.; Viti, S.; Slater, B.; Brown, W. A. Trapping and desorption of complex organic molecules in water at 20 K. *Journal of Chemical Physics* **2015**, *143*, 164704.

[318] Gao, Q.; Leung, K. T. Thermal Evolution of Acetic Acid Nanodeposits over 123-180 K on NanCrystalline ICe and Polycrystalline Ice Studied by FTIR Reflection-Absorption Spectroscopy: Hydrogen-bonding Interactions in Acetic Acid and between Acetic Acid and Ice. *Journal of Physical Chemistry B* **2005**, *109*, 13263–13271.

[319] Allouche, A.; Bahr, S. Acetic Acid-Water Interaction in Solid Interfaces. *Journal of Physical Chemistry B* **2006**, *110*, 8640–8648.

[320] Bertin, M.; Romanzin, C.; Michaut, X.; Jeseck, P.; Fillion, J.-H. Adsorption of Organic Isomers on Water Ice Surfaces: A Study of Acetic Acid and Methyl Formate. *Journal of Physical Chemistry C* **2011**, *115*, 12920–12928.

[321] Burke, D. J.; Puletti, F.; Brown, W. A.; Woods, P. M.; Viti, S.; Slater, B. Glycolaldehyde, methyl formate and acetic acid adsorption and thermal desorption from interstellar ices. *Monthly Notices of the Royal Astronomical Society* **2014**, *447*, 1444–1451.

[322] Bahr, S.; Borodin, A.; Höfft, O.; Kempter, V.; Allouche, A.; Borget, F.; Chiavassa, T. Interaction of Acetic Acid with Solid Water. *Journal of Physical Chemistry B* **2006**, *110*, 8649–8656.

[323] Collings, M. P.; Anderson, M. A.; Chen, R.; Dever, J. W.; Viti, S.; Williams, D. A.; McCoustra, M. R. S. A laboratory survey of the thermal desorption of astrophysically relevant molecules. *Monthly Notices of the Royal Astronomical Society* **2004**, *354*, 1133–1140.

[324] Johnson, T. J.; Sams, R. L.; Profeta, L. T. M.; Akagi, S. K.; Burling, I. R.; Yokelson, R. J.; Williams, S. D. Quantitative IR Spectrum and Vibrational Assignments for Glycolaldehyde Vapor: Glycolaldehyde Measurements in Biomass Burning Plumes. *Journal of Physical Chemistry A* **2013**, *117*, 4096–4107.

[325] Jenniskens, P.; Blake, D. F. Structural transitions in amorphous water ice and astrophysical implications. *Science* **1994**, *265*, 753–756.

[326] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics* **1999**, *110*, 6158–6170.

[327] Peintinger, M. F.; Oliveira, D. V.; Bredow, T. Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *Journal of Computational Chemistry* **2012**, *34*, 451–459.

[328] Leadbetter, A. J.; Ward, R. C.; Clark, J. W.; Tucker, P. A.; Matsuo, T.; Suga, H. The equilibrium low-temperature structure of ice. *Journal of Chemical Physics* **1985**, *82*, 424–428.

[329] Boys, S.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970**, *19*, 553.

[330] Matsumoto, M.; Yagasaki, T.; Tanaka, H. GenIce: Hydrogen-Disordered Ice Generator. *Journal of Computational Chemistry* **2017**, *39*, 61–64.

[331] Matsumoto, M.; Yagasaki, T.; Tanaka, H. Novel Algorithm to Generate Hydrogen-Disordered Ice Structures. *Journal of Chemical Information and Modeling* **2021**, *61*, 2542–2546.

[332] Rosu-Finsen, A.; Davies, M. B.; Amon, A.; Wu, H.; Sella, A.; Michaelides, A.; Salzmann, C. G. Medium-density amorphous ice. *Science* **2023**, *379*, 474–478.

[333] Paiva, M. A. M.; Pilling, S.; Mendoza, E.; Galvã o, B. R. L.; Abreu, H. A. D. Glycolaldehyde formation mediated by interstellar amorphous ice: a computational study. *Monthly Notices of the Royal Astronomical Society* **2022**, *519*, 2518–2527.

[334] Germain, A.; Tinacci, L.; Pantaleone, S.; Ceccarelli, C.; Ugliengo, P. Computer Generated Realistic Interstellar Icy Grain Models: Physicochemical Properties and Interaction with NH3. *ACS Earth and Space Chemistry* **2022**, *6*, 1286–1298.

[335] González, M. A.; Abascal, J. L. F. A flexible model for water based on TIP4P/2005. *Journal of Chemical Physics* **2011**, *135*, 224516.

[336] Duboué-Dijon, E.; Laage, D. Characterization of the Local Structure in Liquid Water by Various Order Parameters. *Journal of Physical Chemistry B* **2015**, *119*, 8406–8418.

[337] Lattelais, M., et al. Differential adsorption of complex organic molecules isomers at interstellar ice surfaces. *Astronomy & Astrophysics* **2011**, *532*, A12.

[338] Mariedahl, D.; Perakis, F.; Späh, A.; Pathak, H.; Kim, K. H.; Camisasca, G.; Schlesinger, D.; Benmore, C.; Pettersson, L. G. M.; Nilsson, A.; Amann-Winkel, K. X-ray Scattering and O–O Pair-Distribution

Functions of Amorphous Ices. *Journal of Physical Chemistry B* **2018**, *122*, 7616–7624.

[339] Chiu, J.; Starr, F. W.; Giovambattista, N. Pressure-induced transformations in computer simulations of glassy water. *Journal of Chemical Physics* **2013**, *139*, 184504.

[340] Min, M.,; Waters, L. B. F. M.,; de Koter, A.,; Hovenier, J. W.,; Keller, L. P.,; Markwick-Kemper, F., The shape and composition of interstellar silicate grains. *Astronomy & Astrophysics* **2007**, *462*, 667–676.

[341] Potapov, A.; Jäger, C.; Henning, T. Ice Coverage of Dust Grains in Cold Astrophysical Environments. *Physical Review Letters* **2020**, *124*, 221103.

[342] Potapov, A.; Bouwman, J.; Jäger, C.; Henning, T. Dust/ice mixing in cold regions and solid-state water in the diffuse interstellar medium. *Nature Astronomy* **2021**, *5*, 78–85.

[343] Fornasier, S., et al. Spectrophotometric properties of the nucleus of comet 67P/Churyumov-Gerasimenko from the OSIRIS instrument onboard the ROSETTA spacecraft. *Astronomy & Astrophysics* **2015**, *583*, A30.

[344] Stenzel, O. J.; Hilchenbach, M.; Merouane, S.; Paquette, J.; Varmuza, K.; Engrand, C.; Brandstätter, F.; Koeberl, C.; Ferrière, L.; Filzmoser, P.; Siljeström, S.; the COSIMA team, Similarities in element content between comet 67P/Churyumov–Gerasimenko coma dust and selected meteorite samples. *Monthly Notices of the Royal Astronomical Society* **2017**, *469*, S492–S505.

[345] Boogert, A. C. A. et al. The c2d Spitzer Spectroscopic Survey of Ices around Low-Mass Young Stellar Objects. I. H2O and the 5-8 νm Bands* **. *Astrophysical Journal* **2008**, *678*, 985–1004.

[346] Boogert, A. C. A.; Huard, T. L.; Cook, A. M.; Chiar, J. E.; Knez, C.; Decin, L.; Blake, G. A.; Tielens, A. G. G. M.; van Dishoeck, E. F. ICE

AND DUST IN THE QUIESCENT MEDIUM OF ISOLATED DENSE CORES*. *Astrophysical Journal* **2011**, *729*, 92.

[347] McClure, M. K. et al. An Ice Age JWST inventory of dense molecular cloud ices. *Nature Astronomy* **2023**, *7*, 431–443.

[348] Zhuravlev, L. The surface chemistry of amorphous silica. Zhuravlev model. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2000**, *173*, 1–38.

[349] Potapov, A.; Theulé, P.; Jäger, C.; Henning, T. Evidence of Surface Catalytic Effect on Cosmic Dust Grain Analogs: The Ammonia and Carbon Dioxide Surface Reaction. *Astrophysical Journal Letters* **2019**, *878*, L20.

[350] Belosludov, R. V.; Subbotin, O. S.; Mizuseki, H.; Rodger, P. M.; Kawazoe, Y.; Belosludov, V. R. Crystal-like low frequency phonons in the low-density amorphous and high-density amorphous ices. *Journal of Chemical Physics* **2008**, *129*, 114507.