



**HAL**  
open science

# Astrochemical reaction networks : bringing order to chaos

Lorenzo Tinacci

► **To cite this version:**

| Lorenzo Tinacci. Astrochemical reaction networks : bringing order to chaos. Astrophysics [astro-ph].  
| Université Grenoble Alpes [2020-..], 2023. English. NNT : 2023GRALY009 . tel-04126162

**HAL Id: tel-04126162**

**<https://theses.hal.science/tel-04126162v1>**

Submitted on 13 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : PHYS - Physique

Spécialité : Astrophysique et Milieux Dilués

Unité de recherche : Institut de Planetologie et d'Astrophysique de Grenoble

## **Mettre de l'ordre dans le chaos des réseaux de réactions astrochimiques**

### **Astrochemical reaction networks: bringing order to chaos**

Présentée par :

**Lorenzo TINACCI**

Direction de thèse :

**Cecilia CECCARELLI**

Astronome, Université Grenoble Alpes

Directrice de thèse

**Ugliengo PIERO**

Professeur, Università degli studi di Torino

Co-encadrant de thèse

Rapporteurs :

**ALEXANDER G. G. M. TIELENS**

Professeur, Universiteit Leiden

**CLAUDIO GRECO**

Professeur, Università degli Studi di Milano-Bicocca

Thèse soutenue publiquement le **31 janvier 2023**, devant le jury composé de :

**ALEXANDER G. G. M. TIELENS**

Professeur, Universiteit Leiden

Rapporteur

**CLAUDIO GRECO**

Professeur, Università degli Studi di Milano-Bicocca

Rapporteur

**PIERRE BECK**

Professeur des Universités, UNIVERSITE GRENOBLE ALPES

Président

**NADIA BALUCANI**

Professeur, Università degli Studi di Perugia

Examinatrice

**GUNNAR NYMAN**

Professeur, Göteborgs Universitet

Examineur







A chi mi ha accompagnato,  
a chi mi ha sopportato  
ed a chi mi ha aiutato.

This thesis has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N°811312 for the project “Astro-Chemical Origins” (ACO).



This thesis was written using  $\LaTeX$  e  $\text{BIB}\TeX$ , edited with  Overleaf and via the   $\text{TeXLive}$  2022. Almost all my figures and plot were obtained via  $\text{TikZ}$  and  $\text{PGFPlot}$  packages. I really thanks once more the  $\LaTeX$  community.



## Abstract

How molecules are formed, survive and thrive in the Interstellar Medium are the fascinating focus of astrochemistry. Astronomical observations have detected a multitude of molecules ( $\geq 270$ ) with up to 13 atoms (not considering the PAHs). On the other side, astrochemical models attempt to simulate the processes that govern the interstellar chemical evolution, providing predictions to reproduce the observations. In between, it lies the understanding of the chemistry that occurs in the very peculiar, with respect to the terrestrial, interstellar conditions.

The reaction networks are the backbone of astrochemical models and one of the most crucial elements where the chemical knowledge is concentrated. However, due to their complexity and non-linear behaviour on the astrochemical modeling, reaction networks can also be the first source of error in the simulations, if incorrect chemical data are present. During my thesis, I tried to “*bring some order*” in the commonly employed astrochemical reaction networks using computational chemistry as a tool. The goal of my thesis can be divided into two immediate objectives: (1) improving the gas-phase chemical reaction networks, identifying possible incorrect reactions, and (2) improving the grain-surface chemistry reaction networks, providing new insights of the interstellar thermal desorption process.

*Objective 1:* Present day gas-phase networks (*e.g.*, those in the KIDA and UMIST databases) contain about 8000 reactions involving more than 500 species. Of these, less than 20% have been studied in laboratories or theoretically, and not always in the range of the interstellar densities ( $10^3$ – $10^{10}$  and temperatures (10–200 K). Thus, my first objective was to “*clean*” the networks from the most obvious possible error: the presence of endothermic reactions. To this end, using quantum mechanical (QM) calculations, I first created a database of structures and properties for almost all the molecules present in the commonly used gas-phase reaction networks. Then, with these data, I performed a control of the endothermicity for each reaction present in the gas-phase reaction network generated by our group over the years, which is based on the KIDA database. I found that  $\sim 5\%$  of the reactions are endothermic not recognised as such, of which  $2/3$  are reported as barrierless. The impact of the incorrect reactions removal was studied via the modeling of typical cold and warm regions, discovering that the reaction network involving Si-bearing species is the most impacted one. After the removal of the incorrect reactions, a new gas-phase reaction network, called GRETOBAPE, was created and made publicly available to the community. In the process, I also made public a reduced network to use when only the major chemical

species need to be computed.

*Objective 2:* Two major quantities govern the thermal desorption of species from interstellar ices: the species binding energy (BE) and the pre-exponential factor ( $\nu$ ). Usually, astrochemical databases provide only a single or very few BEs and the  $\nu$  is computed from approximate formulae. The BEs and, in very few cases, the  $\nu$  are usually obtained via experiments or computational studies that consider only one or a few water molecules to simulate the ice grain mantle. However, the surfaces of the interstellar ices is likely irregular and a multitude of different adsorbing sites must exist. Thus, in order to better simulate the desorption of molecules from interstellar ices, I developed a new methodology for computing an accurate and unbiased distribution of adsorption sites and the corresponding BEs and  $\nu$ . First, a new large ice grain mantle model was obtained via the new ACO-FROST program, developed by our group and to which I substantially contributed. Second, I obtained a distribution of adsorption sites for ammonia and water, where I computed the BE and  $\nu$ . The low- and high- end of the BE distribution of water and ammonia might have a major impact in the predicted gaseous abundance of these two species. For example, it could explain the observed presence in the gas-phase of these two molecules in cold regions, such as prestellar cores and protoplanetary outer disks, where thermal desorption would not be expected to be efficient based on the previous literature single BE values. Furthermore, I studied the impact of the commonly used formulae for  $\nu$  to compute the thermal desorption rate in the Temperature-Programmed Desorption (TPD) experiments. To this end, I simulated TPD experiments using the amorphous ice used to derive the ammonia and water BE distributions, described above, and a crystalline ice model. I found that the use of the different formulae cause important differences in how the desorption process is described.

Finally, my thesis only scratched the tip of the iceberg of the astrochemical data used to simulate the chemical processes occurring in the interstellar medium. I hope that my contribution has paved the way for more systematic studies of the gas-phase reactions networks and the thermal desorption rate of interstellar molecules, to make the astrochemical model predictions more reliable.

**Keywords:** Interstellar Medium, Astrochemistry, Gas Chemistry, Surface Chemistry, Reaction Network, Computational Chemistry.

# Published and accepted articles

## Main author

1. Lorenzo Tinacci, Stefano Pantaleone, Andrea Maranzana, Nadia Balucani, Cecilia Ceccarelli, and Piero Ugliengo. Structures and properties of known and postulated interstellar cations. *The Astrophysical Journal Supplement Series*, 256(2):35, 2021. doi: [10.3847/1538-4365/ac194c](https://doi.org/10.3847/1538-4365/ac194c)
2. Lorenzo Tinacci, Aurele Germain, Stefano Pantaleone, Stefano Ferrero, Cecilia Ceccarelli, and Piero Ugliengo. Theoretical distribution of the ammonia binding energy at interstellar icy grains: A new computational framework. *ACS Earth and Space Chemistry*, 6(6):1514–1526, 2022. doi: [10.1021/acsearthspacechem.2c00040](https://doi.org/10.1021/acsearthspacechem.2c00040)
3. Lorenzo Tinacci, Simon Ferrada-Chamorro, Cecilia Ceccarelli, Stefano Pantaleone, Daniela Ascenzi, Andrea Maranzana, Nadia Balucani, and Piero Ugliengo. The gretobape gas-phase reaction network: the importance of being exothermic. *The Astrophysical Journal Supplement Series*, Accepted for publication and currently under review
4. Lorenzo Tinacci, Aurele Germain, Stefano Pantaleone, Cecilia Ceccarelli, Nadia Balucani, and Piero Ugliengo. Theoretical water binding energy distribution and snowline in protoplanetary disks. *The Astrophysical Journal*, Submitted

## Co-author

1. Aurele Germain, Lorenzo Tinacci, Stefano Pantaleone, Cecilia Ceccarelli, and Piero Ugliengo. Computer generated realistic interstellar icy grain models: Physicochemical properties and interaction with nh<sub>3</sub>. *ACS Earth and Space Chemistry*, 6(5):1286–1298, 2022. doi: [10.1021/acsearthspacechem.2c00004](https://doi.org/10.1021/acsearthspacechem.2c00004)

# Contents

<b>I</b>	<b>Introduction and Methodology</b>	<b>9</b>
<b>1</b>	<b>Scientific context and objectives of the thesis</b>	<b>10</b>
1.1	The chemical history of the interstellar matter . . . . .	12
1.1.1	The journey of a dark cloud to a Solar-type planetary system . . . . .	12
1.1.2	Chemical evolution and inheritance during the formation of a Solar-type planetary system . . . . .	15
1.2	Understanding observations with astrochemical models . . . . .	17
1.2.1	What is an astrochemical model? . . . . .	19
1.2.2	Reaction network: the core of astrochemical models . . . . .	23
1.3	Problems and challenges for astrochemical reaction networks at the beginning of this thesis . . . . .	27
1.3.1	Gas-phase reaction network . . . . .	29
1.3.2	Grain-phase reaction network . . . . .	30
1.4	Thesis context and objectives . . . . .	32
1.4.1	Context: The ITN Astro-ChemicalOrigin (ACO) project . . . . .	32
1.4.2	Objectives of the thesis . . . . .	34
1.4.3	Thesis outline . . . . .	36
<b>2</b>	<b>Methods</b>	<b>37</b>
2.1	Quantum mechanic in chemistry . . . . .	38
2.1.1	The Hamiltonian, principles and approximations in molecular system . . . . .	38
2.1.2	The Hartree-Fock method . . . . .	42
2.1.3	Post Hartree-Fock methods . . . . .	44
2.1.4	DLPNO-CCSD(T) method . . . . .	45
2.1.5	ONIOM method . . . . .	47
2.2	Graph theory for molecules and reaction network . . . . .	48

	7
2.2.1	Molecules . . . . . 51
2.2.2	Reaction Network . . . . . 56
<b>II</b>	<b>Astrochemical Gas-Phase Network 58</b>
<b>3</b>	<b>Introduction 59</b>
<b>4</b>	<b>Structures and properties of known and postulated interstellar cations 62</b>
<b>5</b>	<b>The GRETOBAPE gas-phase reaction network: the importance of being exothermic 73</b>
<b>III</b>	<b>Thermal Desorption of Interstellar Molecules 96</b>
<b>6</b>	<b>Introduction 97</b>
<b>7</b>	<b>Computer generated realistic interstellar icy grain models: physicochemical properties and interaction with ammonia 101</b>
<b>8</b>	<b>Theoretical distribution of the ammonia binding energy at interstellar icy grains: a new computational framework 115</b>
8.1	ACS Earth & Space Chemistry - Issue Cover . . . . . 129
<b>9</b>	<b>Theoretical water binding energy distribution and the water snowline 130</b>
<b>10</b>	<b>The pre-exponential factor importance on desorption rate for species on water surface: a computational study 152</b>
<b>IV</b>	<b>Conclusions &amp; Perspectives 168</b>
<b>11</b>	<b>Conclusions and future perspectives 169</b>
11.1	Summary and conclusions . . . . . 170
11.1.1	Astrochemical gas-phase network . . . . . 170
11.1.2	Thermal desorption of astrochemical molecules . . . . . 171
11.2	Future perspectives . . . . . 172
11.2.1	Astrochemical gas-phase networks, how improving it more? . . . . . 172

11.2.2 Computational studies of thermal desorption: what next? . . . . .	174
11.3 Final remarks . . . . .	175
<b>V Résumé en Français de la Thèse et des Résultats</b>	<b>176</b>
12 Résumé de thèse	177
13 Résumée de résultats Part II	180
14 Résumée de résultats Part III	183
<b>VI Appendix</b>	<b>187</b>
A ACO-VR project: The journey of water to Earth	188
<b>Bibliography</b>	<b>205</b>

# Part I

## Introduction and Methodology



# Scientific context and objectives of the thesis

## Contents

---

<b>1.1</b>	<b>The chemical history of the interstellar matter . .</b>	<b>12</b>
1.1.1	The journey of a dark cloud to a Solar-type planetary system . . . . .	12
1.1.2	Chemical evolution and inheritance during the formation of a Solar-type planetary system . . . . .	15
<b>1.2</b>	<b>Understanding observations with astrochemical models . . . . .</b>	<b>17</b>
1.2.1	What is an astrochemical model? . . . . .	19
1.2.2	Reaction network: the core of astrochemical models	23
<b>1.3</b>	<b>Problems and challenges for astrochemical reaction networks at the beginning of this thesis . . .</b>	<b>27</b>
1.3.1	Gas-phase reaction network . . . . .	29
1.3.2	Grain-phase reaction network . . . . .	30
<b>1.4</b>	<b>Thesis context and objectives . . . . .</b>	<b>32</b>
1.4.1	Context: The ITN Astro-ChemicalOrigin (ACO) project . . . . .	32
1.4.2	Objectives of the thesis . . . . .	34
1.4.3	Thesis outline . . . . .	36

---

The word “*chemistry*” is often associated to scientists working in a laboratory with colored potions, colours linked to Earth’s molecules. However chemistry is ubiquitous and present in all the parts of the universe, also in what it is normally thought to be “*empty*” (*e.g.*, the black zones in Figure 1.1): the interstellar medium (ISM), *i.e.*, the diffuse material between stars. The evolution of the physical conditions in the Universe is linked to the evolution of its chemistry, from the formation of elements to the formation of molecules and more complex structures. In this chapter, I will introduce the astrochemical reaction network, the importance that such object cover in the astrochemical field and how can be improved with computational chemistry techniques.

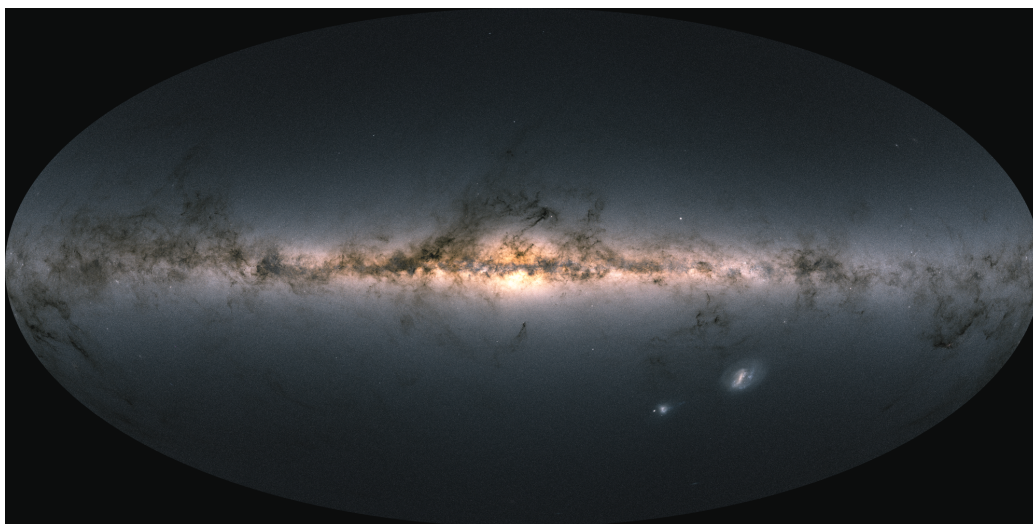


Figure 1.1: Gaia’s all-sky view of the Milky Way (and the two neighboring galaxies, the Magellanic clouds) based on measurements of  $\sim 109$  stars. The map shows the total brightness and color of stars observed by the ESA satellite in each portion of the sky between July 2014 and May 2016. Brighter regions indicate denser concentrations of bright stars, while darker regions correspond to foreground clouds of the interstellar gas and dust, which absorb the light of background stars. (Credit: ESA/Gaia/DPAC)

## 1.1 The chemical history of the interstellar matter

The starting point of time and space, according to the current cosmological model of the Universe ( $\Lambda$ CDM), is the Big Bang. In this event, only hydrogen (H), deuterium (D), helium (He), and a small quantity of lithium (Li) are produced, all the other heavier elements of the periodic table are synthesized in stars over the long history of the Universe ( $\sim 14$  billions years). The creation of new and heavier atoms from lighter elements occurs thanks to thermonuclear reactions inside stars, a phenomenon defined as “*nucleosynthesis*”. The newly created elements, which are the bricks for chemistry, represent only 2% in mass of the ordinary matter. With the life cycles of the stars, heavier atoms are synthesized and ejected or inserted into the ISM via supernova explosion and other mass-loss processes. The star dance between life and death, driven by gravity and thermonuclear reactions, can be schematically represented with all its phases in the ISM cycle, *e.g.*, Fig. 1.2. The ISM, defined as the material which fills the space between the stars, represents by almost 10% in mass of the ordinary matter (baryons) in the Milky Way. It is one of the essential actors in the star formation and evolution (*e.g.*, Tielens, 2005). The ISM is made up  $\sim 99\%$  of gas (ions, neutral atoms and molecules) and  $\sim 1\%$  of dust (whose refractory component is composed by silicates and carbonaceous grains, with sizes from 0.005 to 10  $\mu\text{m}$ ). Understanding and characterizing the ISM and its chemical evolution can help to reconstruct the early stages of our Solar System and, maybe, the origin of life on Earth (Caselli and Ceccarelli, 2012). About the matter transformation through the ISM cycle, an important and interesting aspect is played by its molecular evolution, which is also powerful in tracking all the stages.

### 1.1.1 The journey of a dark cloud to a Solar-type planetary system

Solar-type stars are formed in dense molecular clouds, where the medium is characterized by moderately high density ( $\geq 200 \text{ cm}^{-3}$ ), relatively low temperature (8–15 K) and their size varies from a few to 15 parsecs. Molecular clouds are sometimes also called dark clouds because they can be identified as dark patches that obscure the light from background stars (see Fig. 1.1). Please note that the adjective “*molecular*” means that hydrogen is partially molecular in these regions while other elements may be atomic. This occurs in the inner part of molecular clouds, which is protected by the UV radiation and, consequently, the photo-dissociation of the  $\text{H}_2$  becomes inefficient and

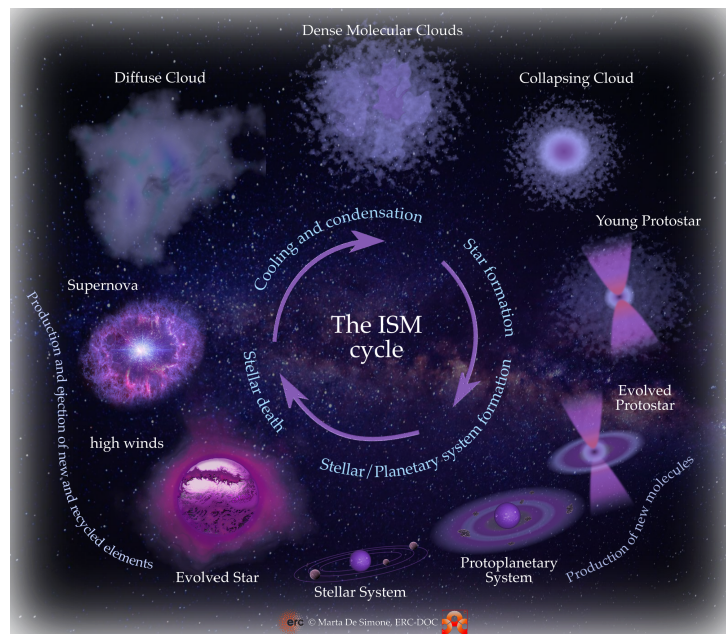


Figure 1.2: Artistic representation of the life cycle of galactic matter. (Credit: De Simone, 2022)

high abundances of  $\text{H}_2$  are preserved (Tielens, 2005; Yamamoto, 2017).

The formation process of a Solar-type planetary system can be schematically divided in five, physically and chemically different, stages as reported in Figure 1.3.

**Prestellar cores** Inside molecular clouds, small denser cores are created due to the slow accretion of mass toward the center, caused by the gravitational force. The cores centers end up having large densities ( $10^5$ – $10^6 \text{ cm}^{-3}$ ) and cold temperature (8–12 K). These structures are believed to be the first stage of star formation, for this reason are called “*prestellar cores*” (Beichman et al., 1986; Bergin and Tafalla, 2007).

**Class 0 and I protostars** After about  $10^5$  yr, the gravitational collapse of the large-scale prestellar core matter starts to dominate over the internal (thermal, magnetic and turbulent) forces that contrast it. This in-falling envelope material feeds the central object, and the system is now defined “*protostar*”. The gravitational energy of the collapse is converted into radiation close to the center, and it heats up the surrounding envelope. The innermost region is heated up to 100 K and it is called “*hot-corino*” (Ceccarelli et al., 2000; Ceccarelli, 2004). At this point, to remove the excess

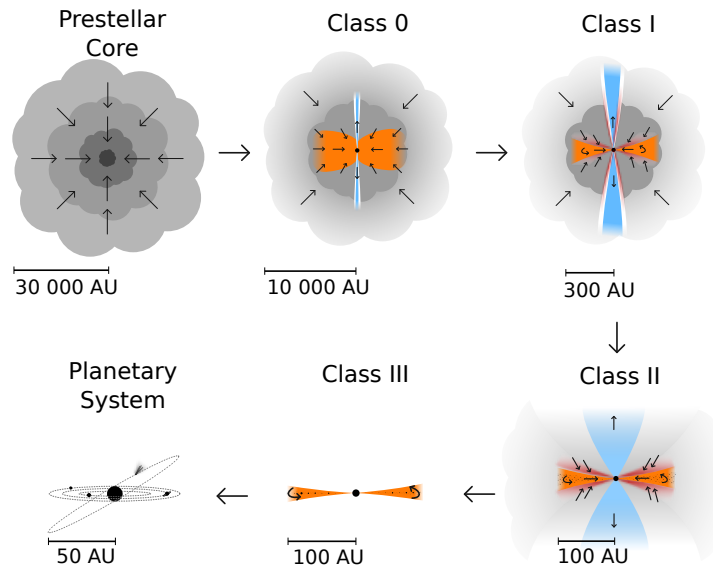


Figure 1.3: Schematic representation of a low-mass star evolution. (Credit: Persson (2014), adapted from Adams et al. (1987); Andre et al. (2000)).

angular momentum, a bipolar outflow emerges from the protostar. During the Class 0 protostar phase, the central object is still obscured by the thick and collapsing envelope. After this phase, the protostar (*i.e.*, central object + disk + envelope) enters the Class I protostellar stage, when the envelope becomes thinner due to the growing protostar and the cavities excavated by the outflow pressure (Garufi et al., 2021).

**Class II and III protostars** The Class II protostar phase is observed after  $\sim 10^6$  years. The majority of the envelope has fallen into the protostar or has been blown away from it, revealing the disk. In this phase, the dust coagulation is efficient, the planetesimals start to form and the typical shape of a protoplanetary disk is now visible (Burrows et al., 1996; Henning and Semenov, 2013). As illustrated in Figure 1.4, the matter in protoplanetary disk experiences different physical conditions depending on the distance from the star and the mid-plane. The protoplanetary disk will ultimately become a debris disk (Class III protostar), when almost all the original envelope gas is injected into the star, gaseous planets, or driven away as a result of the stellar radiation winds. The debris disk will evolve over millions of years, forming planets by collisions with planetesimals and finally creating a planetary system.

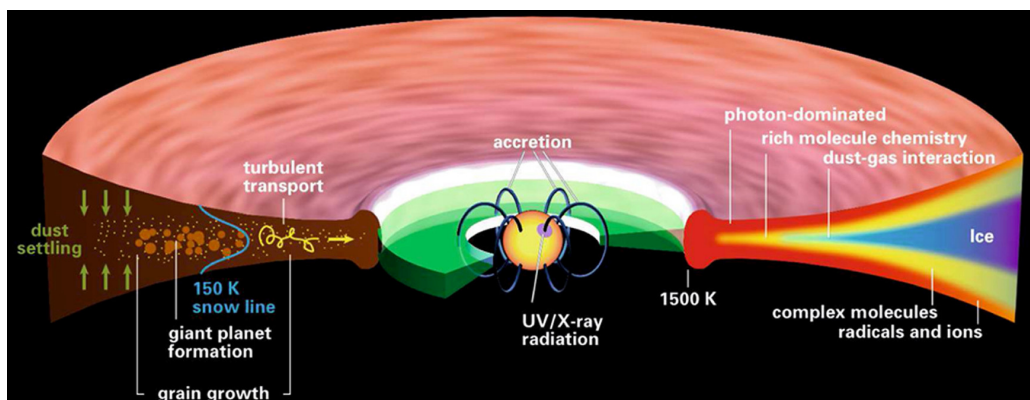


Figure 1.4: Schematic representation of the physical and chemical structure of a protoplanetary disk. (Credit: [Henning and Semenov, 2013](#))

### 1.1.2 Chemical evolution and inheritance during the formation of a Solar-type planetary system

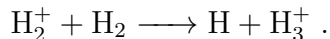
The physical conditions experienced during the Solar-type star formation phases govern the gas chemistry evolution (and vice versa) and, consequently, the molecules observed and observable by astronomers. The chemical processes in the ISM can be divided into two main classes, depending whether they occur in the gas-phase or on the surfaces of the dust grains, for brevity called hereinafter grain-phase. The two classes are strongly coupled and their interaction depends on the physical conditions of the ISM. Since observations mostly detect gaseous species, the two classes interaction also govern which molecules can be and are observed.

As just described in the previous Section, the starting point in the formation of a Solar-type planetary system is the molecular cloud. The first step in the chemical evolution starts with the formation of molecular hydrogen on the surface of the dust grains (*e.g.*, [Van de Hulst, 1948](#); [Hollenbach and Salpeter, 1971](#); [Vidali, 2013](#)).  $\text{H}_2$  is a perfect example of how chemistry proceeds in the ISM. First, under the molecular cloud conditions, where the temperature is around 10 K, only reactions that are exothermic and occur without any activation barrier can play a role. Second, the low gas density ( $\sim 10^4 \text{ cm}^{-3}$ ) (*i*) allows only bi-molecular reactions to take place in the gas-phase (limiting possible chemical combinations) and (*ii*) it does not allow stabilization of the reaction products by collision with a third body. For these reasons, the formation of  $\text{H}_2$  in the gas-phase cannot occur since, once formed, it would dissociate back to the reactants. On the contrary, the dust grain acts as a catalyst for the reactions taking place on its surface, possibly



affecting the reaction energetics and dissipating the excess energy. The latter process is highly efficient in the case of H<sub>2</sub> formation (*e.g.*, [Vidali, 2013](#); [Pantaleone et al., 2021](#)).

Once molecular hydrogen is produced on the grain surfaces and either evaporates from the surface into the gas or is directly ejected during its formation, the next major step in the chemical evolution is the formation of the H<sub>3</sub><sup>+</sup> ion. This species is produced by H<sub>2</sub><sup>+</sup> (obtained via cosmic-ray ionization of H<sub>2</sub>) and another hydrogen molecule in the gas-phase, as described by the following reaction:



Once H<sub>3</sub><sup>+</sup> is produced, a cascade of reactions involving this species enriches the gas-phase chemistry of molecular clouds ([Herbst, 2000](#)).

That said, the gas-phase chemistry presents severe limits, as it is incapable to reproduce the observed chemical complexity already when considering simple species such as water, methane or ammonia (*e.g.*, [Tielens and Hagen, 1982](#)). It is clear that, in addition to the gas-phase reactions, the observed rich chemical complexity needs the contribution of the grain-phase chemistry. This contribution starts with the beginning of the formation of a Solar-type planetary system and evolves with it, as illustrated in [Fig. 1.5](#).

**Molecular cloud and prestellar core phases** Due to the low temperatures ( $\sim 10$  K) of molecular clouds and prestellar cores, at these stages, most of the species in the gas-phase will be adsorbed and remain stuck on the grain surfaces since they do not have enough thermal energy to counteract their binding energy (BE). Once on the grain surface, some species might be efficiently hydrogenated thanks to the H atom quantum tunneling and the H low diffusion barrier ([Senevirathne et al., 2017](#)). This is, for example, the case of the O or O<sub>2</sub> landed on the grain surface, which formed water ([Dulieu et al., 2010, 2013](#)). The species synthesised on the grain-phase or adsorbed from the gas-phase create the so-called icy grain mantles, which consist mainly of H<sub>2</sub>O followed by smaller amounts of CO, CO<sub>2</sub>, NH<sub>3</sub>, making them an important reservoir of molecular species and elements ([Boogert et al., 2015](#)). It is also possible, and postulated, that these mantles also contain some radicals formed by the partial hydrogenation of atoms and species and the UV photons (produced by the interaction of the cosmic-rays with H<sub>2</sub>) dissociation (*e.g.* [Garrod and Herbst, 2006](#); [Taquet et al., 2012](#)). These radicals could also participate to the formation of more complex species on the grain mantles, combining between them (*e.g.* [Garrod and Herbst, 2006](#); [Garrod et al., 2022](#)), even though this is still a debate issue ([Balucani et al., 2015](#); [Enrique-Romero et al., 2021](#); [Ceccarelli et al., 2022](#), *e.g.*).

**Class 0 and I protostars** The increasing temperature in the more evolved low-mass star formation phases makes to become efficient other chemical and physical mechanisms than those occurring in the previous cold phases. The two most notable processes are the thermal desorption of the species frozen onto the grain mantles, previously formed, and the diffusion on the grain surfaces of the frozen radicals and, more generally, species. These two processes enhance the formation of complex species both in the gas- and grain-phases. Particularly remarkable and important in the context of the possible impact on the emergence of life are the so-called “*interstellar Complex Organic Molecules*”, *i.e.*, COMs or iCOMs (Herbst and Van Dishoeck, 2009; Ceccarelli et al., 2017), molecules containing at least six atoms of which at least one is a carbon. Indeed, iCOMs represent about 40% of the identified ISM molecules<sup>1</sup> and are an important reservoir of carbon, the basic element to develop the organic chemistry and the life in the form that we know on Earth. In the same vein, the detection of several and abundant iCOMs in Class 0 and I (see previous Section) suggests a potential link between interstellar chemistry and terrestrial organic life, which makes the study of this class of interstellar molecules particularly important.

**Inheritance?** The increased complexity observed in the Class 0 and I may be conserved in the following phases of planetary system formation, as during the protoplanetary disk phase the molecules freeze-out onto the new grain mantles. Some of these mantles could survive during the agglomeration of the interstellar grains in larger ones (*e.g.*, planetesimals), and the frozen molecules, including the iCOMs, could be incorporated in the eventually formed planets, asteroids and comets. Thus, there may be a link between the interstellar chemistry and that in planets, where iCOMs could act as the bricks for prebiotic chemistry and ultimately life (Caselli and Ceccarelli, 2012).

## 1.2 Understanding observations with astrochemical models

Since the discovery of the first molecules in the ISM in the visible spectrum (Swings and Rosenfeld, 1937; Dunham Jr, 1937; McKellar, 1940), the astrochemical community has tried to understand their origin and how interstellar molecules can survive the apparently prohibitive ISM conditions. Understanding how interstellar molecules are formed and destroyed can also

---

<sup>1</sup><https://cdms.astro.uni-koeln.de/classic/molecules>



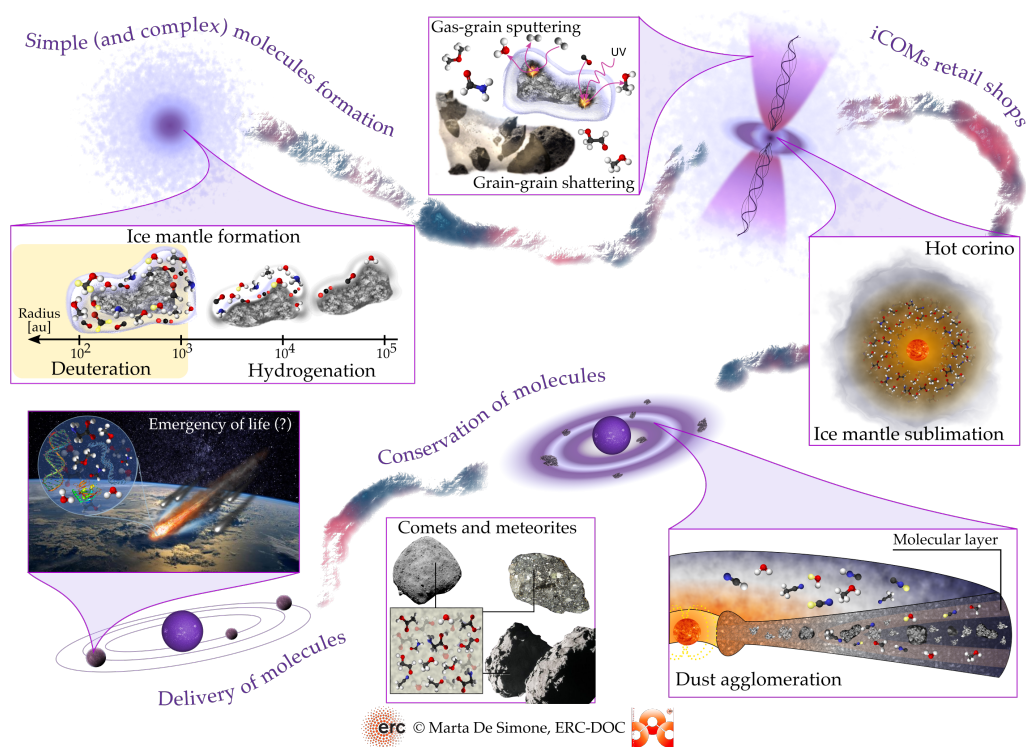


Figure 1.5: Artistic representation of the chemical evolution during the early stages of star formation. (Credit: De Simone, 2022)

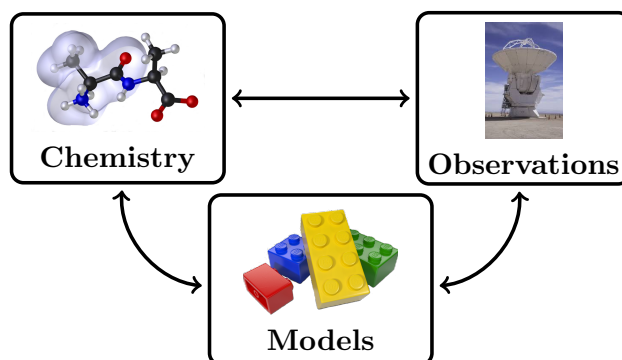


Figure 1.6: Schematic bi-univocal link between astrochemical models, observations and chemistry.

<b>Chemistry</b>	Gas-phase Gas-phase + adsorption and desorption Gas-phase + grain-surfaces
<b>Physical condition</b>	Fixed Evolving with time and/or geometry position
<b>Geometry</b>	0D 1D, 2D and 3D

Table 1.1: Levels of complexity of astrochemical models.

help us to infer the properties, history, and evolution of astronomical objects (see *e.g.*, De Simone et al., 2022). Thus, in parallel with the identification of these molecules from the astrochemical spectra and the measurements of their abundances, soon a community dedicated to developing astrochemical models began to play an important role (*e.g.*, Herbst and Klemperer, 1973; Prasad and Huntress Jr, 1980). For doing that, models need to collect information from astronomical observations, chemical experimental and quantum computational studies, becoming a crucial link between all these sources of information (Fig. 1.6). The final aim of a model is to reproduce the observed abundances, in the chosen ISM object physical conditions, and, if successful, to provide complementary details to observations (*e.g.*, age of the modeled object, important reactions/processes in the chosen conditions, etc.).

### 1.2.1 What is an astrochemical model?

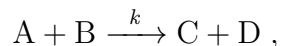
In literature, many astrochemical models with different levels of complexity have been proposed over the years (*e.g.*, Herbst and Klemperer, 1973;

Prasad and Huntress Jr, 1980; Tielens and Hagen, 1982; Hasegawa et al., 1992; Viti et al., 2004; Garrod and Herbst, 2006; Taquet et al., 2012; Garrod et al., 2022). The complexity of a model depends on the considered chemical processes, the physical conditions and system’s geometry, as roughly summarized in Tab. 1.1 (from Agúndez and Wakelam, 2013). The level of complexity depends on the purpose of the study and the astronomical object characteristics that the modeler aims to simulate.

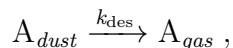
In general, current chemical models compute abundances as a function of time, starting from an initial composition and for a set of parameters. Considering a simplified model for gaseous and volatile species, the most important parameters/inputs are the following:

- temperature of the gas and dust;
- visual extinction ( $A_v$ );
- cosmic-ray ionization rate for  $H_2$  ( $\zeta$ );
- dust grain properties, *e.g.*, dimension, density, number of sites per grain etc..;
- density of the gas, usually assumed to be the Hydrogen nuclei number density ( $2[H_2]+[H]$ );
- initial molecular and atomic abundances;
- **Reaction Network.**

With reaction networks here I mean all the chemical reactions/processes (their reactants, products, reaction rate coefficients, etc..) that are encoded with the so-called “*chemical reaction formalism*”, *e.g.*,



where A and B are the reactants, C and D the products and  $k$  is the rate coefficient. This last element describes the efficiency of the reaction as a function of temperature or other physical parameters (*e.g.*, photon or cosmic-ray). This formalism can be applied not only to reactions but also to other processes. An example is the thermal desorption of a species from the dust grain surface, which can be represented as follow:



where  $A_{dust}$  represents the A species when adsorbed on the dust grain surface,  $A_{gas}$  the A species on the gas-phase and  $k_{des}$  the desorption rate constant.

This compact formalism allows to approximate the complex potential energy surface (PES) of a reactive molecular system, which describes its energy in terms of geometric parameters such as inter-nuclear distances, and its probability to pass from the reactants to products at a given temperature. The whole list of processes encoded within the chemical formalism (*e.g.*, the Reaction Mechanism box in Fig.1.7), usually goes under the name of “*Reaction Network*” (Feinberg, 2019).

Generally, models numerically solve a system of non-linear coupled Ordinary Differential Equations (ODEs) created considering a set of species and the chemical processes that inter-connect them, usually called rate equation model. Each species has an equation based on the Mass Action law (first enunciated by Guldberg CM., 1864), with the following form:

$$\frac{dn_i}{dt} = \overbrace{\sum_{j,k} k_{j,k} n_j n_k}^{\text{bi-molecular reactions forming } n_i} - \overbrace{n_i \sum_j k_{i,j} n_j}^{\text{bi-molecular reactions destroying } n_i} + \underbrace{\sum_j k_j n_j}_{\text{uni-molecular reactions forming } n_i} - \underbrace{n_i \sum_j k_j}_{\text{uni-molecular reactions destroying } n_i},$$

where  $n_i$ ,  $n_j$  and  $n_k$  represent the density of species  $i$ ,  $j$  and  $k$  respectively, while  $k_{j,k}$  and  $k_{i,j}$  are the bi-molecular rate coefficients of reactions producing and destroying  $n_i$ ,  $k_i$  and  $k_j$  the uni-molecular ones. Figure 1.7 presents an example of chemical reactions and how they constitute a set of ODEs, where each species is coupled/dependent on the other species making the solution of the ODE highly non-linear.

The complexity of the system is interesting per se, from a mathematical point of view. Indeed, a field of applied mathematics under the name of chemical reaction network theory arose to rationalise the mathematical properties of such a system. Number of steady states, basin of attraction, stability of steady states, existence of stable periodic solutions, network structure and dynamical properties are some of the properties that can be extracted. The mathematical description of the above-mentioned (and not only) properties of the reaction network from a theoretical point of view can be found in the textbook by Feinberg (2019) meanwhile an introduction to non-linear dynamical system in Strogatz (1994) .

The evolution of the abundances obtained solving the above ODEs is a deterministic approach. Alternatively, also stochastic approaches based on the so-called chemical master equation (Van Kampen, 1992; Gardiner, 2009), even if computationally highly demanding, have been formulated (*e.g.*, Charnley, 1998, 2001). It is worth mentioning that the nature of all these

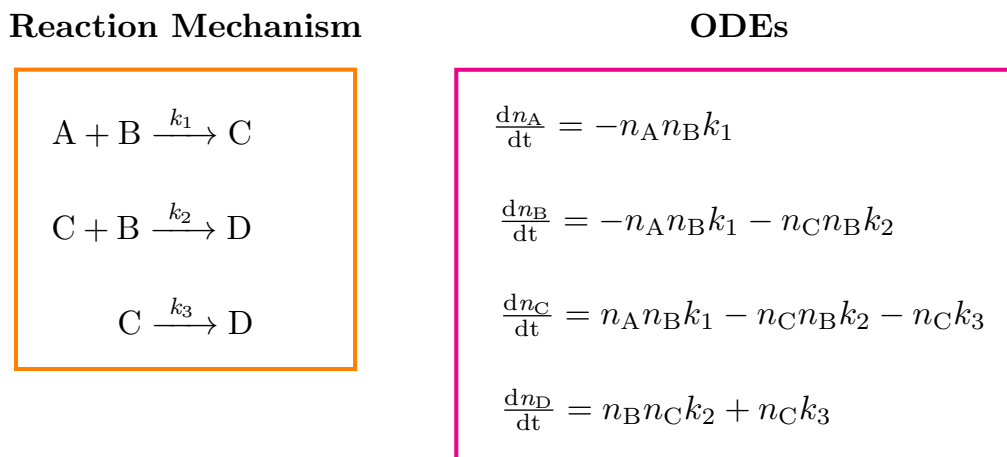


Figure 1.7: Example on how to encode a set of chemical reactions (*left panel*) into Ordinary Differential Equations (ODEs) (*right panel*).

chemical processes is stochastic and the deterministic model is an approximation of the real stochastic microscopic behavior of the species evolution. In fact, starting from the chemical master equation, when the fluctuations are negligible, the deterministic equations can be recovered through the so-called mean-field approximation (see *e.g.*, [Gardiner, 2009](#)). On the other hand, when random fluctuations cannot be ignored, a stochastic treatment is essential for accurate results.

The deterministic rate equation models are widely (almost the only paradigm) used to model the gas-phase chemistry in astrochemistry, since the stochastic approach is not necessary. On the other hand, the resolution of species evolution in grain-phase is not unequivocal, and can be described via both deterministic (*e.g.*, [Hasegawa et al., 1992](#)) and stochastic approaches (*e.g.*, [Tielens and Hagen, 1982](#)). However, the deterministic formalism implicitly assumes an “*infinite*” surface where even small coverages are meaningful without considering the finite size of the system, that is not the ISM case, driving in many cases to inaccurate results ([Tielens, 2005](#); [Stantcheva and Herbst, 2004](#); [Cuppen et al., 2013](#)). For this reason, even if more computational demanding, stochastic methods based on Monte Carlo algorithms are sometimes applied in astrochemistry (*e.g.* [Tielens and Hagen, 1982](#); [Charnley, 2001](#); [Stantcheva and Herbst, 2004](#); [Chang et al., 2007](#); [Cuppen et al., 2009](#); [Tsvetkov and Shematovich, 2010](#); [Cuppen et al., 2013](#)). However the vast majority of models rely on the deterministic approach (*e.g.*, [Hasegawa et al., 1992](#); [Aikawa and Herbst, 2001](#); [Garrod and Herbst, 2006](#); [Hincelin](#)

et al., 2016; Taquet et al., 2012; Garrod et al., 2022)

## 1.2.2 Reaction network: the core of astrochemical models

One of the most important elements in astrochemical models is the reaction network, *i.e.*, the list of chemical reactions, their rate constants and branching ratios (or products). The astrochemical reaction network can be divided into two sub-networks depending on whether the reactions/processes occur in the gas-phase or on the grain-surfaces. These two sub-networks are highly coupled and interdependent, as demonstrated by the fact that many observed abundances cannot be reproduced without considering both networks (*e.g.*, Tielens and Hagen, 1982; Vasyunin and Herbst, 2013; Balucani et al., 2015). An example of the interplay of the two sub-network is illustrated in Figure 1.8 for the water case (van Dishoeck et al., 2013).

Here are some numbers to measure the size of the most widely used reaction network<sup>2</sup>:  $\sim 8000$  is the number of reaction in gas-phase network with respect to  $\sim 800$  grain-phase reactions.

**Gas-phase reaction network** The early attempts to predict the species observed abundances using astrochemical models of cold molecular clouds relied on gas-phase reactions only, with the exception of the formation of  $\text{H}_2$  on the grain surfaces (Herbst and Klemperer, 1973; Mitchell et al., 1978). In general, the most significant gas-phase reactions that contribute to an increase in chemical complexity in the low density regions of Solar-type star formation are bi-molecular reactions. The different number between classes of reactions in the reaction network can give information on the history and importance of these reactions in astrochemistry. Here are listed some indicative numbers from gas-phase reactions in the Wakelam et al. (2015) network: ion-neutral reactions represent about 50% of the network, cation-anion reactions  $\sim 20\%$ , neutral-neutral reactions  $\sim 15\%$ , electron recombination reactions and unimolecular reactions (photo-dissociation or CR-induced dissociation)  $\sim 9\%$  each.

The two major classes of bi-molecular reactions that increase the chemical complexity in the molecular ISM are, therefore, ion-neutral and neutral-neutral. As mentioned in Subsection 1.1.2, due to the low temperature experienced in the studied astrochemical objects, bi-molecular reactions are

---

<sup>2</sup>Data retrieved from the KIDA database (<https://kida.astrochem-tools.org/>), in October 2022.

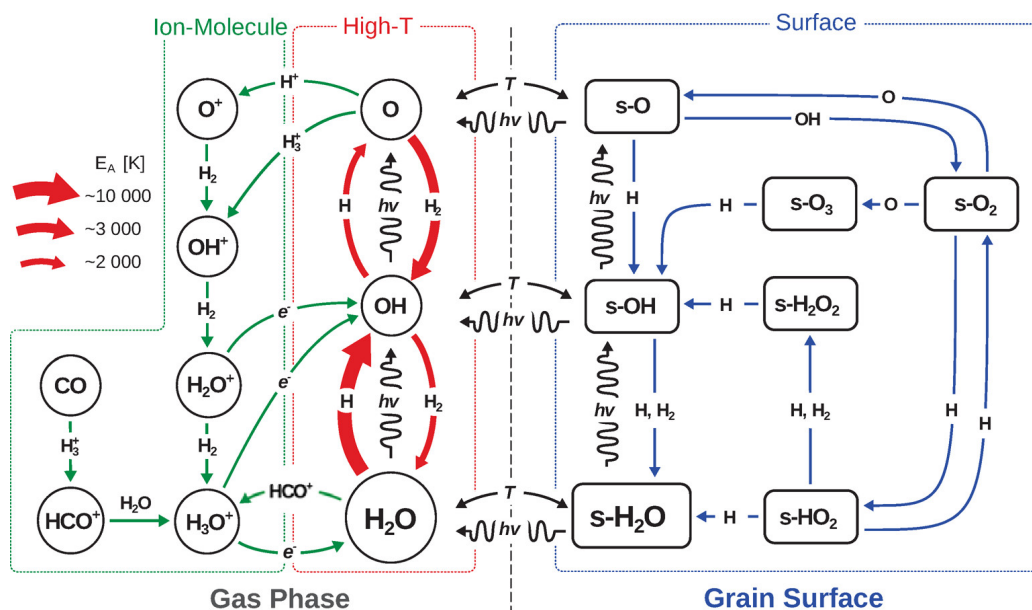


Figure 1.8: Chemical reaction network, gas- and grain-phase, for water formation. (Credit: van Dishoeck et al., 2013)

always exothermic and involve either a radical (species with unpaired electrons) or an ion, the reactions of which are usually characterized by very small or null activation energy. Another constrain to the increase in chemical complexity through gas-phase reactions is the absence of a third body (due to low density) that can collide with the excited species and adsorb excess reaction energy, preventing the reaction from going backwards. In rare cases, the excess energy can be released via emission of photons in the so-called “*radiative association*” reactions (Herbst, 1985; Klippenstein et al., 1996; Tennis et al., 2021). The only (widely applicable) alternative that remains to dissipate the exothermicity of the reactions is to follow the exothermic channels of the potential energy surface (PES) and form products that can adsorb the excess of energy (through translational or vibrational degrees of freedom) without further dissociation. For this reason, in the absence of a third body or the emission of a photon at least two products are always present. An example is provided by the following proton transfer reaction:  $\text{CH}_3\text{O} + \text{H}_3^+ \longrightarrow \text{H}_2 + \text{CH}_3\text{OH}^+$ , where the reaction energy (-282 kJ/mol) is mainly adsorbed by  $\text{H}_2$ .

For the aforementioned reasons, interstellar gas-phase chemistry was initially based mostly on ion-neutral reactions (Herbst and Klemperer, 1973) because of their relatively large rate coefficients at low temperatures. In



general, the latter are obtained via experimental studies, computed with the capture theory (Su and Chesnavich, 1982; Herbst, 2006; Woon and Herbst, 2009) and focused theoretical studies of the PES and kinetics. Only in the last decades, with results achieved by the CRESU experimental method (Smith and Rowe, 2000), neutral-neutral reactions have started to be inserted in astrochemical models and to play a role in the formation of molecules with more than six atoms (Balucani et al., 2015).

**Grain-surfaces reaction network** Despite the larger size of the gas-phase network, the grain-surfaces reaction network is of fundamental importance in explaining the formation and abundances of many molecules. The high efficiency formation of water (Dulieu et al., 2010, 2013), ammonia (Fedoseev et al., 2015), formaldehyde (Molpeceres et al., 2021), methanol (Hiraoka et al., 1994; Watanabe and Kouchi, 2002; Rimola et al., 2014) and ethanol (Chuang et al., 2021; Perrero et al., 2022b) are some examples. The molecular complexity, enhanced by the catalytic effect of the grain surface, is mainly obtained by three classes of reactions:

- Hydrogenation of atoms/molecules by H atom: these reactions can take place efficiently at low temperature thanks to quantum tunnelling of the H atom and its low diffusion barrier (*e.g.*, Tielens and Hagen, 1982; Senevirathne et al., 2017).
- Radical-icy species reactions, in which the gaseous landing radical or atom reacts directly with the icy grain mantle surface after being adsorbed from the gas-phase (*e.g.*, Rimola et al., 2018; Jin and Garrod, 2020; Molpeceres et al., 2021).
- Radical-radical recombination: in this case the radicals are both adsorbed on the surface and need to meet (via thermal and/or not-thermal diffusion) to react (*e.g.*, Garrod et al., 2008; Enrique-Romero et al., 2022).

Besides the reactions (as the above-presented) and the diffusion energy (DE) barriers of the species adsorbed on the grain surface, the processes at the interface between the gas-phase and the grain-surfaces included in the grain-phase network are: (*i*) adsorption of species from the gas-phase to the surfaces and, the other way around, (*ii*) desorption (thermal and non-thermal) of the species from the grain-surfaces to the gas-phase. Adsorption, thermal desorption and diffusion are normally encoded in the model as shown in Figure 1.9.



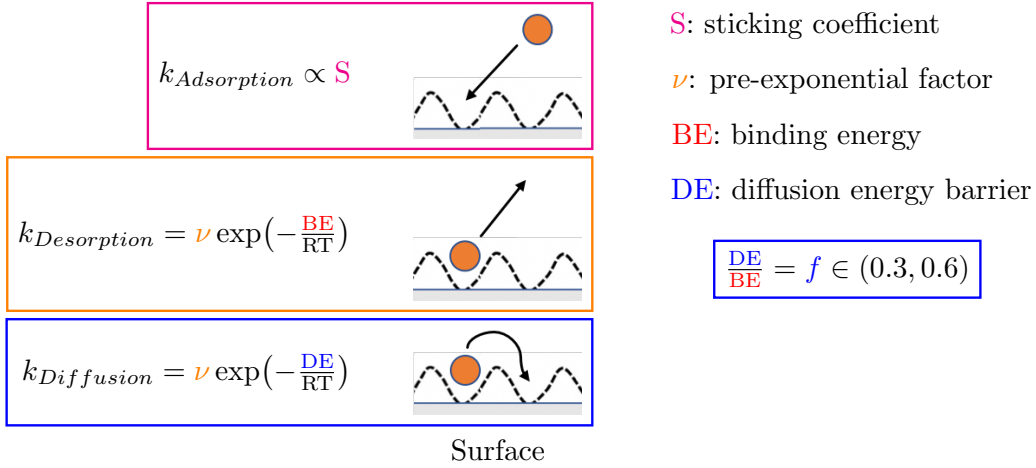


Figure 1.9: Mechanisms of adsorption, desorption and diffusion on a surface and their rate constants. The chemical parameters encoded in a grain-phase reaction network are highlighted in colour. T and R are the grain temperature and the ideal gas constant, respectively.

The chemical information needed to model all these processes are the reaction constants, binding energies (BEs), diffusion energy (DE) barriers, the pre-exponential factors ( $\nu$ ) for desorption and diffusion rate constants, and the sticking coefficients (S), which are all highly dependent on the characteristics of the surface. An example is provided by the ammonia BE, which on a water ice surface is  $\sim 4000$  K (He et al., 2016) with respect to  $\sim 3000$  K on a gold surface (Kruczkiewicz et al., 2021).

Thermal desorption, as just explained before, plays a key role in astrochemical modeling (Minissale et al., 2022; Ferrero et al., 2022). In general, laboratory experiments measure the rate of desorption via the so-called Temperature-Programmed Desorption (TPD) technique, in which BE and  $\nu$  are derived from the inversion of the desorption rate curve (King, 1975; He et al., 2016). This derivation is degenerate in BE and  $\nu$ , hence, in order to obtain the desorption rate constant, what is meaningful is the couple and not their separate values. From the computation point of view, normally only the BE is provided even if methodologies based on the transition state theory (TST) can be used to obtain the value of  $\nu$  (e.g., Fair and Madix, 1980; Tait et al., 2005; Sprowl et al., 2016). The other two, and more often used, values for  $\nu$  in astrochemical models are a constant of  $1 \times 10^{12} \text{ s}^{-1}$  and the following

formula proposed by Hasegawa and Herbst (1993):

$$\nu = \sqrt{\frac{2 \text{BE}}{\pi^2 m A}}, \quad (1.1)$$

where  $A$  is the surface area per adsorbed molecule (normally assumed to be  $10^{-19} \text{ m}^2$ ) and  $m$  is the mass of the adsorbed species. Obviously, the most important role in the desorption rate evaluation is played by the BE value, since it is in the exponent of the formula (see Fig. 1.9).

Furthermore, the value of the BE also impacts the diffusion energy barrier. In fact, many laboratory and theoretical studies express DE in terms of a fraction of BE, *i.e.*  $f = \text{DE}/\text{BE}$  (Fig. 1.9). A value of  $f$  between  $\sim 0.3$  and  $0.6$  is normally assumed depending on the coverage of the adsorbed species, its charge and whether it is a molecule or an atom (Minissale et al., 2016; He et al., 2018).

In summary BE has a huge role in the description of the surface processes. Moreover, it governs the abundance of species in the two phases, gas or solid, and, thus, via astrochemical models, provides important quantitative information to explain astronomical observations. For example, the BE of water regulates the snow-line of protoplanetary discs (Öberg and Bergin, 2021).

### 1.3 Problems and challenges for astrochemical reaction networks at the beginning of this thesis

Many sources of uncertainties are present in astrochemical models, from the physical parameters of the astronomical object and the various intrinsic approximations of the model, to the reaction network. If the parameters and modeling approximations can be evaluated “*more easily*”, the reaction network, for its dimensions and non-linear behaviors, has an intrinsic difficult complexity. Many literature examples are known on the possible huge discrepancy in predicted abundances every time (just) a single reaction for a specific species is added, modified or deleted. An example can be the dimethyl ether gas-phase reaction case (Ascenzi et al., 2019; Skouteris et al., 2019) or the importance that just a single BE value can have to interpret a protoplanetary disc snow-line (Grassi et al., 2020).

The most updated reaction networks are created by a layering process over years, from the pioneering works in the 70s and 80s (*e.g.*, Herbst and Klemperer, 1973; Mitchell et al., 1978; Herbst et al., 1989) up to the main

insertions of reactions (“*borrowed*”) from gas-phase combustion chemistry (Anicich, 2003) and (“*LEGO-like*”) radical-radical grain-surfaces reactions (Garrod et al., 2008). Respect to the backbone of the reaction networks, which has remained almost intact during the years, most of the attention has focused on reactions that can explain newly detected species and their abundances. In this way, the number of reactions, reactants and products have increased each time a new molecule was detected, which also led to an increase in the network size and complexity. This behavior has the consequence of focusing the attention on just a few terminal reactions (for “*just*” explaining the formation and destruction route of new species) in the network without necessarily considering all the possible reactions that a new introduced species could have. For this reason, the new reaction network contributions are mainly oriented to newly detected species without giving enough importance to reviewing or controlling “*old*” reactions via more accurate experiments or computational studies.

To give an idea of the reaction network status, only  $\sim 20\%$  of the present reactions/processes are studied/verified via QM calculations or experimental studies. A meticulous and huge work by astrochemists dedicated to releasing new reaction networks has been done to discriminate and control possible new or old reactions and values (*e.g.*, Wakelam et al., 2010b). Nonetheless, the huge number of reactions makes impossible a case-by-case extensive control. This problem opens the door to new attempts for improving the networks systematically and extensively.

Many techniques, mainly based on Monte Carlo methods, were used during the years in order to estimate uncertainties in models (*e.g.*, Vasyunin et al., 2004; Wakelam et al., 2006, 2010a; Makrymallis and Viti, 2014; Penteado et al., 2017). These techniques are one of the most important tools in modelers hands, since they can give information on the hierarchy of the reactions/processes involved under specific conditions. Consequently, once the most important reactions are found, those can be refined via experiments or theoretical computations. Anyways, the main challenge of such techniques have to face is the huge and non affordable computational time required to sample all the parameter space. This is a consequence of the non-independence of the parameters due to the coupling between the species in the reaction network.

All the above-presented problems and challenges regard the whole reactions network (and applicable to both the sub-networks). Specific arguments for gas-phase and grain-phase reaction networks will be discussed in details in the next Subsections.

### 1.3.1 Gas-phase reaction network

The number of species present in the gas-phase reaction networks are of the order of 500, but only less of half of them were characterized theoretically (Woon and Herbst, 2009) and, often, the only information available for the remaining species is their codification name. This lack of information mainly regards cations or neutral species that belong to intermediate reactions in the formation of more complex and astronomically important species.

Another problem is related to the non-univocal representation of the species, *i.e.*, the univocal connection between a name and a structure. With the addition of larger detected species in the reaction network, a representation problem related to the discrimination of isomers has begun to arise. This process can be seen as a consequence of the layered human-supervised processes limits in the creation of such reaction networks. Many limitations are present in the commonly used astrochemical nomenclature, as evidenced by the fact that no standard is provided. At the moment, for examples, different articles/databases report the same species with different names and moreover no definition is provided to distinguish between cis and trans isomers. This problem, consequently, affects the accuracy of the reaction network, as the latter collects all the chemical information. An example can be represented by the  $\text{SiOH}^+$  cation, which in KIDA<sup>3</sup> database is reported the less stable isomer (*i.e.*,  $\text{HSiO}^+$ , were the H atom is bounded to the Si atom) respect to the more stable isomer (*i.e.*,  $\text{SiOH}^+$ , were the H atom is bounded to the O atom) reported in UMIST<sup>4</sup> database.

Among the various improvements needed to make astrochemical gas-phase networks more reliable, two are essential: (*i*) control the feasibility of each included reaction, and (*ii*) introduce new reactions. One control that has never been applied in astrochemistry would be the exo/endothemicity reaction estimation and the subsequent identification of endothermic reactions encoded as barrierless. To achieve this, as previously underlined at the beginning of this Subsection, a complete characterization of the species from a theoretical point of view is necessary, as the experimental information is usually scarce (especially regarding the enthalpy of formation of the astrochemical species, *i.e.*,  $\Delta H_f$ ).

Regarding the kinetic considerations, accurate rate constants are normally obtained via time-consuming and case-by-case theoretical calculations or experiments. Other less accurate (and frequently source of error) approaches regard constants inferred via educated guess, similarity principle and capture theory. This latter approach provides an estimated upper limit in the case

<sup>3</sup><https://kida.astrochem-tools.org/species/257/HSiO+.html>

<sup>4</sup><http://udfa.ajmarkwick.net/?search=SiOH%2B>

of ion-neutral barrierless reactions (Su and Chesnavich, 1982).

### 1.3.2 Grain-phase reaction network

Many uncertainties are present in the grain-phase reaction network, not only in the surface parameters entering in the description of the various processes (such as BE, reactions, diffusion barriers, etc..) but also in the processes themselves. If the first issue has similarities with the uncertainties of gas-phase reactions (*e.g.*, incorrect reactions, inaccurate values, lack of information, etc.), the second one is more influenced by the approximations of the modeling and, moreover, by the relatively poor knowledge of the processes themselves (He and Vidali, 2014; Minissale et al., 2022). Vibrant discussions are present in the astrochemistry community, mainly driven by experiments, astrochemical models and computational studies to understand better the processes on the grain.

Experimental studies are crucial for leading the community since they can tell us whether a mechanism operates or not under conditions similar to those of the ISM and they can provide extraordinary physicochemical data on molecule and surface properties, but they have limitations too. Firstly, the different time-scales at which experiments and star formation processes in the interstellar medium happen. If low-mass star formation has scales of tens of millions years, the Earth-based laboratory experiments happen on scales of hours or days. Secondly, many microscopic aspects belonging to the atomic scale cannot be investigated due to the “*ensemble*” detection methods, *i.e.*, one cannot detect a single molecule desorbing from a surface or a single reaction. And precisely in relation to this latter problem, experiments such as TPD cannot disentangle the BE value of species that have an interaction with the surface of the same nature (and strength) of the molecules that construct the surface itself. An example is provided by He et al. (2016) in which the ammonia BE on a water ice surface cannot be obtained since starts to desorb with the water molecules of the ice. All these problems can be solved using computational chemistry.

One of the most important tool for studying astrochemical surface processes at atomic level is indeed computational chemistry. As an example, it has been disproved that many radical-radical surface reactions are barrierless, as previously thought (Enrique-Romero et al., 2022). However, computational chemistry applied to surface processes has to deal with two main challenges: (*i*) computational method accuracy and (*ii*) reliability of the surface model. The accuracy of computational results depends on the chosen level of theory. As a consequence, the main problem a computational scientist has to face is the balance between computational cost and accuracy,

depending on the size of the studied system. Many improvements have been made in this direction with new electronic structure methods (*e.g.*, Guo et al., 2018; Grimme et al., 2021) and machine learning (ML) force field methods (Poltavsky and Tkatchenko, 2021), which can be applied to astrochemical problems. The second main challenge concerns the chosen surface model, more specifically the icy-grain mantle model. It is of central importance in simulations, as demonstrated by the fact that different surfaces can give very different results, which raises a reproducibility issue. For many reasons, small surface models should be avoided due to the non-physical significance and inaccurate results (in contrast to some published works, *e.g.*, Wakelam et al., 2017; Piacentino and Öberg, 2022), mainly related to non-negligible (and non-physically) surface mobility and not considered interactions. On this side, new and more complex surface models with many chemically different surface sites are needed.

The binding energy has a key role in astrochemical models, for the importance in modeling thermal desorption and diffusion. For these reasons, a huge effort has been done to compute accurate BEs via computational studies (*e.g.*, Ferrero et al., 2020; Bovolenta et al., 2020; Duflot et al., 2021; Perrero et al., 2022a; Ferrero et al., 2022; Molpeceres et al., 2022) or ML approaches (*e.g.*, Villadsen et al., 2022). Anyway, only a handful molecules of those in astrochemical databases have a measured or computed BE (Minissale et al., 2022; Ceccarelli et al., 2022). Nonetheless, almost all the computed BEs rely on a single or very few sites of adsorption, missing all the statistical information that a binding energy distribution can provide (He and Vidali, 2014). In addition, the theoretical calculation of one or few BEs can be biased by the computational scientist, which, for example, can start geometrical optimisations of the adsorbed molecule close to more stable sites. In recent years, progress has been made to extend the approach from a single value, passing to more values ( $\sim 10$  BEs per molecule Ferrero et al., 2020), to a distribution of BE (Bovolenta et al., 2020, 2022) which can lead to a better description and understanding on the processes that happened on the icy grain mantle surface (He and Vidali, 2014; Grassi et al., 2020).

Last but not least, in the last year a discussion on the commonly used desorption pre-exponential factors in astrochemistry ( $1 \times 10^{12} \text{ s}^{-1}$  or obtained via Eq. 1.1 proposed by Hasegawa and Herbst (1993)) also began in the astrochemical community (Minissale et al., 2022; Ferrero et al., 2022; Ceccarelli et al., 2022). In fact, even if with a lesser role, the  $\nu$  can have a non-negligible impact on the desorption rate (see *e.g.*, Ferrero et al., 2022; Ceccarelli et al., 2022).

## 1.4 Thesis context and objectives

### 1.4.1 Context: The ITN Astro-Chemical Origin (ACO) project

This thesis is part of the “*Astro-Chemical Origins*” (ACO)<sup>5</sup> project, which is financed by the European Union’s Horizon 2020 and Innovation Program, within the Innovative Training Networks (ITN) Marie Skłodowska-Curie Action (MSCA), grant agreement No 811312. ACO is a collaborative network of 13 European institutes and companies. As stated in the ACO website, the project has two main goals:

- to unveil the early history of the Solar System, using the chemical composition of today forming Solar-like planetary systems and comparing it with that of the Solar System primitive bodies;
- to train a new generation of researchers able to tackle this highly interdisciplinary problem, providing them with a wide-range of transferable skills, including the ability to communicate Science to a large audience.

These goals can only be achieved through an interdisciplinary collaboration among different communities. The project is divided into four scientific work packages (WP):

1. (Receivers) to improve the detection capabilities of radio/millimetre receivers, in order to observe weak sources and low abundance molecules;
2. (Observations) to build up a large and homogeneous database of astronomical spectral observations, in order to derive the chemical composition of forming Solar-like systems as a function of evolution and environment;

---

<sup>5</sup>website: <https://aco-itn.oapd.inaf.it/>





3. (Laboratory & Theory) to understand the chemical processes in the exotic interstellar conditions via ad-hoc quantum chemistry computations and laboratory experiments, in order to establish a reliable network of reactions occurring in the forming Solar-like systems;
4. (Model & Tools) to develop sophisticated astrochemical models and tools, using modern techniques in order to predict the chemical composition of forming Solar-like planetary systems and compare it with observations, to finally build up a new model for the Proto Solar Nebula.

Each of the four objectives requires a specific expertise, possessed by different communities: instrumentation, astronomy, chemistry (experimental and theoretical) and computer science. The backbone of the project is constituted by 17 PhD students, who have carried out their thesis, each in different WPs.

Parallel to the scientific research, public engagement and outreach activities were carried out to increase the visibility and impact of astrochemistry among a broad non-scientific audience.

**Outreach activities** Within the ACO project, a part of this thesis was also dedicated to learn and built outreach activities with all the other PhD. The main achievement was obtained with the development of a virtual reality (VR) project about astrochemistry and its presentation in science festival. This product is built by several interactive and immersive games using VR technology, which allows to introduce astrochemistry. The name of this project is “The journey of water to Earth” (See more in Appendix A).

In the framework of outreach activity, Prof. Ugliengo and I created a 3D printed icy grain mantle model (Figure A.1(b)). Which it allows to easy engage and explain to a broad audience the importance of the icy grain mantle in astrochemistry.

**Thesis context** In the ACO multidisciplinary environment, my thesis belongs to the Model & Tools work package. Specifically, it focuses on the improving and cleaning of existing astrochemical reaction networks. The thesis was planned to be carried out at the Université Grenoble Alpes in the Institut de Planétologie et d’Astrophysique de Grenoble (UGA, IPAG, Grenoble, France) under the supervision of Prof. Cecilia Ceccarelli (astronomer and modeler) and at the Department of Chemistry of the Università degli studi di Torino (UniTo, Torino, Italy) under the supervision of Prof. Piero Ugliengo (theoretical and computational chemist). Therefore, I spent 18 months in each institute/university respectively.



The thesis original focus was on the application of Artificial Intelligence (AI) techniques to improve astrochemical reaction networks. During the first eight months of my thesis, an extended literature review, preliminary tests and discussions with experts were carry out on the possible application of AI techniques to our reaction networks. This knowledge made us realized that in order to make any inference on our target system using AI techniques a reliable and huge database of information was needed, which is not the case of astrochemical databases. For this reason in order to pave the way towards this direction, one part of the thesis was based on the creation of a reliable database of chemical information. Moreover, even if not reported in this manuscript, an extensive study of ML force field for molecular simulations (and some preliminary results) were obtained during my PhD.

Of all the ACO PhD thesis, this manuscript is one of the most interdisciplinary. This is because a combination of knowledge in computational chemistry, computer science and programming, astrochemical modeling and astrophysical context was needed to achieve the results of the thesis. It took a lot of effort from my side (and also on the side of my supervisors) to understand each other and deal with such an interdisciplinary context. Ultimately, I would say that I am neither a pure computational/theoretical chemist nor an astrochemical modeler, hence somewhere in between.

## 1.4.2 Objectives of the thesis

The astrochemical reaction network (and its gas-phase and grain-phase sub-networks) is the object that collects all the chemical information and processes coming from experimental and theoretical studies. In the framework of this thesis, the processes at the interface between the gas- and grain-phases are considered part of the grain-phase reaction network.

Before the start of my thesis, a lot of work was done to improve the existing astrochemical reaction networks. However, as described in Section 1.3, several problems remained to be tackled. In particular, using computational chemistry my thesis aimed to tackle specific problems related to the gas and grain-phase reaction networks. For this reason, the objective of this work can be divided in two sub-objectives, focused on the improvement of gas- and grain-phase networks, respectively.

### Gas-phase network improvement

**Objective** The first objective of this thesis is to produce a more reliable gas-phase reaction network. To this end, it is indispensable to acquire a deeper understanding of which class of gas-phase reactions are incorrect.

**Implementation** In order to achieve the above-mentioned goal, the first step was to make some order in the astrochemical species nomenclature and to characterize all species from a theoretical point of view. The species characterization allows the creation of a correct database, which is mandatory to make any inference on reactivity or other aims. In particular, after the characterization of neutral species by [Woon and Herbst \(2009\)](#), the species most affected by this lack of information are cations that act as reaction intermediates in the production of more complex detected species. For this reason, a methodology for characterizing all these species and their collection into a database of chemical information was performed. Once the database built, the second step was to compute the reaction energy of each reaction present in one of the most up-to-date gas-phase reaction networks to find and delete erroneous reactions, specifically endothermic reactions encoded as barrierless, and understanding their origin. Finally, the impact of the new reaction network on the predicted abundances via astrochemical modeling had been estimated.

## Grain-phase network improvement

**Objective** From a pure computational chemistry perspective, the main aim of this part is to provide a more accurate and complete image of the thermal desorption process of astrochemical molecules from ice surface models.

**Implementation** As already mentioned in Subsection 1.3.2, if accurate BEs want to be provided, a reliable surface model and an accurate level of theory have to be used. To this end, a new ice grain model with large surface area and an high variability of chemically different adsorption sites was produced. Once this model was developed, it was used as starting point for the development of an accurate methodology to calculate accurate BE distributions. This methodology was tested for important astrochemical species like ammonia and water. Then, these distributions were commented in the light of the astrochemical implications as well as the comparison with the standard TPD experimental technique. Complementary to the BE, the pre-exponential factor ( $\nu$ ) in desorption rate has a non-negligible role. This quantity has been less studied in astrochemistry compared with studies coming from the material science community. For this reason and in order to understand better the importance and limits of this quantity, a comparison between the commonly used formula in astrochemistry (proposed by [Hasegawa and Herbst, 1993](#)), the free-rotator formula (proposed by [Tait et al., 2005](#)) and its harmonic oscillator extension (both the latter based on transition state theory) was performed.

### 1.4.3 Thesis outline

The results of the thesis are divided into two parts, one relating to gas-phase network improvement (Part II) and the second to the grain-phase network (Part III). Each one of these two parts has their own short introduction with the scientific context, the specific goals, the novelties, the methods and a summary of the results. Then, the research carried out in each part is presented in different Chapters. Finally, the conclusions and the perspectives of this doctoral thesis work are presented in Part IV.

Before entering in the core of the thesis, the Chapter 2 provides an overview of the methods and theories used to achieve the objectives of the thesis.

# Methods

## Contents

---

<b>2.1</b>	<b>Quantum mechanic in chemistry</b>	<b>38</b>
2.1.1	The Hamiltonian, principles and approximations in molecular system	38
2.1.2	The Hartree-Fock method	42
2.1.3	Post Hartree-Fock methods	44
2.1.4	DLPNO-CCSD(T) method	45
2.1.5	ONIOM method	47
<b>2.2</b>	<b>Graph theory for molecules and reaction network</b>	<b>48</b>
2.2.1	Molecules	51
2.2.2	Reaction Network	56

---

This chapter will present the most important theories and methods that have been applied to achieve the aims of this thesis. The discussed methodologies can be found in standard textbooks and articles, therefore the level of details is kept at its minimum to provide the essential framework within which discuss the main results.

## 2.1 Quantum mechanic in chemistry

This section introduces the theoretical fundamentals about electronic structure theory and methods for computing molecular properties. The theory in this section, up to sec. 2.1.4, is inspired upon the Szabo and Ostlund (2012) textbook, Enrique-Romero (2021) and Jacopo Lupi Lupi (2022) PhD thesis.

### 2.1.1 The Hamiltonian, principles and approximations in molecular system

Many chemical properties can be deeply understood and characterized only if electronic structure calculations for the molecular system are carry out. Quantum mechanics is the fundamental theory of physics that can explain the microscopical behaviour of matter. The fundamental brick of this theory is the Schrödinger equation:

$$\hat{\mathcal{H}} |\Psi(\mathbf{r}, t)\rangle = i\hbar \frac{\partial}{\partial t} |\Psi(\mathbf{r}, t)\rangle ,$$

where  $\hbar = h/2\pi$ , with  $h$  Planck's constant,  $\hat{\mathcal{H}}$  the Hamiltonian operator and  $|\Psi\rangle$  the wavefunction which fully describes a state of the system. This linear partial differential equation describes the time evolution of a N particles system, with coordinates  $\mathbf{r} = \{\mathbf{r}_i\}$ ,  $i = 1, \dots, N$ . In this section, we use the Bracket, or Dirac, notation in which the wavefunction is represented by a state vector in a complex vector space.

All the physical information of the system are inside the wavefunction, but itself the wavefunction has no physical meaning. What has a physical meaning is the square norm,  $\langle \Psi(\mathbf{r}, t) | \Psi(\mathbf{r}, t) \rangle = |\Psi(\mathbf{r}, t)|^2$  that is the probability density of finding each particle at a given point and at a given time. The other physical properties can be obtained from the wavefunction by applying the proper operators. In the realm of quantum mechanics, any measurable property has an associated linear operator  $\hat{A}$ , which is Hermitian, *i.e.*

$\hat{A}^\dagger = \hat{A}$ . In the case of  $N$  charged particles,  $\hat{A}$  becomes:

$$\hat{\mathcal{H}} = - \underbrace{\sum_{i=1}^N \frac{\hbar^2}{2m_i} \nabla_i^2}_{\hat{T}} + \underbrace{\sum_{i=1}^N \sum_{j>i}^N \frac{q_i q_j}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}}, \quad (2.1)$$

where the first term is the kinetic energy  $\hat{T}$ , with  $m_i$  being the mass of the  $i$ -th particle and  $\nabla^2$  is the Laplacian operator acting on the coordinates of the  $i$ -th particle. The second term is the potential energy  $\hat{V}$ , including the sum of pair-wise particle Coulomb's interactions.

Since Eq. 2.1 has no explicit dependence on time, the wavefunction can be factorized as  $|\Psi(\mathbf{x}_i, t)\rangle = |\Psi(t)\rangle |\Psi(\{\mathbf{x}_i\})\rangle$ , where can be easily demonstrated that  $|\Psi_t(t)\rangle = A \exp(-iEt/\hbar)$ . This latter quantity is an oscillating term and therefore it carries no information on the physics of the system, as in fact the  $|\Psi(\mathbf{r}, t)|^2$  is time independent. Therefore one can keep only the space-dependent part, reaching the time-independent (non-relativistic) Schrödinger equation:

$$\hat{\mathcal{H}} |\Psi(\{\mathbf{x}_i\})\rangle = E |\Psi(\{\mathbf{x}_i\})\rangle ,$$

from which the energy is obtained through the expectation value of the Hamiltonian, *i.e.*  $E = \langle \Psi | \hat{\mathcal{H}} | \Psi \rangle$ . Molecular systems consist of two different types of particles: electrons and nuclei. For a system with  $n$  electrons and  $N$  nuclei, the kinetic energy contributions can be separated into the electronic and the nuclear terms,  $\hat{T}_e$  and  $\hat{T}_n$ . Regarding the potential energy term, there are crossed terms: electron-electron, nucleus-nucleus and electron-nucleus potential energy terms ( $\hat{V}_{ee}$ ,  $\hat{V}_{nn}$ ,  $\hat{V}_{en}$ ). The last term couples electrons and nuclei, making the system not separable into electron-dependent and nuclei-dependent wavefunctions. Consequently, the molecular Hamiltonian (in atomic units) is:

$$\begin{aligned} \hat{\mathcal{H}} = & - \underbrace{\sum_{i=1}^n \frac{1}{2} \nabla_i^2}_{\hat{T}_e} - \underbrace{\sum_{k=1}^N \frac{1}{2m_k} \nabla_k^2}_{\hat{T}_n} + \underbrace{\sum_{i<j}^n \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_{ee}} + \\ & + \underbrace{\sum_{k<l}^N \frac{Z_k Z_l}{|\mathbf{r}_k - \mathbf{R}_l|}}_{\hat{V}_{nn}} - \underbrace{\sum_{i=1}^n \sum_{k=1}^N \frac{Z_k}{|\mathbf{r}_i - \mathbf{R}_k|}}_{\hat{V}_{en}}, \end{aligned} \quad (2.2)$$

where  $Z_k$  and  $m_k$  is the atomic number and the atomic mass of the  $k$ -sim nucleus,  $\mathbf{r}$  and  $\mathbf{R}$  are the electronic and nuclear coordinates respectively.

The non-separability of the wavefunction into an electronic and a nuclear part limits the applicability of Eq. 2.2 for a multi-electron system, for this reason some additional approximations have to be introduced.

**The Born-Oppenheimer approximation** The Born-Oppenheimer (BO) approximation, called also “adiabatic approximation”, is crucial to simplify many system in quantum chemistry. This approximation is based on the fact that nuclei are much heavier than electrons (about 1800 times more), consequently there is a time scale separation between electrons and nuclei motions. In this scenario, it can be assumed that the electrons experience the mean field of the nuclei, or in other words the electron Hamiltonian has a parametric dependency from the nuclei configuration. In this framework, the kinetic energy of the nuclei can be neglected and the nuclei-nuclei repulsion term can be considered to be a constant. The remaining terms determine the so called electronic Hamiltonian, which describes the motion of  $n$  electron in the field of  $N$  fixed point nuclear charges, as:

$$\hat{\mathcal{H}}_{elec} = - \sum_{i=1}^n \frac{1}{2} \nabla_i^2 - \sum_{i=1}^n \sum_{k=1}^N \frac{Z_k}{r_{ik}} + \sum_{i<j}^n \frac{1}{r_{ij}},$$

in which  $r_{ik} = |\mathbf{r}_i - \mathbf{R}_k|$  and  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ . Within this approximation, the electronic dynamics can be considered separated by the nuclear one. The latter only appears as a parameter inside the electronic wavefunction, as can be seen by the following equation:

$$\hat{\mathcal{H}}_{elec} |\psi(\mathbf{r}; \mathbf{R})\rangle = E_{elec} |\psi(\mathbf{r}; \mathbf{R})\rangle . \quad (2.3)$$

The  $|\psi(\mathbf{r}; \mathbf{R})\rangle$  are eigenfunctions of  $\hat{\mathcal{H}}_{elec}$  and therefore they form an orthonormal basis set in the electronic Hilbert space, *i.e.*:

$$\langle \psi_q(\mathbf{r}; \mathbf{R}) | \psi_s(\mathbf{r}; \mathbf{R}) \rangle = \delta_{qs} .$$

Assuming to be able to solve Eq. 2.3 for each nuclear configuration  $\{\mathbf{R}\}$ , the total molecular wavefunction  $|\Psi(\mathbf{r}; \mathbf{R})\rangle$  can be expanded on the basis of  $|\psi_q(\mathbf{r}; \mathbf{R})\rangle$ :

$$|\Psi(\mathbf{r}; \mathbf{R})\rangle = \sum_q \chi_q(\mathbf{R}) |\psi_q(\mathbf{r}; \mathbf{R})\rangle . \quad (2.4)$$

Inserting Eq. 2.4 into Eq. 2.3 and projecting onto the  $\psi_q$  basis set (integrating only on electronic coordinates), we can evaluate how the different terms of  $\hat{\mathcal{H}}$  act on  $\chi_q(\mathbf{R})$  and  $\psi_q(\mathbf{r}; \mathbf{R})$ . In particular the kinetic energy operator for the nuclei ( $-\sum_k \frac{\hbar^2}{2m_k} \nabla_k^2$ ) is not diagonal on the  $|\psi_q(\mathbf{r}; \mathbf{R})\rangle$  basis. The Born-Oppenheimer approximation now consists of neglecting the off diagonal terms

of this operator and therefore decoupling the electronic dynamics from the nuclear one: this can be assumed if the electronic energy difference between different Born-Oppenheimer surface is large enough (excited states). In general, this assumption is an extremely mild one, and it is entirely justified in most cases.

**The Pauli exclusion principle** All the particles that follow Fermi-Dirac statistics are called Fermion, and generally these particles have half-odd-integer spin ( $1/2$ ,  $3/2$ , etc.). Electrons belongs to this class of particles. Consequently, to the cited statistics, any observable derived from the wavefunction, like their probability density, must be invariant given any permutation of the electronic coordinates:

$$|\Psi(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots, \mathbf{r}_n)|^2 = |\Psi(\mathbf{r}_1, \dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots, \mathbf{r}_n)|^2 .$$

Introducing the permutation operator  $\hat{P}$  of parity  $p$ , which permutes two electrons in the wavefunction, we can formalize the following:  $\hat{P} |\Psi\rangle = (-1)^p |\Psi\rangle$ . This is the Pauli exclusion principle, which states that a  $n$ -electrons wavefunction must be antisymmetric with respect to the exchange of any two electrons. By consequence, any electron must differ from the other by at least one quantum number, in other words two electron can occupy the same orbital only if they have opposite half-integer spin projections of  $1/2$  and  $-1/2$ . Any fermionic wavefunction must obey this principle.

**The Variational principle** The variational principle is an important theorem in multi-body quantum chemistry, in which many computational techniques are rooted to achieve approximate solutions to the ground state. The variational theorem states that the expectation value of any normalized trial wavefunction ( $\Psi_t$ ) with respect to the energy represents an upper bound to the ground state energy ( $E_g$ ), in formula:

$$\frac{\langle \Psi_t | \hat{\mathcal{H}} | \Psi_t \rangle}{\langle \Psi_t | \Psi_t \rangle} = E_t \geq E_g .$$

In this framework, when the trial wavefunction is written as function of a set of parameters, the problem is to find the best approximation to the wavefunction for the ground state. In essence, it becomes an energy minimizing constraints problems for the wavefunction trial parameters. The constraints are imposed by the physical boundary conditions of the wavefunction.



**The Slater determinant** The definition of the wavefunction for a multi-electron system is not trivial, at first one could think about expressing it as the product of independent single electron spin-orbitals.  $|\chi(\mathbf{x})\rangle = |\Psi(\mathbf{r})\rangle |\omega(\sigma)\rangle$ , where the spin function  $|\omega(\sigma)\rangle$  can assume two values,  $|\alpha\rangle$  or  $|\beta\rangle$ . However this formulation does not respect the Pauli principle, it can be easily seen that the wavefunction is not antisymmetric respect to the exchange of any two electrons. A clever choice is provided by the so called Slater determinant:

$$\Psi(\mathbf{x}_1 \cdots \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \equiv |\chi_1 \cdots \chi_N| .$$

### 2.1.2 The Hartree-Fock method

The Hartree-Fock method (HF) is a method for finding approximated solutions to the multi-electronic Hamiltonian. It is based on the assumptions that each electron belongs to a single-particle functions (*spin-orbitals*) and behaves in the mean field generated by the presence of other electrons. The definition of the wavefunction through the Slater determinant is the starting point of the HF method. The second step is calculating the expectation value of the molecular Hamiltonian, under the BO approximation, of the wavefunction, that in formula yields:

$$\langle \Psi | \hat{\mathcal{H}} | \Psi \rangle = V_{nn} + \sum_{i=0}^N h_i + \sum_{i=0}^N \sum_{j>i}^N J_{ij} - K_{ij} ,$$

where  $h_i$  collects one-electron terms and  $J_{ij} - K_{ij}$  two-electrons ones. These terms are defined as:

$$\begin{aligned} h_i &= \langle \chi_i | \hat{\mathbf{T}}_e + \hat{\mathbf{V}}_{en} | \chi_i \rangle = \langle \chi_i | \hat{h} | \chi_i \rangle , \\ J_{ij} &= \langle \chi_i \chi_j | r_{ij}^{-1} | \chi_i \chi_j \rangle , \\ K_{ij} &= \langle \chi_i \chi_j | r_{ij}^{-1} | \chi_j \chi_i \rangle . \end{aligned}$$

The two latter equations come from electron-electron repulsion Hamiltonian term ( $\hat{\mathbf{V}}_{ee}$ ), where  $r_{ij}$  is the distance between the  $i$ -th and  $j$ -th electrons ( $|\mathbf{x}_i - \mathbf{x}_j|$ ).  $J_{ij}$ , called Coulomb integral, represents the electrostatic interaction between two electronic densities  $|\chi_i|^2$  and  $|\chi_j|^2$  while  $K_{ij}$ , called the exchange integral, arises from the antisymmetric nature of the wavefunction and do not has any classical meaning.  $J_{ij}$  and  $K_{ij}$  are positive, and therefore, in the  $i = j$  case they cancelled each other preventing any self-interaction spurious

effect (which would come from  $J_{ii}$  elements). These two matrix elements can be further expressed as single-electron operators, defined by the spin-orbitals themselves:

$$\begin{aligned}\hat{J}_j \chi_i(\mathbf{x}) &= \chi_i(\mathbf{x}) \int \frac{|\chi_j(\mathbf{x}')|^2}{|\mathbf{x} - \mathbf{x}'|} d\mathbf{x}' , \\ \hat{K}_j \chi_i(\mathbf{x}) &= \chi_j(\mathbf{x}) \int \frac{\chi_j^*(\mathbf{x}') \chi_i(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} d\mathbf{x}' .\end{aligned}$$

From this expression, it is now clear why the HF theory is a mean-field theory, *i.e.* each electron feels the averaged effect of the rest by means of these one-electron operators. Collecting these two operators and the mono-electronic term, the Fock operator  $\hat{F}$  is obtained, which may be defined for a given electronic spin value  $\sigma$ , either up or down ( $\alpha, \beta$ ), as:

$$\hat{F}^\sigma |\chi_i^\sigma\rangle = \left[ \hat{h} + \sum_{j,\sigma'} \hat{J}_j^{\sigma'} - \delta_{\sigma\sigma'} \hat{K}_j^\sigma \right] |\chi_i^\sigma\rangle = \varepsilon_i^\sigma |\chi_i^\sigma\rangle ,$$

where  $\varepsilon_i^\sigma$  is the energy associated with the  $i$ -th spin-orbital with spin  $\sigma$ . These are the unrestricted HF (UHF) equations, in which each electron is allowed to occupy a different spatial orbital thanks to the separation into the electron's spin. Nevertheless the two sets of orbitals are still coupled by the Coulomb's term. For a closed-shell molecule, *i.e.* same number of spin  $\alpha$  and  $\beta$  electrons, the previous equation simplified to

$$\hat{F} |\chi_i\rangle = \left[ \hat{h} + \sum_j \hat{J}_j - \hat{K}_j \right] |\chi_i\rangle = \varepsilon_i |\chi_i\rangle ,$$

in which each spin-orbital is filled by two electrons, one spin up and one spin down. This gives rise to the restricted HF (RHF) theory. One could also take a midway by forcing orbitals to behave as RHF as far as possible and the remaining ones as UHF, which is the restricted open-shell HF (ROHF) method. RHF and ROHF wavefunctions are also eigenfunctions of the total squared spin operator  $\hat{S}^2$ , while the UHF is not. Consequently, the UHF can suffer from spin contamination issues contrary to the ROHF method.

The core of the HF method is minimizing the expectation value of the Fock operator, based on the variational principle, with respect to the coefficients of a finite basis set of atomic coefficients in which each  $\chi$  are expanded. The method is defined "Self-Consistent Field", as the minimization is an iterative procedure until the solution satisfied a stop criteria, *e.g.* the energy difference between the  $i$  and  $i+1$  minimization steps is less than a given threshold.

### 2.1.3 Post Hartree-Fock methods

Despite the importance the Hartree-Fock method as an electronic solution method, it lacks the “electronic correlation” energy. This is due to the mean field approximation (or independent-particle approximation), as each electron are treated independently and only feels the averaged effect of the remaining electrons. In this approach, the information of pair- and higher-electron correlation are not taken into account. In accordance with the variational principle, the effect of neglecting these terms is to raise the energy above the exact energy. Consequently, the correlation energy is defined as the magnitude of the energy difference between the exact (non-relativistic) energy and the Hartree-Fock energy:

$$E_{corr} = E_{exact} - E_{HF} .$$

Electron correlation can be classified in two ways: dynamic and static correlation. Dynamic correlation takes into account the instantaneous correlation of electrons while the static one is associated to near-degenerate states (spin-orbitals of similar energy). The dynamic correlation is more important for electrons in the same orbital, while static one is more important between electrons in different orbitals.

In order to recover the  $E_{corr}$  many methods, defined as post Hartree-Fock, were proposed during the years. In the framework of this thesis we will discussed only the Coupled Cluster method and its DLPNO variant.

**Coupled cluster theory** The initial formulation of coupled cluster (CC) theory was for the quantum-chemical treatment of the nuclear matter. The electron correlation treatment and high-accuracy computations made use of one of the most powerful quantum chemistry techniques when it was embedded into the theory of electronic structure.

The underlying assumption of the CC theory is the definition of the wave function as follows:

$$|\Psi_{CC}\rangle = \exp(\hat{T}) |\Psi_{HF}\rangle , \quad (2.5)$$

where  $\hat{T}$  is the cluster operator, which is an excitation operator and consists of the weighted sum of all excitations, *i.e.*

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_{N_{elec}} .$$

$\hat{T}_1 + \hat{T}_2 + \dots$  are the weighted sums of single, double, etc.. excitations with the unknown parameters given by the weighting coefficients that are usually referred to as amplitudes. The CC wavefunction (Eq. 2.5) is often not

computed using the variational principle because of the exponential ansatz. Instead, one employs a projection method in which the CC wavefunction is added to the Schrödinger equation and a projection onto the reference determinant is used to derive a formula for the energy:

$$E = \langle \Psi_{HF} | \exp(-\hat{T}) \hat{\mathcal{H}} \exp(\hat{T}) | \Psi_{HF} \rangle .$$

Nonlinear equations for the amplitudes are then obtained by projection onto the excited determinants  $|\Psi_Q\rangle$

$$0 = \langle \Psi_Q | \exp(-\hat{T}) \hat{\mathcal{H}} \exp(\hat{T}) | \Psi_{HF} \rangle ,$$

which has to be solved for every possible  $|\Psi_Q\rangle$ . Only when used in conjunction with a truncated cluster operator does this theory reveal its advantages. Starting from the CCSD most common implementation (CC singles and doubles, with  $\hat{T} = \hat{T}_1 + \hat{T}_2$ ), the CCSD(T) (where third order excitations are added perturbationally) variant of CCSDT (CC singles, doubles, triples, with  $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$ ) reaches the best balance between accuracy and computational cost. In CCSD(T) scheme, the triples contribution is calculated separately using the fourth-order perturbation theory and added to the CCSD results. An additional term that describes the coupling between singles and triples and originates from the fifth-order perturbation is also included.

### 2.1.4 DLPNO-CCSD(T) method

The “gold standard” model of computational chemistry is the CCSD(T) method for the high accuracy and ability to recover a large fraction of the electron correlation. Non-covalent interactions (hydrogen bonds, dispersion interactions, etc..) can be well reproduced by wavefunction methods if correlation is recovered, as it is in CCSD(T). The drawback of CCSD(T) is the increasing computational cost with respect to the system size, which scales roughly as  $\mathcal{O}(N^7)$ . In this framework, the Domain Local Pair Natural Orbital (DLPNO) approximation to CCSD(T) (Riplinger et al., 2016; Guo et al., 2018) method allows to reach an almost linear scaling  $\mathcal{O}(N)$  behaviour in the computational cost, while keeping the accuracy of CCSD(T). In few words, this method, instead of the canonical delocalized orbitals, used localized pair natural orbitals. Consequently, these orbitals can be classified into domains for proper sorting and selection of the most important excitations accounting for electronic correlation.

The most important features of this technique are summarized below, following the implementation in the program ORCA (see Neese, 2018). The starting point in the DLPNO-CCSD(T) method is the determination and

localization of the Hartree–Fock (HF) occupied orbitals. The correlation energy ( $E_{corr}$ ), is then expressed as a sum of electron pair correlation energies  $\epsilon_{ij}$ , where  $i$  and  $j$  referred to the occupied HF orbitals, as in the formula:

$$E_{corr} = 2 \sum_{i, \tilde{a}_{ii}} F_{i, \tilde{a}_{ii}} t_{\tilde{a}_{ii}} + \sum_{i \geq j} \epsilon_{ij} + E^{(T)}. \quad (2.6)$$

The first term represents the contribution from the single excitations and vanishes if the Brillouin’s theorem is satisfied. In the pair natural orbitals (PNOs) approximation, the expression for the pair energy is:

$$\epsilon_{ij} = \sum_{\tilde{a}_{ij}, \tilde{b}_{ij}} (i\tilde{a}_{ij} | j\tilde{b}_{ij}) \tilde{\tau}_{\tilde{a}_{ij}, \tilde{b}_{ij}}^{ij},$$

where  $\tilde{a}_{ij}$  and  $\tilde{b}_{ij}$  are PNOs that belong to pair  $ij$ ,  $(i\tilde{a}_{ij} | j\tilde{b}_{ij})$  are the two-electron integral in Mulliken (11|22) notation. The cluster amplitudes in the PNO basis are:

$$\tau_{\tilde{a}_{ij}, \tilde{b}_{ij}}^{ij} = t_{\tilde{a}_{ij}, \tilde{b}_{ij}}^{ij} + t_{\tilde{a}_{ij}}^i t_{\tilde{b}_{ij}}^j,$$

where  $t_{\tilde{a}_{ij}, \tilde{b}_{ij}}^{ij}$ ,  $t_{\tilde{a}_{ij}}^i$  and  $t_{\tilde{b}_{ij}}^j$  denote the doubles and the singles amplitudes of the coupled cluster equations. Meanwhile:

$$\tilde{\tau}_{\tilde{a}_{ij}, \tilde{b}_{ij}}^{ij} = \frac{2}{1 + \delta_{ij}} (2\tau_{\tilde{a}_{ij}, \tilde{b}_{ij}}^{ij} - \tau_{\tilde{b}_{ij}, \tilde{a}_{ij}}^{ij})$$

are the contravariant amplitudes. In Eq. 2.6, the singles amplitudes are expanded by the singles PNOs which are essentially a set of diagonal PNOs truncated by a much tighter threshold than the doubles PNOs.

The pair correlation energies decay very quickly with the distance  $R_{ij}$  between the centers of the localized occupied orbitals  $i$  and  $j$ , *i.e.* at least as  $R_{ij}^{-6}$ . This, with other criteria, permits the discrimination between “*weak pairs*”, with negligible contributions to the correlation energy treated at MP2 level, and “*strong pairs*”, treated at CCSD level. The external correlation space of the latter pairs are spanned by an highly compact set of PNOs that are localized, different for each electron pair, and represent the dominant part of the overall correlation energy. The last term in Eq. 2.6 corresponds to the perturbative triples contribution to the correlation energy, which involves a combination of weak and strong pairs.

The main source of error of the DLPNO-CCS(T) method, excluding the incompleteness basis set error, lies mainly in the choice of the two parameters controlling the size of the correlation space. A solution can be obtain using an easy extrapolation procedure, similar, but conceptually different, from

the two-point CBS strategy, to recover the Complete PNO Space (CPS) and correct the additive error on the electron pair truncation correlation energy (Altun et al., 2020). The size of the correlation space is controlled by two parameters:  $T_{CutPairs}$  and  $T_{CutPNO}$ .  $T_{CutPairs}$  defines weak and strong pairs.  $T_{CutPNO}$  controls the dimension of the compact virtual space, in function of pair natural orbitals (PNOs), for which the coupled cluster (strong pairs) equations are solved. The equation used to recover the correlation energy in the CPS extrapolation limit ( $E_{cps}$ ) is:

$$E_{cps} = E^X + 1.5(E^Y - E^X) ,$$

where  $E^X$  and  $E^Y$  are the correlation energies with  $T_{CutPNO} = 10^X$  and  $10^Y$  respectively and  $Y = X + 1$ . To calculate the total absolute energy the  $E_{cps}$  must be added to the DLPNO-CCSD energy, which is indifferent to the PNO cutoff.

### 2.1.5 ONIOM method

Despite electronic structure methods having improved in recent decades, it is still challenging and extremely time-consuming (as well as disk/memory-consuming) to accurately predict the structure and characteristics of large molecular systems using QM methods alone. Several attempts have been made to make such calculations possible. One solution is the use of a hybrid approach, in which a large molecular system is divided into several fragments. Then, different theoretical methods are used on the various fragments, ranging from very expensive and accurate to less expensive and accurate techniques.

The most commonly used hybrid methods can be divided into two categories: “*additive*” and “*subtractive*” schemes. Here, the difference between the two schemes is presented with a system divided into two fragments (A and B) and each of them studied with two different electronic methods ( $M_a$  and  $M_b$ ). The energy calculated with an “*additive*” method is:

$$E_{tot} = E_A^{M_a} + E_B^{M_b} + E_{A-B} .$$

Here, the difficulty lies in determining the  $E_{A-B}$  term describing the interactions at the interface between the two subsystems. Normally, these interactions are based on Lennard-Jones and electrostatic potentials. On the other hand, the energy calculated with a “*subtractive*” method, in which B is the target system, is:

$$E_{tot} = E_{A+B}^{M_a} + E_B^{M_b} - E_B^{M_a} . \quad (2.7)$$

In this case, the entire system (A+B) must be calculated at  $M_a$  level, consequently,  $M_a$  must be less computationally demanding than  $M_b$  to reduce the computational cost. From the above equations it is clear why the “*subtractive*” scheme is also called “*extrapolative*”, as extending the B zone to become A or adopting the  $M_b$  method also for the A zone corresponds to treat the whole system with the most accurate  $M_b$  method.

Belongs to this latter scheme the ONIOM method (Svensson et al., 1996; Humbel et al., 1996; Chung et al., 2015). Its name stands for “*Our own N-layered Integrated molecular Orbital and Molecular mechanics*” and enables different *ab initio*, semi-empirical or classical mechanics-based methods to be combined to different parts of a system to give reliable energy, geometry and vibrational frequencies at a reduced computational cost. Furthermore, ONIOM can be used as a two-layer method (two subsystem fragments, ONIOM2) or can easily be extended to multiple layers (more than two).

In a two-layer setup, where the B system is studied with a more accurate  $M_b$  method (high-level layer) and A is the rest of the whole system where a less accurate  $M_a$  method is used (low-level layer), the total energy (E), gradient vector ( $\mathcal{G}$ ) and Hessian matrix ( $\mathcal{H}$ ) are:

$$\begin{aligned} E &= E_{A+B}^{M_a} + E_B^{M_b} - E_B^{M_a} , \\ \mathcal{G} &= \mathcal{G}_{A+B}^{M_a} \times \mathcal{J} + \mathcal{G}_B^{M_b} \times \mathcal{J} - \mathcal{G}_B^{M_a} \times \mathcal{J} , \\ \mathcal{H} &= \mathcal{H}_{A+B}^{M_a} \times \mathcal{J} + \mathcal{J}^T \times \mathcal{G}_B^{M_b} \times \mathcal{J} - \mathcal{J}^T \times \mathcal{G}_B^{M_a} \times \mathcal{J} , \end{aligned}$$

where  $\mathcal{J}$  is the Jacobian matrix between the Model A and the B system. The role of the low-level method in ONIOM2 is to reproduce the difference between the whole (A+B) and the more accurate (B) systems, consequently, more accurate results are expected when the low-level theory give results close to the high-level and can well reproduce the interactions at the interface with the high-level zone. On the other hand, another important aspect limiting the accuracy is the size of the high-level zone (Chung et al., 2015; Morokuma et al., 2006; Das et al., 2018; Chen et al., 2021).

## 2.2 Graph theory for molecules and reaction network

This section briefly introduces some fundamentals about Graph theory and the application used in this thesis. More information can be found in standard network science (Barabási, 2013; Newman, 2018) textbooks.

The foundations of graph theory is well illustrated by the “*Seven Bridges of Königsberg*” mathematical problem. In the 18th century Königsberg, the

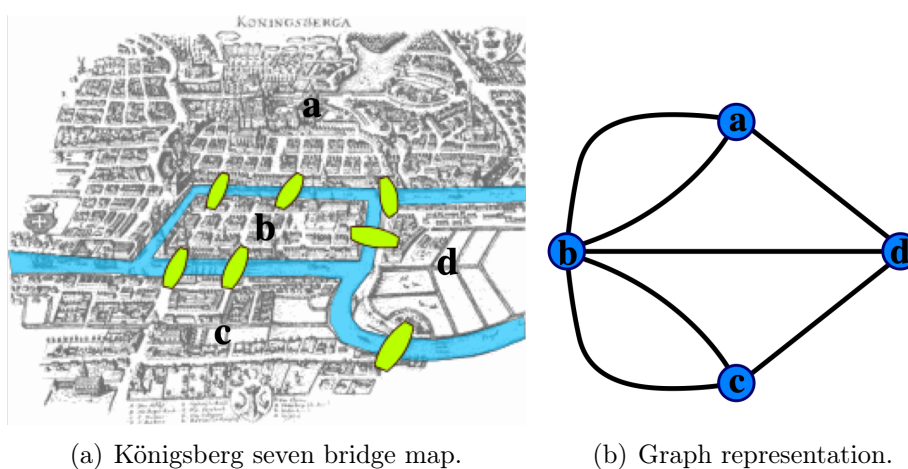


Figure 2.1: Map of Königsberg with the seven bridge highlighted and its graph representation. (Credit: Bogdan Giușcă - <https://commons.wikimedia.org/w/index.php?curid=112920>)

capital of Eastern Prussia and the modern Kaliningrad, had a special arrangement. Seven bridges across the river Pregel surrounded the town, five of these bridge connect the mainland to Kneiphof island meanwhile the other two crossed the branches of the river (left panel of Fig. 2.1). This peculiar structure gave birth to a mathematical puzzle: *Can one walk across all seven bridges and never cross the same one twice?*

Leonard Euler proved in 1736 that the problem has no solution, *i.e.* no such path exist. Euler used the letters a, b, c, and d to symbolize each of the four geographical regions divided by the river. He then connected each piece of land with a bridge between them using lines (right panel of Fig. 2.1). Thus, he created a network in which the nodes represented lands and the edges represented bridges. Euler then made a straightforward deduction: if a path exists that crosses every bridge, but never the same bridge twice, then nodes with odd numbers of connections must be either the starting or the end point of this path. A walking path through all bridges can only have one beginning and one ending. In fact, you could find that there is not a free link to exit if you arrive to a node with an odd number of links. The Euler solution was the first time in which a graph was used to solve a mathematical problem.

**The mathematical representation** To fully describe a graph all the nodes and the links (edges) should be provided. The simplest way is to represent a graph through its adjacency matrix. The adjacency matrix of a



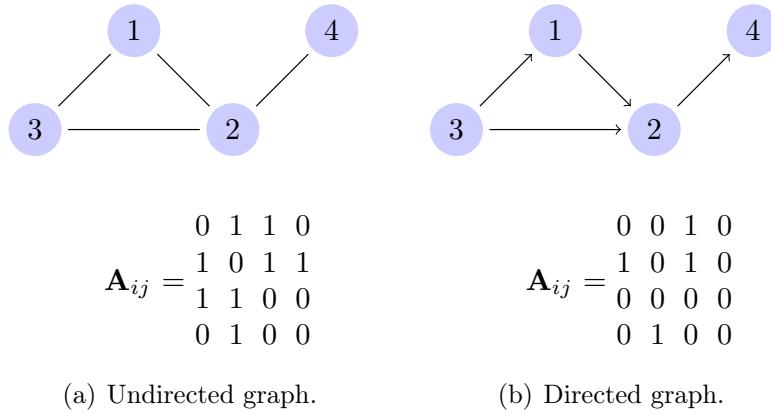


Figure 2.2: Graphs and related adjacency matrices of a directed and undirected graph.

graph of  $N$  nodes has  $N$  rows and  $N$  columns:

$A_{ij} = 1$  if there is a link from node  $j$  to node  $i$  ;

$A_{ij} = 0$  if nodes  $i$  and  $j$  are not connected to each other.

This is the condition for an undirected graph, where the edges do not have a direction, and consequently the adjacency matrix is symmetric,  $A_{ij} = A_{ji}$ . In the case of a directed graph, the adjacency matrix is not symmetrical and the first condition above is reformulated as:

$A_{ij} = 1$  if there is a link pointing from node  $j$  to node  $i$  ;

Example of a directed and undirected graph, with the related adjacency matrices, are shown in Fig. 2.2. An important property of the graph is the degree of a node, that represents the number of links it has to other nodes. The degree  $k$  of a node  $i$ , *i.e.*  $k_i$ , can be easily computed via the adjacency matrix. For undirected graph the  $i$ -node degree is the sum over either the row  $i$  or the column  $i$  of the matrix:

$$k_i = \sum_{j=1}^N A_{ij} = \sum_{j=1}^N A_{ji} .$$

Meanwhile, for a directed graph the sums over the row  $i$  and column  $i$  provide respectively the incoming and outgoing degrees:

$$k_i^{in} = \sum_{j=1}^N A_{ij} ,$$

$$k_i^{out} = \sum_{j=1}^N A_{ji} .$$

In the case of the node 2 in Fig. 2.2, the degree for the undirected graph is  $k_2 = 3$  meanwhile the incoming and outgoing degrees for the directed graph are  $k_2^{in} = 2$  and  $k_2^{out} = 1$  respectively.

Many extensions can be formulated to the standard directed/undirected graph formalism in order to fully represent the specific processes/object (Energy Network, social media interactions, Reaction Network, etc.). Furthermore, almost any property can be inferred or obtained numerically through the adjacency matrix, edges and nodes properties. In the following sections it is illustrated how to use the graph theory to deal with problems related to molecules and reaction networks.

### 2.2.1 Molecules

Molecules can be represented as graphs, the so-called chemical graphs (Trinajstić, 2018). One application is related to the Machine Learning or Deep Learning field, where molecular graph representations can be easily handled increasing the accuracy of these models (Coley et al., 2019). Another simpler application, developed during this thesis and described below, shows the power of such chemical graphs codification in the constitutional isomers discrimination.

A chemical graph is a non-directed graph where atoms and bonds in molecules correspond to nodes and edges, respectively. The different chemical elements present in the periodic table are represented in graphs as colors assigned to the vertices: the graph is, therefore, defined as a multi-colored graph (a chemical graph is presented in Fig. 2.3). Depending on the objective, information on the bond distances or atoms characteristics can be added to the edges and nodes properties. The codification normally passes via the conversion from a .XYZ (coordinate) molecular file<sup>1</sup>. This process is possible using the covalent radii and distance functions, *e.g.* using the functions implemented in the Atomic Simulation Environment (ASE) python

<sup>1</sup>[https://en.wikipedia.org/wiki/XYZ\\_file\\_format](https://en.wikipedia.org/wiki/XYZ_file_format)

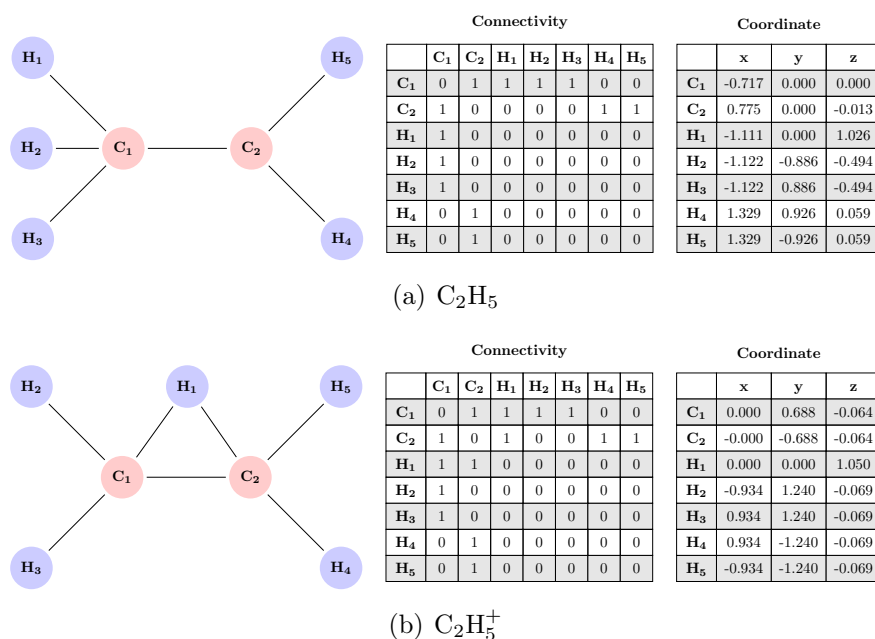


Figure 2.3: Graph, connectivity matrix and coordinates (in Å) of neutral ethyl radical  $C_2H_5$  and its cation  $C_2H_5^+$ . (Credit: Tinacci et al. (2021))

package (Larsen et al., 2017). The Python 3.8 script used to create a possible molecules (in .XYZ format) starting from a molecular graph is here reported<sup>2</sup>:

```

1 from rdkit import Chem
2 from rdkit.Chem import AllChem
3 import networkx as nx
4 from ase import io, neighborlist, atoms
5
6 #Function that encode the molecules from graph object to XYZ
  format
7 def MolFromGraphs(G):
8     '''
9     Function that takes as input the networkx graph (each node
    must have an atom property H,N,C,O etc) and return the mol
    (rdkit) object
10    the function dont discriminate between different type of
    bond, it care only about the connectivity.
11    '''
12    adjacency_matrix = nx.convert_matrix.to_numpy_matrix(G,
    weight='bond').tolist()
13    node_list = []
14    for i,node in enumerate(G):
15        node_list.append(G.nodes[i]['atom'])

```

<sup>2</sup><https://github.com/TinacciL/FromGraphToXYZ>

```

16 #Create empty editable mol object
17 mol = Chem.RWMol()
18 #Add atoms to mol and keep track of index
19 node_to_idx = {}
20 for i in range(len(node_list)):
21     a = Chem.Atom(node_list[i])
22     molIdx = mol.AddAtom(a)
23     node_to_idx[i] = molIdx
24 #Add bonds between adjacent atoms
25 for ix, row in enumerate(adjacency_matrix):
26     for iy, bond in enumerate(row):
27         #Only traverse half the matrix
28         if iy >= ix:
29             break
30         #Add relevant bond type (there are many more of these)
31         if bond == 0:
32             continue
33         elif bond == 1:
34             bond_type = Chem.rdchem.BondType.SINGLE
35             mol.AddBond(node_to_idx[ix], node_to_idx[iy],
36 bond_type)
37         elif bond == 2:
38             bond_type = Chem.rdchem.BondType.DOUBLE
39             mol.AddBond(node_to_idx[ix], node_to_idx[iy],
40 bond_type)
41         elif bond == 3:
42             bond_type = Chem.rdchem.BondType.TRIPLE
43             mol.AddBond(node_to_idx[ix], node_to_idx[iy],
44 bond_type)
45 mol.UpdatePropertyCache(strict=False)
46 #Add the possibility in N bearing species cations to
47 have valence 4 for the N center
48 for at in mol.GetAtoms():
49     if at.GetAtomicNum() == 7 and at.GetExplicitValence() == 4
50     and at.GetFormalCharge() == 0:
51         at.SetFormalCharge(1)
52 Chem.SanitizeMol(mol)
53 return mol
54
55 #INPUT - Change the above line in order to save the computed
56 .xyz in the choose path
57 path = '/Users/aaa/Documents/FromGraphToXYZ/'
58 #INPUT - Name of the molecules for the .xyz file
59 name = 'CH4'
60 #INPUT - Ordinated list of atoms in the choose molecule
61 atoms = ['H', 'H', 'H', 'H', 'C']
62 #INPUT - Ordinated list of bond in the choose molecule (1 =
63 sigma, 2 = pi, 3 = 2 * pi)
64 bonds = [1,1,1,1]

```

```

58 #INPUT - Ordinated order of Connettivity between list of atom
59 edges = [[0,4],[1,4],[2,4],[3,4]]
60
61 #Upload all the info in graph object
62 connectivity = []
63 for i,item in enumerate(edges):
64     connectivity.append((edges[i][0],edges[i][1], {'bond':
65         bonds[i]}))
66 n_dim = len(atoms)
67 G = nx.Graph()
68 for i,atom in enumerate(atoms):
69     G.add_node(i)
70     tmp_attr = {'atom': atom}
71     G.nodes[i].update(tmp_attr.copy())
72 G.add_edges_from(connectivity)
73 #Encode the molecules from graph object to XYZ format
74 tmp_mol = MolFromGraphs(G)
75 #Read the molecules in Rdkit
76 AllChem.EmbedMolecule(tmp_mol)
77 #Optimized the guessed structure with UFF in Rdkit
78 AllChem.UFFOptimizeMolecule(tmp_mol)
79 #Save the .XYZ file of the optimized structure
80 with open(path + name + '.xyz', "w+") as file_mol:
81     file_mol.write(Chem.MolToXYZBlock(tmp_mol))

```

Using .XYZ molecular file makes it difficult to discriminate between two constitutional isomers due to the non-univocal atoms order inside the file and the possible atom permutations. A way to overcome such a problem is using a property of the graph, *i.e.* the isomorphism. Two graphs that contain the same number of graph colored-vertices connected in the same way are considered isomorphic, consequently two molecules are not constitutional isomers if their molecular graphs satisfied such conditions. An example of two chemical graphs that are not isomorphic is shown in Fig. 2.3. Here below is presented the Python 3.8 script used to encode molecules to graphs and perform an isomorphism control (the NetworkX python package (Hagberg et al., 2008) is used to deal with graph objects and properties<sup>3</sup>):

```

1 import networkx as nx
2 from ase import io, neighborlist, atoms
3
4 #function that encode the molecules from xyz file to graph
  object
5 def FromXYZtoGraph(input_file):
6     atoms = ['H','He','Li','C','N','O','F','Na','Si','P','S',
7         'Cl']
8     atomic_num = [1,2,3,6,7,8,9,11,14,15,16,17]

```

<sup>3</sup>[https://github.com/TinacciL/Connectivity\\_chemical\\_graph](https://github.com/TinacciL/Connectivity_chemical_graph)

```

8     mol = io.read(input_file)
9     #compute neighbor of the atoms in xyz format
10    cutOff = neighborlist.natural_cutoffs(mol)
11    neighborList = neighborlist.NeighborList(cutOff,
12    self_interaction=False, bothways=True)
13    neighborList.update(mol)
14    #compute adjacency matrix and atoms list
15    adj_matrix = neighborList.get_connectivity_matrix(sparse=False)
16    Natom_list = mol.get_atomic_numbers()
17    atoms_list = []
18    for i,item in enumerate(Natom_list):
19        for k in range(len(atomic_num)):
20            if item == atomic_num[k]:
21                atoms_list.append(atoms[k])
22    #convert in networkx-molecules graph
23    G=nx.from_numpy_matrix(adj_matrix)
24    for i,item in enumerate(atoms_list):
25        tmp_attr = {'atom': item}
26        G.nodes[i].update(tmp_attr.copy())
27    return(G)
28 #encode the molecules from xyz file to graph object
29 mol_0 = FromXYZtoGraph(pathToMolecules/mol_0.xyz)
30 mol_1 = FromXYZtoGraph(pathToMolecules/mol_1.xyz)
31
32 #test the isomorphism
33 prop = 'atom'
34 nm = nx.algorithms.isomorphism.categorical_node_match(prop,
35 prop)
36 if nx.is_isomorphic(mol0,mol_1,node_match=nm):
37     print('Are isomorphic!')
38 else:
39     print('Are NOT isomorphic!')

```

Another application of graph theory in chemistry is the possibility to create all possible constitutional isomers starting from the number and type of atoms in a molecule. This application can be achieved by using two graph properties: the isomorphism and the definition of a connected graph<sup>4</sup>. A non-optimized program was developed in the framework of this thesis. It can be found at the following link: [https://github.com/TinacciL/Isomers\\_generator](https://github.com/TinacciL/Isomers_generator).

<sup>4</sup>A graph is connected if there is a path from any point to any other point in the graph.

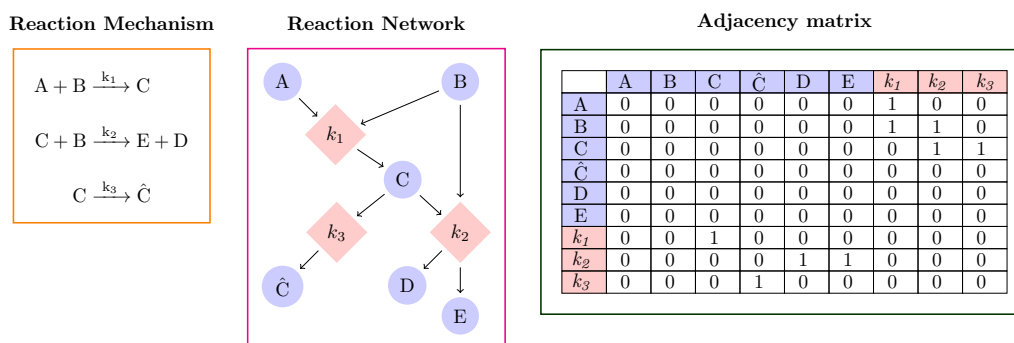


Figure 2.4: Three different ways to visualize and work with a reaction network: reactions list (*left panel*), colored directed graph (or colored di-graph), where circles and squares are species and rate constants, respectively (*left panel*), and non symmetric adjacency (or connectivity) matrix representation of the colored digraph (*right panel*). (Credit: XXX)

## 2.2.2 Reaction Network

As stated in its own name, a reaction network is a network that can be encoded as a graph. With this codification, the reaction network can be easily handled and many properties can be obtained (*e.g.* node connectivity, node centrality, sinks, etc.). Different graphs types can be used to codify reaction networks, *e.g.* multi-directed-graph, colored directed graph, hypergraph etc.. In the framework of this thesis, a colored directed graph (also known as a colored di-graph) is used. In a colored di-graph, each node has a color attribute and the connections are directed. A colored node can have different attributes depending on the color: in our case, it can be a species-node or a rate constant-node. In addition, in a directed graph (as described in paragraph 2.2), the edges have a direction, *i.e.* the adjacency (or connectivity) matrix is not symmetric, so that the connection between nodes is oriented. Fig. 2.4 shows three ways to represent a generic reaction network: the simple reaction list, the reaction network visualization and, finally, the corresponding adjacency matrix of the colored directed graph.

The codification of a reaction network with graph theory has show very powerful in dealing with complex reaction pathways computed with electronic structure calculations. An example can be found in Garay Ruiz et al. (2022) and McDermott et al. (2021) articles.

A code to encode an Astrochemical reaction network (in KIDA<sup>5</sup> formalism) using a Python 3.8 script and NetworkX python package (Hagberg et al.,

<sup>5</sup>[https://github.com/TinacciL/GreToBaPe\\_Cleaning](https://github.com/TinacciL/GreToBaPe_Cleaning)

2008) is shown below:

```

1 import pandas as pd
2 import numpy as np
3 import networkx as nx
4 #The following function can be found in https://github.com/
   TinacciL/GreToBaPe_Cleaning Github repository
5 from lib import FromNetworkDataFrameToNetworkClassList,
   ExtractMolFromNetwork, FromNetworkKidaDATtoCSV
6 #Encode KIDA .dat format into pandas dataframe
7 df_net = FromNetworkKidaDATtoCSV(path_to_network,save=False)
8 #Extract all the molecules inside the Network inside a
   dataframe and then to a vector
9 df_mol = ExtractMolFromNetwork(df_net)
10 mol_vec = df_mol['species'].to_numpy()
11 mol_prop = np.zeros(len(mol_vec),dtype=float)
12 #Add properties to species (EXAMPLE)
13 for i in range(len(mol_vec)):
14     mol_prop[i] = len(mol_vec[i])
15 df_mol['properties'] = mol_prop
16 #Network information and adding fake energy to Network (
   EXAMPLE)
17 net_id = df_net['Number'].to_numpy()
18 net_rec = np.zeros(len(net_id),dtype=float)
19 net = FromNetworkDataFrameToNetworkClassList(df_net)
20 #Define the network type
21 net_g = nx.DiGraph()
22 #create nodes and add attribute
23 for i,mol in enumerate(mol_vec):
24     net_g.add_node(mol)
25     tmp_attr = {'type': 'species','energy': mol_prop[i]}
26     net_g.nodes[mol].update(tmp_attr.copy())
27 #create edges and attribute
28 for i,rec in enumerate(net):
29     net_g.add_node(net_id[i])
30     tmp_attr = {'type': 'reaction','energy': net_rec[i]}
31     net_g.nodes[net_id[i]].update(tmp_attr.copy())
32     for j,item_j in enumerate(set(net[i].reactants)):
33         net_g.add_edge(item_j,net_id[i])
34     for k,item_k in enumerate(set(net[i].products)):
35         net_g.add_edge(net_id[i],item_k)

```



Part II

Astrochemical Gas-Phase  
Network

# Introduction

**Context** Gas-phase reactions are of fundamental importance for predicting molecular abundances in astrochemical models. The role and focus that this sub-reaction network has in astrochemistry can be quantified by the  $\sim 8000$  reactions and  $\sim 500$  species present in the KIDA database<sup>1</sup>, which is from my point of view the most up-to-date database of chemical information in astrochemistry. Astrochemists have gradually added new reactions as new species have been detected, increasing the complexity of the reaction network. Within this framework, I realized that almost all the attention is focused on explaining the newly identified species, focusing on the formation reactions of these species or their precursors, while the backbone of the reaction network (the oldest part) has less consideration.

The growing number of reactions collaterally increases the number of species. These species are constituted mainly by positive ions and neutral species. Of the entire number of species present in the astrochemical database (or network), more than half are positive ions, which thus play a key role in cold molecular cloud chemistry. Despite their number, before my thesis, many cations were not characterized by experiments or theoretical studies, as evidenced by the fact that, for many of them, the only information available in the KIDA database is their name. I think that this lack of information is mainly due to their precursor role in the gas-phase reaction network and the small number of detection: of  $\sim 270$  interstellar detected species, about 30 are cations (McGuire, 2022). The cations in the networks mainly behave as intermediates in the formation of more (“*important*”) detected species, and their paucity of detection can also be explained by the difficulty of deriving their spectroscopic properties. On the other hand, lots of attention is addressed to neutral species and in fact almost all of them were characterized theoretically by Woon and Herbst (2009).

The reactions reported in the astrochemical networks are based on laboratory studies, quantum mechanics (QM) calculations, general principles (*e.g.*

---

<sup>1</sup><https://kida.astrochem-tools.org/>

capture theory for estimate reaction coefficients), as well as reactions inferred from chemical intuition and similarity principle. However, sometimes no investigation of their exothermicity, or feasibility in the interstellar conditions, were performed. In fact, this last statement is supported by the fact that only  $\sim 20\%$  of reactions have been adequately studied via experiments or computational studies. As already mentioned, I think that control should be devoted to the backbone reactions of the network. Indeed, considering the non-linear behaviour of the reaction network, it could be the case that a reaction considered unimportant could make a drastic contribution to the outputs of the model (Wakelam et al., 2006), and consequently validating the importance of the control.

**Goals and novelty of the research** The main goal of this Part II of my thesis is to identify erroneous reactions in gas-phase reaction network performing an extended reactions exo-/endothermicity control and, in practice, to provide a new reaction network, which is cleaned from endothermic reactions not explicitly recognized as such. This extended control has never been applied to astrochemical reactions, also due to the lack of specific chemical information. In fact, not of secondary importance, the creation of a database of physicochemical data obtained via accurate QM chemical method is the first goal of this Part II. Thus it is crucial to achieve our goals and to pave the way for new approaches.

**Summary of the results** The first objective was to create a database of accurate physicochemical data for cations, comparable in terms of methodology to those available for neutral species (Woon and Herbst, 2009). To achieve this objective, I developed a computational methodology to link a structure to each species name<sup>2</sup> and, consequently, all the physicochemical information at CCSD(T)/aug-cc-pVTZ//M06-2X/cc-pVTZ level. For each cation, I computed the symmetry group, electronic state, electronic spin multiplicity, rotational constants, dipole moment, absolute electronic energy, zero-point energy. In total, I performed ab initio calculations of 262 cations. All these data were merged with the neutral species properties to create the database GreToBaPe<sup>3</sup>.

The second and major objective of this thesis's Part II is to produce a gas-phase reaction network "*cleaned*" from endothermic reactions not reported as such. The new reaction network is called GRETObAPE, built from the publicly available KIDA2014 (Wakelam et al., 2015) with the addition of

---

<sup>2</sup><https://github.com/TinacciL/FromGraphToXYZ>

<sup>3</sup>[https://github.com/TinacciL/GreToBaPe\\_Cleaning](https://github.com/TinacciL/GreToBaPe_Cleaning)

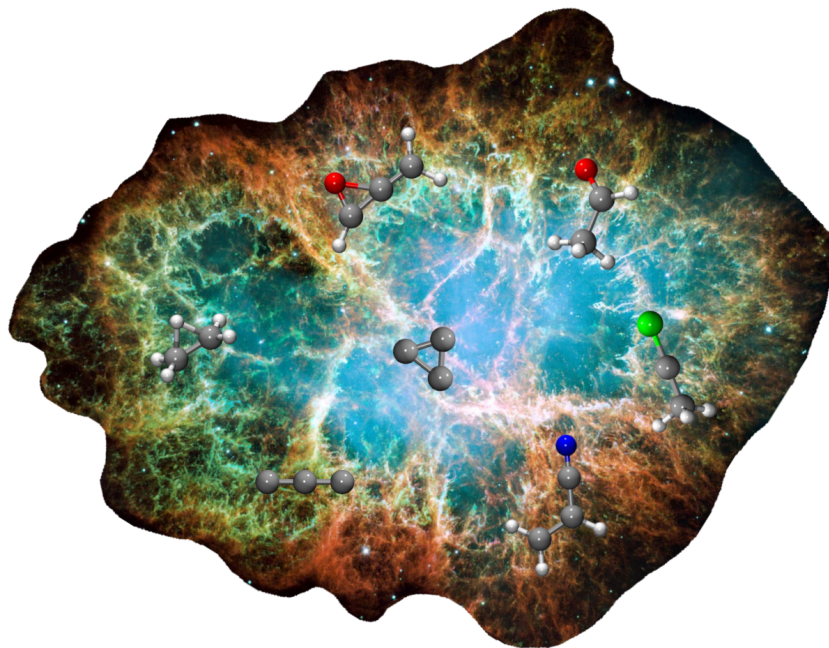
several new reactions and corrections from more recent studies. With the information coming from the above-mentioned database, I created a procedure based on graph theory to easily handle the cleaning of the reaction network and visualized/studied its properties<sup>3</sup>. I found that the  $\sim 5\%$  (342) of the original network reactions with thermodynamic data (6275) are endothermic reactions not recognised as such, and almost 2/3 of them are reported as barrierless. The silicon reaction network is the most affected by the network cleaning process. As shown by the number of reactions removed involving silicon-bearing species ( $\sim 15\%$  of total deleted reactions), the species removed (6 over the total 11) and the comparison of astrochemical simulations using the reaction network before and after cleaning, in molecular cloud and hot molecular outflow shock conditions. Furthermore, a review of the possible missing destruction reactions between neutral species and the most abundant and reactive cations (*e.g.*  $\text{H}^+$ ,  $\text{He}^+$ ,  $\text{H}_3^+$ ,  $\text{H}_3\text{O}^+$  and  $\text{HCO}^+$ ) was done.

A reduced version of the cleaned reaction network was also provided for those cases in which a large reaction network is not necessary while increasing the computational time, *e.g.* astrochemical models coupled with hydrodynamical simulations of cloud collapse or protoplanetary disk chemistry.

**Structure of this thesis's part** In order to achieve the above mentioned goals one article was published (Tinacci *et al.*, 2021) and another one submitted in October 2022. Each of these articles constitutes the two following Chapters of this thesis's part. The first Chapter (Chap. 4) focuses on the creation of the chemical database for cations, while the second Chapter (Chap. 5) deals with the identification of exothermic reactions, their removal from the network, the study of their impact on astrochemical models and, finally, the publication of a new “*clean*” reaction network.

# Structures and properties of known and postulated interstellar cations

The hereafter article has been published in The Astrophysical Journal Supplement Series, 256:35 (10pp), 2021 October.





# Structures and Properties of Known and Postulated Interstellar Cations

Lorenzo Tinacci<sup>1,2</sup> , Stefano Pantaleone<sup>1,3</sup> , Andrea Maranzana<sup>1</sup> , Nadia Balucani<sup>2,3,4</sup> , Cecilia Ceccarelli<sup>2</sup> , and Piero Ugliengo<sup>1</sup>

<sup>1</sup> Dipartimento di Chimica and Nanostructured Interfaces and Surfaces (NIS) Centre, Università degli Studi di Torino, via P. Giuria 7, I-10125 Torino, Italy

[Lorenzo.Tinacci@univ-grenoble-alpes.fr](mailto:Lorenzo.Tinacci@univ-grenoble-alpes.fr)

<sup>2</sup> Université Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France

<sup>3</sup> Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, I-06123 Perugia, Italy

<sup>4</sup> Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy

Received 2021 May 12; revised 2021 July 14; accepted 2021 July 18; published 2021 September 28

## Abstract

Positive ions play a fundamental role in interstellar chemistry, especially in cold environments where chemistry is believed to be mainly ion driven. However, in contrast with neutral species, most of the cations present in the astrochemical reaction networks are not fully characterized in the astrochemical literature. To fill this gap, we have carried out new accurate quantum chemical calculations to identify the structures and energies of 262 cations with up to 14 atoms that are postulated to have a role in interstellar chemistry. Optimized structures and rotational constants were obtained at the M06-2X/cc-pVTZ level, while electric dipoles and total electronic energies were computed with CCSD(T)/aug-cc-pVTZ//M06-2X/cc-pVTZ single-point energy calculations. The present work complements the study by Woon & Herbst, who characterized the structure and energies of 200 neutral species also involved in interstellar chemistry. Taken together, the two data sets can be used to estimate whether a reaction, postulated in present astrochemical reaction networks, is feasible from a thermochemistry point of view and, consequently, to improve the reliability of the present networks used to simulate the interstellar chemistry. We provide an actual example of the potential use of the cations plus neutral data sets. It shows that two reactions, involving Si-bearing ions and present in the widely used reaction networks KIDA and UMIST, cannot occur in the cold interstellar medium because they are endothermic.

*Unified Astronomy Thesaurus concepts:* [Astrochemistry \(75\)](#); [Interstellar molecules \(849\)](#)

*Supporting material:* data behind figure, machine-readable table

## 1. Introduction

Soon after the first detection in the late 1960s of polyatomic molecules in interstellar cold (10–20 K) molecular clouds (Cheung et al. 1968, 1969; Snyder et al. 1969), the dominant role of cations in the chemistry leading to them became clear (Watson 1973; Herbst & Klemperer 1973). The reason is relatively simple: molecular clouds are too cold for reactions that present activation barriers to take place, and the vast majority of neutral–neutral reactions possess activation barriers (insurmountable at 10 K). Therefore, in cold molecular clouds, chemistry is believed to be mainly driven by cations, whose root is the ionization of hydrogen (both in the atomic and molecular forms) by the cosmic rays that permeate the Milky Way.

The first confirmation of this theoretical prediction came with the detection of HCO<sup>+</sup> by Snyder et al. (1976).<sup>5</sup> To date, out of slightly more than 200 interstellar detected species, about 30 are cations,<sup>6</sup> the last ones discovered being HC<sub>3</sub>S<sup>+</sup> and CH<sub>3</sub>CO<sup>+</sup> (Cernicharo et al. 2021a, 2021b). Interestingly, all the 13 detected cations with more than three atoms are, so far, protonated forms of stable and abundant molecules. It is important to emphasize that the relatively low number of detected cations is not due to their real paucity, at least based on the astrochemical theoretical predictions, but on their low abundance and the difficulty of deriving their spectroscopic properties (e.g., McGuire et al. 2020).

As a matter of fact, of the about 500 species involved in the present astrochemical gas-phase reaction networks (e.g., KIDA<sup>7</sup> and UMIST), more than half are cations. In the same vein, of the 8000 or so reactions in the same reaction networks, the majority, about 5500, involve cations. Yet, despite the obvious importance of cations in the modeling of interstellar chemistry, no systematic study exists in the literature on the structure and energy of the cations involved in these networks. Indeed, it is worth noting that the abovementioned reaction networks list cations whose structure has seldom been characterized and often appear as chemical formulae guessed on the basis of the reactions that involve them. In contrast, a systematic theoretical study of many neutral species present in the astrochemical reaction networks was carried out more than a decade ago by Woon & Herbst (2009).

The goal of the present work is to provide accurate physicochemical data for cation species, comparable in terms of methodology with those available for neutral species, to ultimately improve the accuracy of the astrochemical models. An obvious example of the impact of having reliable data of all species present in the astrochemical networks is that this will allow for quick verification of the exo-/endothermicity of the reaction, and, if relevant, excluding it from the network without the need for the very time-consuming characterization of the transition states of the reaction.

To reach the goal of providing reliable data on the cations, accurate estimates of the electronic spin multiplicity,

<sup>5</sup> Actually, the theory followed the suggestion by Klemperer (1970) that an unidentified line observed in a few sources was to be attributed to HCO<sup>+</sup>.

<sup>6</sup> <https://cdms.astro.uni-koeln.de/classic/molecules>

<sup>7</sup> <http://kida.astrophy.u-bordeaux.fr/>

<sup>8</sup> <http://udfa.ajmarkwick.net/>



geometrical structure, and absolute electronic energy of each cation are needed (Herzberg 1966; Lattelais et al. 2009, 2010; Chabot et al. 2013). Here we present new computations of the physicochemical properties of the 262 cations present in the KIDA astrochemical gas-phase reaction network (Wakelam et al. 2012).

The article is organized as follows. In Section 2, we provide details on the adopted computational methodology. In Section 3, we report the results of the new computations. In Section 4, we provide two examples of the possible application of the two data sets (the neutral one from Woon & Herbst (2009) and a cationic one from the present work) to identify and consequently exclude endothermic reactions present in the KIDA and UMIST reaction networks. Section 5 concludes the article and includes the hyperlink to the online database from this work, which is publicly available.

## 2. Methodology

### 2.1. Initial Guessed Geometrical Structures

In the astrochemical reaction networks, cations fall into two general classes of whether or not they are produced by ionization of a mother neutral species. Therefore, we adopted two different approaches to recover the initial guessed structures of the 262 cations.

For the 128 cations belonging to the first class (i.e., from ionization of a mother species), we started from the structure of the neutral species calculated by Woon & Herbst (2009), removing one electron and then optimizing the structure after assigning the proper charge and spin multiplicity. The reason behind this choice is that although cations are not produced by electron abstraction processes in most cases, astrochemical networks usually postulate that cation structures have the same connectivity of their neutral counterparts.

For about two-thirds of the remaining 134 cations, we retrieved the starting structures from the KIDA database<sup>7</sup> and the NIST Computational Chemistry Comparison and Benchmark DataBase (CCCBDB).<sup>9</sup> Finally, when only the brute formulas were available (about 50 cations), we guessed the starting structure case by case, looking at the products of the reactions forming and destroying the cations. To automatize the initial geometric guess for the unknown chemical structure, we developed a graph-theory-based software tool coupled with the Universal Force Field (Rappé et al. 1992) implemented in RdKit (Landrum 2016). The script can be found in the online-only material and at the Astro-Chemistry Origin (ACO) Cations Scripts website.<sup>10</sup>

We emphasize that the procedures and choices described above stem from the fact that, very often, only simple connectivity or a cation name are available in the reaction networks and not the structure itself. In the few uncertain cases where *cis-trans* isomers are possible and no further information is available in the network to differentiate them, we assumed the most stable one (in the specific case, the *trans* isomer, based on the general rules of organic chemistry).

### 2.2. Computational Details

Once the 262 guessed geometrical structures were obtained, a sequence of geometric optimizations at the density functional theory (DFT) level were made, considering both electronic spin multiplicities in the ground and first excited states.

All calculations were carried out with the Gaussian16 program (Frisch et al. 2016). For the DFT calculations, we adopted the Minnesota method M06-2X (Zhao & Truhlar 2008) coupled with the triple- $\zeta$  Dunning's correlation consistent basis set (cc-pVTZ) (Kendall 1992; Woon & Dunning 1993) for geometry optimization. We kept the default values set up in Gaussian16 for: the DFT integration grid (i.e., 99,590 grid points), self-consistent field convergence (i.e.,  $\Delta E = 10^{-8}$  Ha on the rms density matrix and  $\Delta E = 10^{-6}$  Ha on the maximum density matrix value and total energy), and geometry optimization tolerances (i.e.,  $3 \times 10^{-4}$  Ha and  $1.2 \times 10^{-3}$  a<sub>0</sub>, on rms gradients and displacements, respectively). We carefully explored symmetry constraints to maximize the number of symmetry elements compatible with the most stable structure. Harmonic frequency calculations were performed for all considered cases to ensure that each structure was a minimum of the potential energy surface (PES). Dipole moments and absolute electronic energies were refined at the coupled-cluster level with full single and double excitations and a perturbative treatment of triple excitations (CCSD(T) and ROCCSD(T); Knowles et al. 1993; Watts et al. 1993) in conjunction with an augmented triple- $\zeta$  correlation consistent basis set (aug-cc-pVTZ; Kendall 1992).

Since unrestricted electronic solutions are affected by spin contamination (i.e., the artificial mixing of different spin states) and this contribution is not negligible nor automatically corrected in Gaussian16 via the spin annihilation procedure, we adopted the Restricted Open (RO) formalism (except for singlet open, since the RO formalism is not applicable), whose wave function is the eigenfunction of the  $\hat{S}^2$  operator, for all open shell configurations. Spin contamination causes problems in recovering dynamic correlation mainly in post-HF methods that are based on many-body perturbation theory (MP2, CCSD), because the perturbation through high-spin states is too large to be correctly accounted for by these methods (Watts et al. 1993).

Moreover, a stability analysis on the converged wave function was applied to the singlet states computed via the M06-2X/cc-pVTZ level (Bauernschmitt & Ahlrichs 1996) using the specific keywords (*opt = stable*) provided in Gaussian16 (see Figure 1;  $\Psi_{1c}$  stability block).

The scheme summarizing the adopted strategy is shown in Figure 1, where the various steps needed to reach the final minimum structure are shown.

Rendering of molecule images have been obtained via the VMD software (Humphrey et al. 1996), and the graphics elaboration and plots via the TikZ and PGFPlots LATEX packages.

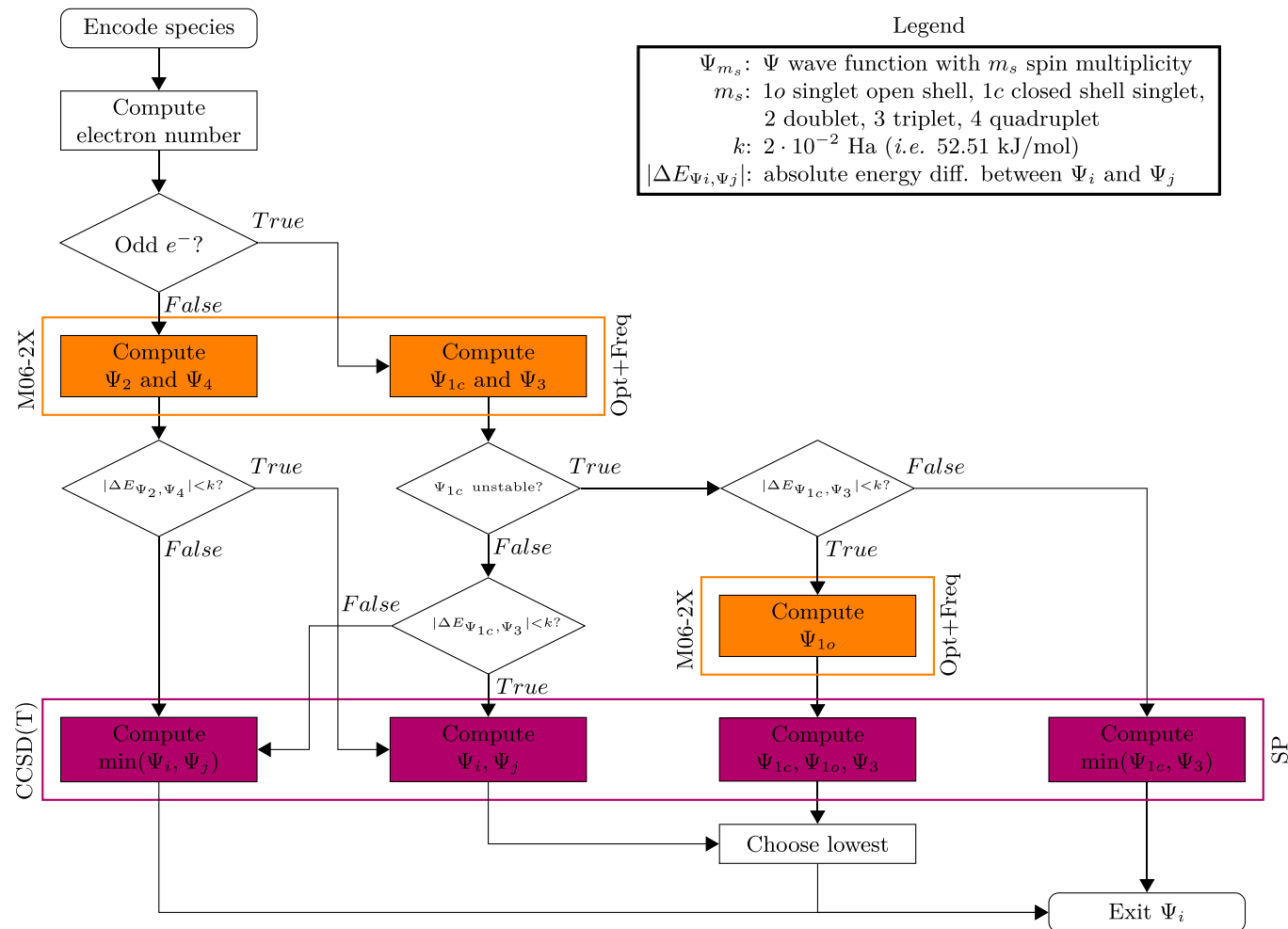
### 2.3. Benchmark Method

In order to test the accuracy of the above-described methodology, a benchmark on both structure and wave function optimizations using different methods was carried out.

*Geometry optimization*—We checked the M06-2X/cc-pVTZ level of theory by comparing our results obtained for a subset of molecules with those computed at the CCSD/aug-cc-pVTZ level, as shown in Table 1. Both the geometry rms deviation (RMSD) and the energy difference are small enough (less than  $\sim 0.045$  Å

<sup>9</sup> <http://cccbdb.nist.gov/>

<sup>10</sup> [aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/software-packages/cations-structures-scripts](http://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/software-packages/cations-structures-scripts)



**Figure 1.** Adopted procedure to define the ground electronic state for a species and to achieve the related optimized geometry and cation properties. To this end, we used two levels of theory: M06-2X/cc-pVTZ and CCSD(T)/aug-cc-pVTZ. In the M06-2X block, a geometric optimization (Opt) and vibrational frequencies calculation (Freq) are performed to ensure that a minimum PES is achieved. In the CCSD(T) block, the final energy is then refined as a single-point evaluation at the CCSD(T) level together with the corresponding properties. The  $\Psi_{1c}$  stability is tested with the Gaussian16 wave function stability tool to find restricted  $\rightarrow$  unrestricted wave function instability.

and  $\sim 1.7$  kJ mol $^{-1}$ , respectively) to validate the present methodology. Moreover, the average  $\Delta E$  is 0.522 kJ mol $^{-1}$ , which is lower than the commonly accepted quantum mechanics calculation accuracy of  $\sim 4$  kJ mol $^{-1}$  (i.e., 1 kcal mol $^{-1}$ ). The extended internal coordinates geometry optimization comparison is available in the [Appendix](#).

**Wave function optimization**—The procedure shown in Figure 1 was first tested on two very common molecules: ethylene (C<sub>2</sub>H<sub>6</sub>) and methylene (CH<sub>2</sub>). These molecules present different ground states, i.e., singlet closed shell ( $^1A_g$ ) and triplet ( $^3B_1$ ) for ethylene and methylene, respectively. Their corresponding excited states are triplet ( $^3A_1$ ) for ethylene and singlet closed shell ( $^1A_1$ ) for methylene. Our calculated transition energies are in good agreement with the experimental data: 282 versus 272 kJ mol $^{-1}$  (Douglas et al. 1955) and 40 versus 38 kJ mol $^{-1}$  (Shavitt 1985) for ethylene and methylene, respectively. For the singlet closed shell methylene ( $^1A_1$ ) the stability analysis was also performed revealing a preference for the unrestricted solution with respect to the restricted one, as expected.

**Dipole moment evaluation**—Woon & Herbst (2009) showed that the CCSD(T)/aug-cc-pVTZ electronic dipole for neutral species is in good agreement with the experimental data. We

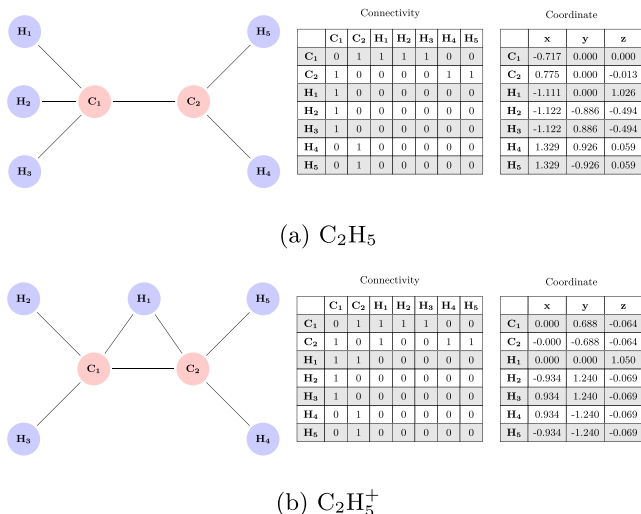
**Table 1**  
Rms Displacement (RMSD) of Atomic Positions and the Absolute Energy Difference

Species	State	$\Delta E$ [kJ mol $^{-1}$ ]	RMSD [Å]
C <sub>2</sub> <sup>+</sup>	$^4\Sigma_g$	0.247	0.003
NH <sub>4</sub> <sup>+</sup>	$^1A_1$	-0.022	0.000
H <sub>2</sub> CO <sup>+</sup>	$^2B_2$	0.247	0.007
PNH <sub>2</sub> <sup>+</sup>	$^2B_2$	0.512	0.005
l-C <sub>3</sub> H <sub>2</sub> <sup>+</sup>	$^2A'$	0.055	0.045
H <sub>3</sub> CS <sup>+</sup>	$^3A_1$	0.509	0.012
c-C <sub>3</sub> H <sub>3</sub> <sup>+</sup>	$^1A_1'$	0.714	0.004
C <sub>4</sub> H <sub>3</sub> <sup>+</sup>	$^1A_1$	1.681	0.008
CH <sub>3</sub> CHOH <sup>+</sup>	$^1A'$	0.432	0.006
H <sub>2</sub> C <sub>3</sub> O <sup>+</sup>	$^2B_2$	0.845	0.005

**Note.** Calculated at CCSD(T)/aug-cc-pVTZ//M06-2X/cc-pVTZ compared with calculations at CCSD(T)/aug-cc-pVTZ//CCSD/aug-cc-pVTZ.

expect a similar or better agreement for cations, in virtue of the more contracted nature of the electron density compared to the more diffuse one in neutral species.





**Figure 2.** Graphs, connectivity matrices, and coordinates (in Å) of the neutral ethyl radical  $C_2H_5$  and its cation  $C_2H_5^+$ .

#### 2.4. Neutral versus Cation Structure Connectivity

Since many cation structures are derived from the neutral counterpart, we have calculated the ionization energy and followed the connectivity change (if any) after the geometry relaxation on the proper PES. In order to check if a cation retains the same connectivity of the neutral counterpart after geometric relaxation, the graph theory approach was used. Finally, the adiabatic ionization energy for the species that have a neutral counterpart will be presented in the dedicated subsection 3.2 with a comparison with available experimental data.

First, all the coordinate files are converted into chemical graphs (Trinajstić 2018) using the covalent radii and distance functions implemented in the Atomic Simulation Environment (ASE) python package (Larsen et al. 2017). A chemical graph is a nondirected graph where atoms and bonds in molecules correspond to nodes and edges, respectively. The different chemical elements present in the periodic table are represented in graphs as colors assigned to the vertices: the graph is, therefore, defined as a multicolored graph. Second, the graphs related to the neutral and ionized molecules are tested for isomorphism. Two graphs which contain the same number of colored vertices connected in the same way are considered isomorphic. The NetworkX python package (Hagberg et al. 2008) was used in order to deal with graph objects. Examples of chemical graphs are shown in Figure 2.

### 3. Results

The structures of the cations computed following the procedures and methodology described in the previous section are minima of PES when starting from the guessed ones, i.e., we did not explore the full PES in search of the global minimum. Indeed, our goal is to consider cations whose connectivity derives from the astrochemical reaction networks with structures based on the reactions giving rise to their formation. When more isomeric forms exist, we adopted the most stable one when specific information in the reaction network was missing. Finally, we cross-checked our computed cation structures with the literature ones, in the relatively few cases where they are available, and generally found a very good agreement.

**Table 2**

Predicted Equilibrium Structures and Properties of Diatomic Species

Species	State	Symmetry	$\mu$	$B_e$	$r_e$	$r_e$ (exp)
$H_2^+$	$2\Sigma_g$	$D_{\infty h}$	0.000	827.202	1.101	1.057 <sup>a</sup>
$HeH^+$	$1\Sigma$	$C_{\infty v}$	1.343	1010.927	0.788	
$CH^+$	$1\Sigma$	$C_{\infty v}$	1.565	428.701	1.126	1.131 <sup>b</sup>
$NH^+$	$2\Pi$	$C_{\infty v}$	1.834	466.167	1.074	
$OH^+$	$3\Sigma$	$C_{\infty v}$	2.127	501.498	1.031	1.029 <sup>a</sup>
$HF^+$	$2\Pi$	$C_{\infty v}$	2.478	518.136	1.010	
$C_2^+$	$4\Sigma_g$	$D_{\infty h}$	0.000	43.077	1.398	
$CN^+$	$1\Sigma$	$C_{\infty v}$	2.846	57.192	1.169	1.290 <sup>a</sup>
$CO^+$	$2\Sigma$	$C_{\infty v}$	3.286	60.629	1.103	
$N_2^+$	$2\Sigma_g$	$D_{\infty h}$	0.000	59.719	1.099	1.113 <sup>a</sup>
$SiH^+$	$1\Sigma$	$C_{\infty v}$	0.330	228.988	1.506	1.499 <sup>a</sup>
$NO^+$	$1\Sigma$	$C_{\infty v}$	0.603	61.361	1.050	1.062 <sup>a</sup>
$CF^+$	$1\Sigma$	$C_{\infty v}$	1.305	51.829	1.151	1.263 <sup>a</sup>
$PH^+$	$2\Pi$	$C_{\infty v}$	0.741	255.260	1.424	
$O_2^+$	$2\Pi_g$	$D_{\infty h}$	0.000	52.927	1.093	1.116 <sup>a</sup>
$HS^+$	$3\Sigma$	$C_{\infty v}$	1.170	278.193	1.364	
$HCl^+$	$2\Pi$	$C_{\infty v}$	1.634	295.614	1.321	
$SiC^+$	$4\Sigma$	$C_{\infty v}$	1.026	18.276	1.815	
$SiN^+$	$3\Pi$	$C_{\infty v}$	2.860	17.154	1.777	
$CP^+$	$3\Sigma$	$C_{\infty v}$	0.120	22.542	1.610	
$CS^+$	$2\Sigma$	$C_{\infty v}$	0.945	26.449	1.480	
$PN^+$	$2\Sigma$	$C_{\infty v}$	2.040	24.210	1.471	
$NS^+$	$1\Sigma$	$C_{\infty v}$	2.209	25.735	1.420	
$PO^+$	$1\Sigma$	$C_{\infty v}$	4.030	23.802	1.419	
$SiF^+$	$1\Sigma$	$C_{\infty v}$	3.599	18.802	1.541	
$CCl^+$	$1\Sigma$	$C_{\infty v}$	0.165	24.050	1.534	
$SiO^+$	$2\Sigma$	$C_{\infty v}$	3.253	21.372	1.524	
$SO^+$	$2\Pi$	$C_{\infty v}$	2.785	23.644	1.416	
$ClO^+$	$3\Sigma$	$C_{\infty v}$	0.569	21.249	1.472	
$SiS^+$	$2\Sigma$	$C_{\infty v}$	4.576	9.240	1.915	
$S_2^+$	$2\Pi_g$	$D_{\infty h}$	0.000	9.641	1.811	

**Notes.**  $\mu$  is the electric dipole moment in Debye units.  $B_e$  is the rotational constant expressed in gigahertz referring to the equilibrium structure.  $r_e$  is the calculated internuclear equilibrium distance expressed in angstroms, meanwhile  $r_e(\text{exp})$  are the available experimental data.

<sup>a</sup> From Chase (1996).

<sup>b</sup> From Huber (2013).

#### 3.1. Cations Properties and Geometries

Tables 2–6 list the calculated properties of the 262 cations studied in this work. In the tables, we grouped the cations into five categories: (1) diatomic species (Table 2); (2) linear species with 3–12 atoms (Table 3); (3)  $C_{2v}$  symmetry species (Table 4); (4) planar species (excluding those belonging to point 3) (Table 5); (5) nonplanar species (Table 6). Within each table, cations are ordered by increasing number of atoms and, within each subset, by increasing molecular mass. A sample of the derived structures is shown in Figure 3.

Online-only material with extended information on all 262 cations is provided as data behind Figure 3. The center of the coordinate frame with respect to the dipole moment components is referred to is the center of nuclear charge. We also make the data publicly available on the website of the ACO project site.<sup>11</sup> The web-based ACO cation structure database<sup>12</sup> is based on the molecule hyperactive JSmol<sup>13</sup> plugin. The ACO

<sup>11</sup> <https://aco-itn.oapd.inaf.it/home>

<sup>12</sup> [aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/cations-database](https://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/cations-database)

<sup>13</sup> JSmol is an open-source Java viewer for chemical structures in 3D (<http://www.jmol.org/>).

**Table 3**  
Predicted Equilibrium Properties of Linear Polyatomic Species

Species	State	Symmetry	$\mu$	$B_e$	Species	State	Symmetry	$\mu$	$B_e$
C <sub>2</sub> H <sup>+</sup>	<sup>3</sup> Π	C <sub>∞v</sub>	0.755	41.665	SiC <sub>4</sub> <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	7.013	1.516
HCN <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	3.586	41.202	C <sub>4</sub> P <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.101	1.530
HNC <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	0.148	47.950	C <sub>4</sub> S <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	2.294	1.531
HCO <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	4.172	45.359	C <sub>4</sub> H <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>g</sub>	D <sub>∞h</sub>	0.000	4.431
HOC <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.380	44.893	HC <sub>2</sub> NCH <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	3.484	4.689
N <sub>2</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	3.154	47.597	HC <sub>3</sub> NH <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	1.246	4.350
CNC <sup>+</sup>	<sup>1</sup> Σ <sub>g</sub>	D <sub>∞h</sub>	0.000	13.661	C <sub>5</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.677	2.420
C <sub>2</sub> N <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.660	12.004	HC <sub>4</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	6.436	2.327
CHSi <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	0.246	15.767	HC <sub>4</sub> O <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	4.504	2.246
NCO <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	1.117	11.196	C <sub>6</sub> <sup>+</sup>	<sup>2</sup> Π <sub>u</sub>	D <sub>∞h</sub>	0.000	1.445
HNSi <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	3.821	16.886	C <sub>5</sub> N <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	3.890	1.391
HCP <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	0.985	18.914	SiC <sub>4</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.487	1.440
CO <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>g</sub>	D <sub>∞h</sub>	0.000	11.600	PC <sub>4</sub> H <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	0.274	1.444
HCS <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	1.964	21.653	HC <sub>4</sub> S <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	2.736	1.463
HSiO <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	6.769	19.171	C <sub>6</sub> H <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	3.410	1.393
HPN <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	0.267	20.749	HC <sub>5</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	7.395	1.345
NO <sub>2</sub> <sup>+</sup>	<sup>1</sup> Σ <sub>g</sub>	D <sub>∞h</sub>	0.000	12.903	HC <sub>5</sub> O <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.586	1.308
SiC <sub>2</sub> <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	1.472	6.125	C <sub>7</sub> <sup>+</sup>	<sup>2</sup> Σ <sub>u</sub>	D <sub>∞h</sub>	0.000	0.912
SiNC <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	4.015	6.653	C <sub>6</sub> H <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>u</sub>	D <sub>∞h</sub>	0.000	1.343
C <sub>2</sub> S <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	1.112	6.489	H <sub>2</sub> C <sub>5</sub> N <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	3.489	1.300
OCS <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	1.653	5.849	C <sub>7</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.299	0.885
HSiS <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	5.168	8.661	HC <sub>6</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	7.813	0.850
C <sub>2</sub> H <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>u</sub>	D <sub>∞h</sub>	0.000	33.595	C <sub>8</sub> <sup>+</sup>	<sup>2</sup> Σ <sub>g</sub>	D <sub>∞h</sub>	0.000	0.610
HCNH <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	0.566	37.543	C <sub>7</sub> N <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	4.860	0.587
C <sub>3</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	2.552	11.262	C <sub>8</sub> H <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	3.461	0.592
HSiNH <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	3.460	17.494	HC <sub>7</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	8.764	0.570
C <sub>4</sub> <sup>+</sup>	<sup>2</sup> Π <sub>g</sub>	D <sub>∞h</sub>	0.000	4.836	HC <sub>7</sub> O <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	1.022	0.555
C <sub>3</sub> O <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	2.809	4.879	C <sub>9</sub> <sup>+</sup>	<sup>2</sup> Σ <sub>u</sub>	C <sub>∞v</sub>	0.011	0.429
C <sub>2</sub> N <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>g</sub>	C <sub>∞v</sub>	0.000	4.716	C <sub>8</sub> N <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	7.020	0.414
SiC <sub>2</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	0.623	5.582	C <sub>8</sub> H <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>g</sub>	D <sub>∞h</sub>	0.000	0.574
PC <sub>2</sub> H <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	0.219	5.803	C <sub>7</sub> H <sub>2</sub> N <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	6.191	0.556
HC <sub>2</sub> S <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	2.502	6.049	C <sub>9</sub> H <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	1.405	0.418
l-SiC <sub>3</sub> <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	1.129	2.651	HC <sub>8</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	8.870	0.403
C <sub>3</sub> S <sup>+</sup>	<sup>2</sup> Σ	C <sub>∞v</sub>	0.962	2.858	C <sub>10</sub> <sup>+</sup>	<sup>2</sup> Π <sub>u</sub>	D <sub>∞h</sub>	0.000	0.312
C <sub>4</sub> H <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	3.140	4.640	C <sub>9</sub> N <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	5.942	0.301
HC <sub>3</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	6.378	4.568	C <sub>10</sub> H <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	3.258	0.304
HC <sub>3</sub> O <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	3.635	4.484	C <sub>9</sub> HN <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	10.189	0.294
C <sub>5</sub> <sup>+</sup>	<sup>2</sup> Σ <sub>u</sub>	D <sub>∞h</sub>	0.000	2.596	HC <sub>9</sub> O <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	1.055	0.286
C <sub>4</sub> N <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	3.900	2.438	C <sub>11</sub> <sup>+</sup>	<sup>2</sup> Σ <sub>u</sub>	D <sub>∞h</sub>	0.000	0.235
SiC <sub>3</sub> H <sup>+</sup>	<sup>3</sup> Σ	C <sub>∞v</sub>	0.661	2.556	C <sub>10</sub> N <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	8.995	0.227
PC <sub>3</sub> H <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	1.002	2.662	C <sub>10</sub> H <sub>2</sub> <sup>+</sup>	<sup>2</sup> Π <sub>u</sub>	D <sub>∞h</sub>	0.000	0.296
HC <sub>3</sub> S <sup>+</sup>	<sup>1</sup> Σ	C <sub>∞v</sub>	1.892	2.753	HC <sub>10</sub> N <sup>+</sup>	<sup>2</sup> Π	C <sub>∞v</sub>	9.935	0.222

**Note.**  $\mu$  is the electric dipole moment in Debye units.  $B_e$  is the rotational constant expressed in gigahertz referring to the equilibrium structure.

cation structure website will be periodically updated when new species will be added to the database. On the same site, we also provide the python script tool<sup>14</sup> to convert molecules into a chemical graph and to exert control if two species share the same connectivity using the isomorphic function. The script can also be found in the online-only material.

Finally, while the rotational constants are reported in Tables 2 and 3 and in the web-based ACO cations database (for all the other nonlinear species), we warn the reader that the level of theory is not sufficiently accurate to allow their use for assigning bands in experimental spectra.

<sup>14</sup> [aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/software-packages/cations-structures-scripts](http://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/software-packages/cations-structures-scripts)

### 3.2. Adiabatic Ionization Energy

Table 7 shows the adiabatic ionization energies (IE) for neutral species computed at CCSD(T)/aug-cc-pVTZ by Woon & Herbst (2009) along with those of the cation counterparts. All computed IE are without zero-point energy (ZPE) correction. Note that the cation counterparts are part of the group whose structures are derived from the ionization of neutral species, described in Section 2.1. As some of the data on larger molecules provided by Woon & Herbst (2009) are computed with a lower-quality basis set, we recomputed the energy for those species with the aug-cc-pVTZ.

The rms error between the computed ionization energies (adiabatic IE) and the experimental available data (vertical IE)

**Table 4**  
Predicted Equilibrium Properties of Planar Polyatomic Species with  $C_{2v}$  Symmetry

Species	State	$\mu$	Species	State	$\mu$	Species	State	$\mu$
$\text{CH}_2^+$	$^2A_1$	0.488	$\text{H}_2\text{SiO}^+$	$^2B_2$	3.737	$\text{C}_4\text{H}_3^+$	$^1A_1$	0.297
$\text{NH}_2^+$	$^3B_1$	0.559	$\text{PNH}_2^+$	$^2B_2$	1.140	$\text{PC}_4\text{H}_2^+$	$^1A_1$	3.268
$\text{H}_2\text{O}^+$	$^2B_1$	2.201	$\text{H}_2\text{PO}^+$	$^1A_1$	5.344	$\text{C}_5\text{H}_2^+$	$^2B_2$	4.696
$\text{H}_2\text{F}^+$	$^1A_1$	2.342	$\text{H}_2\text{CCl}^+$	$^1A_1$	3.330	$\text{H}_2\text{C}_4\text{N}^+$	$^1A_1$	8.037
$\text{NaH}_2^+$	$^1A_1$	1.294	$\text{HSO}_2^+$	$^1A_1$	3.802	$\text{C}_5\text{H}_5^+$	$^1A_1$	0.815
$\text{SiH}_2^+$	$^2A_1$	0.116	$\text{H}_2\text{S}_2^+$	$^2A_2$	2.154	$\text{C}_5\text{H}_3^+$	$^1A_1$	1.907
$\text{PH}_2^+$	$^1A_1$	0.915	$\text{C}_2\text{H}_3^+$	$^1A_1$	0.785	$\text{CH}_3\text{OCH}_3^+$	$^2B_1$	0.791
$\text{H}_2\text{S}^+$	$^2B_1$	1.466	$c\text{-C}_3\text{H}_2^+$	$^2A_1$	1.128	$\text{C}_6\text{H}_3^+$	$^1A_1$	1.721
$\text{C}_3^+$	$^2B_2$	0.731	$\text{CH}_2\text{CN}^+$	$^1A_1$	5.472	$\text{C}_7\text{H}_2^+$	$^2B_2$	6.388
$\text{H}_2\text{Cl}^+$	$^1A_1$	2.010	$\text{H}_2\text{CCN}^+$	$^1A_1$	5.471	$\text{H}_2\text{C}_6\text{N}^+$	$^1A_1$	10.568
$\text{SO}_2^+$	$^2A_1$	1.871	$\text{H}_2\text{CCO}^+$	$^2B_1$	3.539	$\text{C}_7\text{H}_3^+$	$^1A_1$	3.426
$\text{CH}_4^+$	$^2B_2$	1.478	$\text{SiC}_2\text{H}_2^+$	$^2B_2$	0.038	$\text{C}_6\text{H}_5^+$	$^1A_1$	1.330
$\text{H}_2\text{NC}^+$	$^1A_1$	2.335	$\text{PC}_2\text{H}_2^+$	$^1A_1$	1.973	$\text{C}_8\text{H}_3^+$	$^1A_1$	3.505
$\text{H}_2\text{CO}^+$	$^2B_2$	3.025	$\text{CH}_2\text{NH}_2^+$	$^1A_1$	0.112	$\text{C}_6\text{H}_2^+$	$^2B_2$	8.177
$\text{H}_2\text{NO}^+$	$^1A_1$	3.639	$l\text{-C}_3\text{H}_3^+$	$^1A_1$	0.727	$\text{H}_2\text{C}_8\text{N}^+$	$^1A_1$	13.241
$\text{NaH}_2\text{O}^+$	$^1A_1$	2.463	$\text{NH}_2\text{CNH}^+$	$^1A_1$	0.180	$\text{C}_9\text{H}_3^+$	$^1A_1$	5.329
$\text{CH}_2\text{Si}^+$	$^2A_1$	2.280	$\text{H}_2\text{C}_3\text{O}^+$	$^2B_2$	4.275	$\text{C}_9\text{H}_2\text{N}^+$	$^1A_1$	13.766
$\text{PCH}_2^+$	$^1A_1$	0.588	$\text{SiC}_3\text{H}_2^+$	$^2B_1$	0.681	$\text{C}_{10}\text{H}_3^+$	$^1A_1$	5.697
$\text{H}_2\text{CS}^+$	$^2B_2$	2.013	$\text{C}_2\text{H}_5^+$	$^1A_1$	0.612	$\text{H}_2\text{C}_{10}\text{N}^+$	$^1A_1$	16.153

**Note.**  $\mu$  is the electric dipole moment in Debye units.

**Table 5**  
Predicted Equilibrium Properties of Planar Polyatomic Species without  $C_{2v}$  Symmetry

Species	State	Symmetry	$\mu$	Species	State	Symmetry	$\mu$
$\text{H}_3^+$	$^1A_1'$	$D_{3h}$	0.000	$\text{HSiO}_2^+$	$^1A'$	$C_s$	5.494
$\text{HNO}^+$	$^2A'$	$C_s$	2.908	$\text{H}_2\text{COH}^+$	$^1A'$	$C_s$	2.352
$\text{HO}_2^+$	$^3A$	$C_s$	2.170	$l\text{-C}_3\text{H}_2^+$	$^2A'$	$C_s$	3.107
$\text{HNS}^+$	$^2A'$	$C_s$	1.439	$\text{HCOOH}^+$	$^2A'$	$C_s$	0.432
$\text{C}_2\text{O}^+$	$^2A$	$C_s$	1.733	$\text{H}_3\text{SiO}^+$	$^1A'$	$C_s$	2.278
$\text{HPO}^+$	$^2A'$	$C_s$	2.962	$\text{PNH}_3^+$	$^1A'$	$C_s$	3.093
$\text{HSO}^+$	$^1A'$	$C_s$	3.336	$c\text{-C}_3\text{H}_3^+$	$^1A_1'$	$D_{3h}$	0.000
$\text{CCP}^+$	$^1A'$	$C_s$	2.381	$\text{SiC}_2\text{H}_3^+$	$^1A'$	$C_s$	0.224
$\text{HS}_2^+$	$^1A'$	$C_s$	1.496	$\text{NH}_2\text{CH}_2\text{O}^+$	$^1A'$	$C_s$	2.190
$\text{CH}_3^+$	$^1A_1'$	$D_{3h}$	0.000	$\text{C}_3\text{H}_3\text{N}^+$	$^2A''$	$C_s$	6.293
$\text{NH}_3^+$	$^2A_2''$	$D_{3h}$	0.000	$c\text{-C}_3\text{H}_2\text{OH}^+$	$^1A'$	$C_s$	2.492
$\text{SiH}_3^+$	$^1A_1'$	$D_{3h}$	0.000	$\text{HCCCHOH}^+$	$^1A'$	$C_s$	1.234
$\text{HC}_2\text{N}^+$	$^2A'$	$C_s$	5.002	$\text{C}_2\text{H}_3\text{CO}^+$	$^1A'$	$C_s$	2.743
$\text{C}_2\text{HO}^+$	$^1A'$	$C_s$	3.384	$\text{C}_6\text{N}^+$	$^1A'$	$C_s$	5.324
$\text{HNCO}^+$	$^2A''$	$C_s$	3.446	$\text{C}_4\text{H}_4^+$	$^2A''$	$C_s$	1.078
$\text{HOCO}^+$	$^1A'$	$C_s$	3.446	$\text{C}_3\text{H}_3\text{NH}^+$	$^1A'$	$C_s$	1.707
$\text{HN}_2\text{O}^+$	$^1A'$	$C_s$	3.706	$\text{COOCH}_4^+$	$^2A'$	$C_s$	3.196
$\text{C}_3\text{N}^+$	$^3A''$	$C_s$	2.921	$\text{C}_5\text{H}_3^+$	$^1A'$	$C_s$	1.219
$\text{SiNCH}^+$	$^2A'$	$C_s$	2.156	$\text{C}_6\text{H}_4^+$	$^2A_u$	$C_{2h}$	0.000
$\text{HOCS}^+$	$^1A'$	$C_s$	2.274				

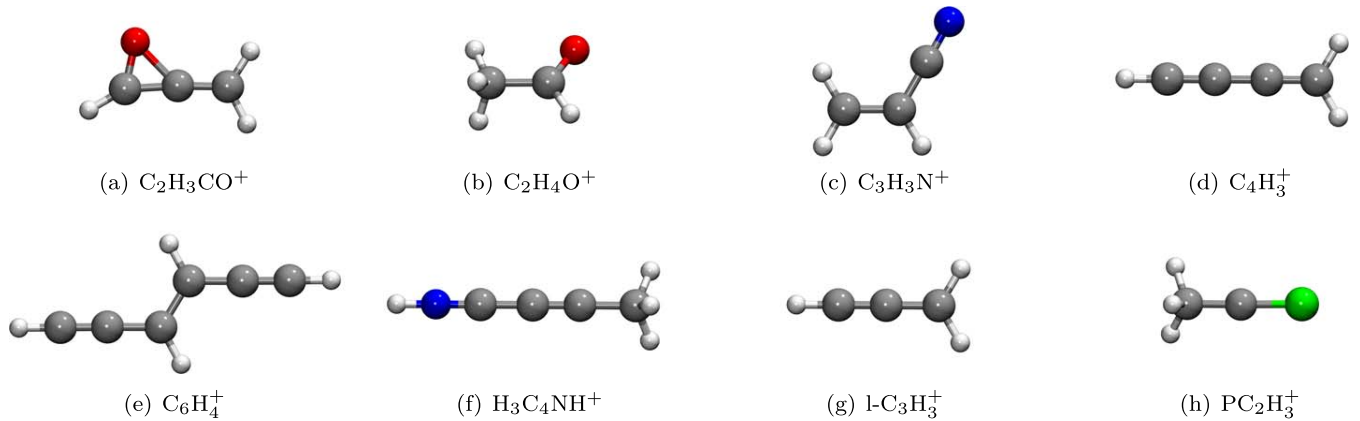
**Note.**  $\mu$  is the electric dipole moment in Debye units.

is 0.308 eV. The difference between these two quantities is that in vertical IE (experimental) the geometry of the ionized molecule is not allowed to relax in its corresponding PES, while in adiabatic IE (computed) the energy of the ion is taken from the minimum of its PES. Therefore, adiabatic IE are expected to be always lower with respect to vertical IE, if calculated with the same methodology.

*Connectivity change between neutral and cationic partners*—Using the tools described in the Section 2.4, the atomic connectivity check was performed on the cations reported in

Table 7 and compared to the corresponding neutral species. The only cations that change their connectivity after the ionization and the subsequent geometric optimization are:  $\text{C}_3^+$ ,  $\text{C}_2\text{H}_5^+$ , and  $\text{C}_4\text{H}_5^+$ , as shown in Figure 4. For the  $\text{C}_3^+$  this rearrangement explains the large difference ( $\sim 1.4$  eV) between our computed IE and the experimental vertical IE value.

This result corroborates our adopted methodology (Section 2), because it also demonstrates that the assumption in the astrochemical networks that many cations share the same connectivity as their neutral counterparts is respected.



**Figure 3.** M06-2X/cc-pVTZ optimized structures of a sample of the cations studied in the present work. The structure of all the 262 cations is available online as data behind the figure and at the ACO Cations Database website ([aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/cations-database](http://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/cations-database)).

(The data used to create this figure are available.)

**Table 6**  
Predicted Equilibrium Properties of Nonplanar Polyatomic Species

Species	State	Symmetry	$\mu$	Species	State	Symmetry	$\mu$
$H_3O^+$	$^1A_1$	$C_{3v}$	1.414	$CH_3NH_3^+$	$^1A_1$	$C_{3v}$	2.166
$PH_3^+$	$^2A_1$	$C_{3v}$	0.359	$CH_3CHOH^+$	$^1A'$	$C_s$	2.569
$H_3S^+$	$^1A_1$	$C_{3v}$	1.554	$H_3C_4N^+$	$^2A'$	$C_s$	5.327
$NH_4^+$	$^1A_1$	$T_d$	0.000	$C_2H_5OH^+$	$^2A$	$C_1$	2.083
$SiH_4^+$	$^2A'$	$C_s$	1.222	$C_4H_5^+$	$^1A'$	$C_s$	0.860
$SiCH_3^+$	$^1A_1$	$C_{3v}$	0.750	$H_5C_2O_2^+$	$^1A'$	$C_s$	0.972
$H_3CS^+$	$^3A_1$	$C_{3v}$	0.559	$C_5H_4^+$	$^2B_3$	$D_2$	0.000
$PCH_3^+$	$^2A''$	$C_s$	0.249	$H_3C_4NH^+$	$^1A_1$	$C_{3v}$	2.635
$H_3S_2^+$	$^1A$	$C_1$	2.292	$HCOCH_2OH_2^+$	$^1A'$	$C_s$	2.474
$CH_5^+$	$^1A'$	$C_s$	1.630	$C_5H_3N^+$	$^2A''$	$C_s$	6.270
$C_2H_4^+$	$^2B_3$	$D_2$	0.000	$C_2H_5OH_2^+$	$^1A$	$C_1$	3.312
$CH_3OH^+$	$^2A''$	$C_s$	1.394	$CH_3OCH_4^+$	$^1A'$	$C_s$	1.177
$CH_3CO^+$	$^1A_1$	$C_{3v}$	2.977	$C_2H_6CO^+$	$^2B$	$C_2$	1.567
$SiH_5^+$	$^1A'$	$C_s$	1.284	$C_5H_4N^+$	$^1A$	$C_1$	5.401
$SiCH_4^+$	$^2A'$	$C_s$	1.224	$C_4H_7^+$	$^1A'$	$C_s$	1.480
$PCH_4^+$	$^3A''$	$C_s$	0.858	$C_3H_6OH^+$	$^1A$	$C_1$	1.636
$PC_2H_3^+$	$^2A'$	$C_s$	0.375	$C_7H_4^+$	$^2B_2$	$D_2$	0.000
$CH_3O_2^+$	$^1A$	$C_1$	4.737	$H_3C_6NH^+$	$^1A_1$	$C_{3v}$	4.646
$CH_3CN^+$	$^2A'$	$C_s$	2.604	$H_3C_7N^+$	$^2A$	$C_1$	5.120
$CH_3NH_2^+$	$^2A'$	$C_s$	1.974	$C_7H_5^+$	$^1A'$	$C_s$	2.172
$CH_3OH_2^+$	$^1A'$	$C_s$	1.835	$C_8H_4^+$	$^2B_2$	$D_2$	0.000
$C_3H_4^+$	$^2B_2$	$D_2$	0.000	$C_6H_7^+$	$^1A'$	$C_s$	0.780
$CH_3CNH^+$	$^1A_1$	$C_{3v}$	1.036	$C_8H_5^+$	$^1A'$	$C_s$	3.950
$C_2H_4O^+$	$^2A'$	$C_s$	2.056	$C_9H_4^+$	$^2A$	$C_1$	0.391
$PC_2H_4^+$	$^1A'$	$C_s$	0.573	$C_8H_4N^+$	$^1A_1$	$C_{3v}$	7.166
$C_2H_6^+$	$^2A_g$	$C_{2h}$	0.000	$C_9H_5^+$	$^1A'$	$C_s$	3.960

**Note.**  $\mu$  is the electric dipole moment in Debye units.

#### 4. Discussion

As explained in the Introduction, having the physicochemical properties of the species involved in the astrochemical networks is a first basic step toward achieving accurate and reliable modeling. In this article, we complemented the work by Woon & Herbst (2009) on neutral species adding the physicochemical properties of the totality of cations present in the KIDA network<sup>7</sup>. In the following, we discuss a first, immediate application of the combined neutral plus cations

data sets: how the reaction energy, derived from the neutral and cations data sets, could be used to identify reactions that cannot happen in the conditions in the interstellar medium.

This example concerns two reactions, reported in Table 8, which involve  $SiS^+$ . This cation has been postulated to be a precursor leading to  $SiS$ , a species observed in star-forming regions and associated with the molecular shocks of young forming protostars (Tercero et al. 2011; Podio et al. 2017). In

**Table 7**  
Adiabatic Ionization Energy of the Corresponding Neutral Species

Species	eV	eV <sub>(exp)</sub>	Species	eV	eV <sub>(exp)</sub>	Species	eV	eV <sub>(exp)</sub>	Species	eV	eV <sub>(exp)</sub>
H <sub>2</sub> <sup>+</sup>	15.528	15.426 <sup>a</sup>	HNC <sup>+</sup>	11.961	12.500 <sup>ai</sup>	C <sub>3</sub> N <sup>+</sup>	11.816		C <sub>6</sub> H <sup>+</sup>	9.336	
CH <sup>+</sup>	10.562	10.640 <sup>b</sup>	HCO <sup>+</sup>	8.037	8.140 <sup>al</sup>	C <sub>3</sub> O <sup>+</sup>	10.772		HC <sub>5</sub> N <sup>+</sup>	10.558	
NH <sup>+</sup>	13.420	13.490 <sup>c</sup>	SiH <sub>2</sub> <sup>+</sup>	9.101		SiC <sub>2</sub> H <sup>+</sup>	7.289		C <sub>7</sub> <sup>+</sup>	10.293	
OH <sup>+</sup>	12.899	13.017 <sup>d</sup>	HNO <sup>+</sup>	10.165	10.100 <sup>am</sup>	C <sub>3</sub> S <sup>+</sup>	10.188		C <sub>6</sub> N <sup>+</sup>	8.866	
HF <sup>+</sup>	15.998	15.980 <sup>e</sup>	PH <sub>2</sub> <sup>+</sup>	9.826	9.824 <sup>m</sup>	CH <sub>4</sub> <sup>+</sup>	12.749	12.610 <sup>bh</sup>	C <sub>2</sub> H <sub>6</sub> <sup>+</sup>	11.619	11.570 <sup>bw</sup>
C <sub>2</sub> <sup>+</sup>	11.728	11.920 <sup>f</sup>	HO <sub>2</sub> <sup>+</sup>	11.306	11.350 <sup>an</sup>	C <sub>2</sub> H <sub>3</sub> <sup>+</sup>	8.678	8.250 <sup>bi</sup>	C <sub>3</sub> H <sub>5</sub> <sup>+</sup>	8.029	
CN <sup>+</sup>	13.715	14.170 <sup>b</sup>	H <sub>2</sub> S <sup>+</sup>	10.451	10.453 <sup>ao</sup>	SiH <sub>4</sub> <sup>+</sup>	11.050	11.200 <sup>bl</sup>	CH <sub>3</sub> CHOH <sup>+</sup>	6.662	
CO <sup>+</sup>	13.959	14.014 <sup>g</sup>	C <sub>3</sub> <sup>+</sup>	11.606	11.610 <sup>ap</sup>	c-C <sub>3</sub> H <sub>2</sub> <sup>+</sup>	9.118	9.150 <sup>bm</sup>	C <sub>5</sub> H <sub>3</sub> <sup>+</sup>	8.124	
N <sub>2</sub> <sup>+</sup>	15.533	15.581 <sup>h</sup>	C <sub>2</sub> N <sup>+</sup>	10.730	12.000 <sup>aq</sup>	CH <sub>2</sub> CN <sup>+</sup>	10.211	10.300 <sup>bn</sup>	C <sub>6</sub> H <sub>2</sub> <sup>+</sup>	9.477	
SiH <sup>+</sup>	7.900	7.890 <sup>i</sup>	C <sub>2</sub> O <sup>+</sup>	10.912		H <sub>2</sub> CCO <sup>+</sup>	9.538	9.614 <sup>bo</sup>	C <sub>7</sub> H <sup>+</sup>	8.062	
NO <sup>+</sup>	9.191	9.264 <sup>l</sup>	NCO <sup>+</sup>	11.620	11.759 <sup>ar</sup>	SiCH <sub>3</sub> <sup>+</sup>	7.066		HC <sub>6</sub> N <sup>+</sup>	2.494	
PH <sup>+</sup>	10.160	10.149 <sup>m</sup>	HNSi <sup>+</sup>	11.718		HCOOH <sup>+</sup>	11.282	11.310 <sup>bp</sup>	C <sub>8</sub> <sup>+</sup>	11.376	
O <sub>2</sub> <sup>+</sup>	12.053	12.070 <sup>n</sup>	HCP <sup>+</sup>	10.786	10.790 <sup>as</sup>	C <sub>4</sub> H <sup>+</sup>	10.057		C <sub>7</sub> N <sup>+</sup>	9.659	
HS <sup>+</sup>	10.310	10.421 <sup>o</sup>	CO <sub>2</sub> <sup>+</sup>	13.765	13.778 <sup>at</sup>	HC <sub>3</sub> N <sup>+</sup>	11.620		CH <sub>3</sub> OCH <sub>3</sub> <sup>+</sup>	10.004	10.025 <sup>bz</sup>
HCl <sup>+</sup>	12.716	12.790 <sup>p</sup>	HCS <sup>+</sup>	7.588	7.412 <sup>au</sup>	C <sub>5</sub> <sup>+</sup>	11.116	12.300 <sup>bq</sup>	C <sub>4</sub> H <sub>5</sub> <sup>+</sup>	7.066	
SiC <sup>+</sup>	8.853	9.000 <sup>q</sup>	NO <sub>2</sub> <sup>+</sup>	9.458	9.600 <sup>av</sup>	C <sub>4</sub> N <sup>+</sup>	9.514		C <sub>8</sub> H <sup>+</sup>	8.836	
SiN <sup>+</sup>	10.342		HPO <sup>+</sup>	10.548		SiC <sub>3</sub> H <sup>+</sup>	8.181		HC <sub>7</sub> N <sup>+</sup>	9.825	
CP <sup>+</sup>	10.914	10.500 <sup>r</sup>	SiNC <sup>+</sup>	7.851		SiC <sub>4</sub> <sup>+</sup>	10.099		C <sub>9</sub> <sup>+</sup>	3.734	
CS <sup>+</sup>	11.381	11.330 <sup>s</sup>	C <sub>2</sub> S <sup>+</sup>	10.186		C <sub>4</sub> P <sup>+</sup>	8.447		C <sub>8</sub> N <sup>+</sup>	8.430	
PN <sup>+</sup>	11.933	11.880 <sup>t</sup>	OCS <sup>+</sup>	11.189	11.185 <sup>aw</sup>	C <sub>4</sub> S <sup>+</sup>	9.199		C <sub>2</sub> H <sub>6</sub> CO <sup>+</sup>	9.675	9.700 <sup>ca</sup>
NS <sup>+</sup>	8.916	8.870 <sup>u</sup>	HSiS <sup>+</sup>	8.260		C <sub>2</sub> H <sub>4</sub> <sup>+</sup>	10.485	10.510 <sup>br</sup>	C <sub>5</sub> H <sub>5</sub> <sup>+</sup>	7.718	
PO <sup>+</sup>	8.476	8.390 <sup>v</sup>	SO <sub>2</sub> <sup>+</sup>	12.680	12.500 <sup>az</sup>	CH <sub>3</sub> OH <sup>+</sup>	10.928	10.850 <sup>bs</sup>	C <sub>6</sub> H <sub>4</sub> <sup>+</sup>	8.993	
CCl <sup>+</sup>	8.834	8.900 <sup>w</sup>	HS <sub>2</sub> <sup>+</sup>	9.376		CH <sub>3</sub> CN <sup>+</sup>	12.246	12.201 <sup>bt</sup>	C <sub>8</sub> H <sub>2</sub> <sup>+</sup>	8.978	
SiO <sup>+</sup>	11.521	11.300 <sup>z</sup>	CH <sub>3</sub> <sup>+</sup>	9.736	9.843 <sup>ba</sup>	C <sub>4</sub> H <sub>2</sub> <sup>+</sup>	10.131		C <sub>9</sub> H <sup>+</sup>	7.742	
SO <sup>+</sup>	10.428	10.294 <sup>aa</sup>	NH <sub>3</sub> <sup>+</sup>	10.148	10.020 <sup>bb</sup>	C <sub>5</sub> H <sup>+</sup>	8.358		HC <sub>8</sub> N <sup>+</sup>	8.387	
SiS <sup>+</sup>	10.526	10.530 <sup>ab</sup>	C <sub>2</sub> H <sub>2</sub> <sup>+</sup>	11.330	11.410 <sup>bc</sup>	HC <sub>4</sub> N <sup>+</sup>	9.425		C <sub>10</sub> <sup>+</sup>	10.647	
S <sub>2</sub> <sup>+</sup>	9.503	9.400 <sup>ac</sup>	H <sub>2</sub> CO <sup>+</sup>	10.900	10.880 <sup>bd</sup>	C <sub>6</sub> <sup>+</sup>	12.441		C <sub>9</sub> N <sup>+</sup>	9.168	
CH <sub>2</sub> <sup>+</sup>	10.338	10.350 <sup>ad</sup>	SiH <sub>3</sub> <sup>+</sup>	8.109	8.170 <sup>be</sup>	C <sub>5</sub> N <sup>+</sup>	10.705		C <sub>10</sub> H <sup>+</sup>	8.510	
NH <sub>2</sub> <sup>+</sup>	11.131	10.780 <sup>ae</sup>	HNCO <sup>+</sup>	11.557	11.595 <sup>bf</sup>	SiC <sub>4</sub> H <sup>+</sup>	7.082		C <sub>11</sub> <sup>+</sup>	8.984	
H <sub>2</sub> O <sup>+</sup>	12.574	12.650 <sup>af</sup>	H <sub>2</sub> CS <sup>+</sup>	9.411	9.376 <sup>bg</sup>	C <sub>2</sub> H <sub>5</sub> <sup>+</sup>	8.063	8.117 <sup>bu</sup>	C <sub>10</sub> N <sup>+</sup>	8.026	
C <sub>2</sub> H <sup>+</sup>	11.267	11.610 <sup>ag</sup>	H <sub>2</sub> SiO <sup>+</sup>	10.674		CH <sub>3</sub> NH <sub>2</sub> <sup>+</sup>	9.061	8.900 <sup>bv</sup>	C <sub>10</sub> H <sub>2</sub> <sup>+</sup>	8.722	
HCN <sup>+</sup>	13.543	13.590 <sup>ah</sup>	C <sub>4</sub> <sup>+</sup>	10.734		C <sub>4</sub> H <sub>3</sub> <sup>+</sup>	7.960		HC <sub>10</sub> N <sup>+</sup>	8.094	

**Notes.** Adiabatic ionization energy (eV) without ZPE correction, compared with the vertical experimental ionization energy (eV<sub>(exp)</sub>).

<sup>a</sup> (Shiner et al. 1993). <sup>b</sup> (Huber 2013). <sup>c</sup> (Dyke et al. 1980). <sup>d</sup> (Wiedmann et al. 1992). <sup>e</sup> (Tiedemann et al. 1979). <sup>f</sup> (Plessis & Marnet 1987). <sup>g</sup> (Erman et al. 1993). <sup>h</sup> (Trickl et al. 1989). <sup>i</sup> (Boo & Armentrout 1987). <sup>l</sup> (Reiser et al. 1988). <sup>m</sup> (Berkowitz & Cho 1989). <sup>n</sup> (Tonkyn et al. 1989). <sup>o</sup> (Milan et al. 1996). <sup>p</sup> (Wang et al. 1984). <sup>q</sup> (Verhaegen et al. 1964). <sup>r</sup> (Smoes et al. 1971). <sup>s</sup> (Drowart et al. 1978). <sup>t</sup> (Bulgin et al. 1977). <sup>u</sup> (Dyke et al. 1977). <sup>v</sup> (Dyke et al. 1982). <sup>w</sup> (Hepburn et al. 1982). <sup>z</sup> (Nakasgawa et al. 1981). <sup>aa</sup> (Norwood & Ng 1989a). <sup>ab</sup> (Cockett et al. 1989). <sup>ac</sup> (Bender et al. 1988). <sup>ad</sup> (Reineke & Strein 1976). <sup>ae</sup> (Qi et al. 1995). <sup>af</sup> (Snow & Thomas 1990). <sup>ag</sup> (Norwood & Ng 1989b). <sup>ah</sup> (Dibeler & Liston 1968). <sup>ai</sup> (Bieri & Jonsson 1978). <sup>al</sup> (Dyke 1987). <sup>am</sup> (Lias et al. 1988). <sup>an</sup> (Tang et al. 2020). <sup>ao</sup> (Walters & Blais 1984). <sup>ap</sup> Rohlfling et al. (1984); only the upper limit value was reported. <sup>aq</sup> (Lias et al. 1988). <sup>ar</sup> (Dyke 1987). <sup>as</sup> (Frost et al. 1973). <sup>at</sup> (Wang et al. 1988). <sup>au</sup> (Ruscic & Berkowitz 1993). <sup>av</sup> (Clemmer & Armentrout 1992). <sup>aw</sup> (Wang et al. 1988). <sup>az</sup> (Snow & Thomas 1990). <sup>ba</sup> (Berkowitz et al. 1994). <sup>bb</sup> (Qi et al. 1995). <sup>bc</sup> (Plessis & Marnet 1986). <sup>bd</sup> (Ohno et al. 1995). <sup>be</sup> (Nagano et al. 1993). <sup>bf</sup> (Ruscic & Berkowitz 1994). <sup>bg</sup> (Ruscic & Berkowitz 1993). <sup>bh</sup> (Berkowitz et al. 1987). <sup>bi</sup> (Blush & Chen 1992). <sup>bj</sup> (Shin et al. 1990). <sup>bm</sup> (Clauberg et al. 1992). <sup>bn</sup> (Holmes et al. 1993). <sup>bo</sup> (Vogt et al. 1978). <sup>bp</sup> (Traeger 1985). <sup>bq</sup> (Ramanathan et al. 1993). <sup>br</sup> (Ohno et al. 1995). <sup>bs</sup> (Tao et al. 1992). <sup>bt</sup> (Gochel-Dupuis et al. 1992). <sup>bu</sup> (Ruscic et al. 1989). <sup>bv</sup> (Aue & Bowers 1979). <sup>bw</sup> (Plessis & Marnet 1987). <sup>bz</sup> (Butler et al. 1984). <sup>ca</sup> (Traeger et al. 1982).

The calculated adiabatic ionization energies for C<sub>4</sub>, C<sub>6</sub>, C<sub>8</sub>, and C<sub>10</sub> have been computed with respect to the triplet state of these species and not to the singlet state, as reported in Woon & Herbst (2009).

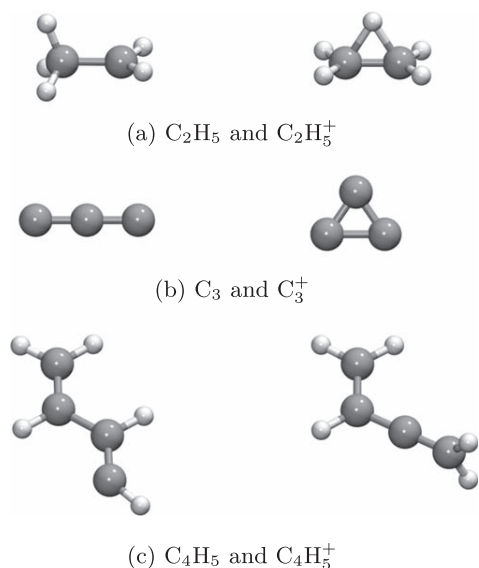
contrast to the competing S-bearing species SiO, which is likely extracted from the shattered grains, SiS is thought to be a product of gas-phase reactions (Podio et al. 2017). The KIDA database only lists one reaction forming SiS: HSiS<sup>+</sup> + e → SiS + H. In turn, according to KIDA, HSiS<sup>+</sup> is formed by the first reaction in Table 8: H<sub>2</sub> + SiS<sup>+</sup> → H + HSiS<sup>+</sup>. Our computations, coupled with those by Woon & Herbst (2009), clearly show that this reaction is highly endothermic (~104 kJ mol<sup>-1</sup>) and, consequently, rule out the formation of SiS by the recombination of HSiS<sup>+</sup>. Previous experimental work by Wlodek & Bohme (1989) support our conclusion. Therefore,

these two reactions (the formation of SiS from HSiS<sup>+</sup>) should be removed by the astrochemical reaction databases.

The second reaction in Table 8 involves the formation of SiS<sup>+</sup>, which would be the step before the above SiS reaction formation, according to the KIDA database (on the contrary, the UMIST database does not report the reaction). Also, in this case, our calculations show that the reaction is endothermic (~36 kJ mol<sup>-1</sup>) and, therefore, should be removed from the database.

For the curious reader, other routes of SiS formation, involving neutral-neutral reactions, have been explored in the literature





**Figure 4.** Connectivity change between cations and the neutral species.

**Table 8**

Two Examples of Endothermic Reactions Found in the Astrochemical Reaction Network for Molecular Clouds

Reaction	$\Delta E$ [kJ mol <sup>-1</sup> ]
$H_2 + SiS^+ \rightarrow H + HSiS^+$	103.6
$SiS + S^+ \rightarrow SiS^+ + S$	36.1

**Note.** The reaction energy (i.e.,  $\Delta E$ , no ZPE correction) is computed with data from this work and Woon & Herbst (2009).

since the work of Podio et al. (2017) and found to be plausible (Rosi et al. 2018; Zanchet et al. 2018; Rosi et al. 2019).

## 5. Conclusions

In this work, we present new ab initio calculations of the structure and energy of 262 cations, all appearing in the used

astrochemical reaction network databases KIDA and UMIST. Our calculations complement the previous work by Woon & Herbst (2009), who reported the same properties for an ensemble of 200 neutral species. The rationale behind our new calculations is that accurate knowledge of the physico-chemical properties of the species in the reaction network databases is a first mandatory step to improving the reliability of the astrochemical models.

All the computed data can be found on the ACO project site.<sup>15</sup>

Finally, we discussed two practical examples to illustrate the potentiality of using our new cations database, coupled with the Woon & Herbst (2009) one, to identify and exclude endothermic reactions from the astrochemical reaction networks.

This project has received funding within the European Union’s Horizon 2020 research and innovation program from the European Research Council (ERC) for the project “The Dawn of Organic Chemistry” (DOC), grant agreement No. 741002, and from the Marie Skłodowska-Curie for the project “Astro-Chemical Origins” (ACO), grant agreement No. 811312. S.P., N.B., and P.U. acknowledge the Italian Space Agency for cofunding the Life in Space Project (ASI N. 2019-3-U.O). CINES-OCCIGEN HPC is kindly acknowledged for the generous allowance of supercomputing time through the A0060810797 project. Finally, we wish to acknowledge the extremely useful discussions with Professor Grotobape and the LATEX community for the insights on TikZ and PGFPlots packages.

**Software:** Rdkit (Landrum 2016), ASE (Larsen et al. 2017), NetworkX (Hagberg et al. 2008), VMD (Humphrey et al. 1996), JSmol, Gaussian16 (Frisch et al. 2016).

## Appendix

### Extended Internal Coordinates Geometry Optimization Comparison

An extended internal coordinates geometry optimization comparison is provided in Table 9. The coordinates given in Table 9 follow those expressions in Gaussian16 for specifying generalized internal coordinates.<sup>16</sup> The complete set of coordinate comparisons is available in a machine-readable format.

**Table 9**

Extended Internal Coordinates Geometry Optimization Comparison for  $1-C_3H_2^+$

Species	Internal Z Vec.	Internal Coord.	M06-2X/cc-pVTZ	CCSD/aug-cc-pVTZ	$\Delta$	Unit
$1-C_3H_2^+$	(1,6,6,6,1)	R(1,2)	1.093	1.092	0.001	Å
$1-C_3H_2^+$	(1,6,6,6,1)	R(2,3)	1.312	1.314	-0.002	Å
$1-C_3H_2^+$	(1,6,6,6,1)	R(2,5)	1.094	1.092	0.002	Å
$1-C_3H_2^+$	(1,6,6,6,1)	R(3,4)	1.340	1.351	-0.011	Å
$1-C_3H_2^+$	(1,6,6,6,1)	A(1,2,3)	120.665	120.424	0.241	Å
$1-C_3H_2^+$	(1,6,6,6,1)	A(1,2,5)	118.903	119.111	-0.208	Å
$1-C_3H_2^+$	(1,6,6,6,1)	A(3,2,5)	120.431	120.464	-0.032	Å
$1-C_3H_2^+$	(1,6,6,6,1)	L(2,3,4,5, -1)	172.977	179.999	-7.021	Å
$1-C_3H_2^+$	(1,6,6,6,1)	L(2,3,4,5, -2)	180.0	180.0	0	degrees
$1-C_3H_2^+$	(1,6,6,6,1)	D(1,2,5,3)	180.0	180.0	0	degrees

**Note.** Table 9 is published in its entirety in a machine-readable format. A portion is shown here for guidance regarding its form and content.

(This table is available in its entirety in machine-readable form.)

<sup>15</sup> [aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/cations-database](http://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations/cations-database)

<sup>16</sup> Gaussian16 Revision C.01, <https://Gaussian.com/gic/>.

## ORCID iDs

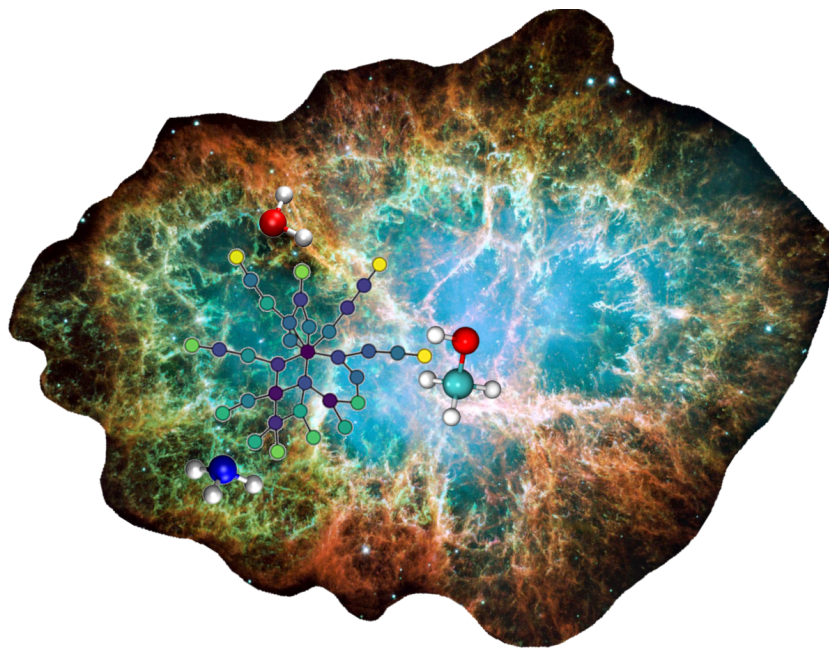
Lorenzo Tinacci  <https://orcid.org/0000-0001-9909-9570>  
 Stefano Pantaleone  <https://orcid.org/0000-0002-2457-1065>  
 Andrea Maranzana  <https://orcid.org/0000-0002-5524-8068>  
 Nadia Balucani  <https://orcid.org/0000-0001-5121-5683>  
 Cecilia Ceccarelli  <https://orcid.org/0000-0001-9664-6292>  
 Piero Ugliengo  <https://orcid.org/0000-0001-8886-9832>

## References

- Aue, D. H., & Bowers, M. T. 1979, *Gas Phase Ion Chemistry* (Amsterdam: Elsevier)
- Bauenschmitt, R., & Ahlrichs, R. 1996, *JChPh*, **104**, 9047
- Bender, H., Carnovale, F., Peel, J. B., & Wentrup, C. 1988, *JChS*, **110**, 3458
- Berkowitz, J., & Cho, H. 1989, *JChPh*, **90**, 1
- Berkowitz, J., Ellison, G. B., & Gutman, D. 1994, *JPhCh*, **98**, 2744
- Berkowitz, J., Greene, J., Cho, H., & Ruscic, B. 1987, *JChPh*, **86**, 674
- Bieri, G., & Jonsson, B.-Ö. 1978, *CPL*, **56**, 446
- Blush, J. A., & Chen, P. 1992, *JPhCh*, **96**, 4138
- Boo, B. H., & Armentrout, P. 1987, *JChS*, **109**, 3549
- Bulgin, D. K., Dyke, J. M., & Morris, A. 1977, *J. Chem. Soc., Faraday Trans.*, **73**, 983
- Butler, J. J., Holland, D. M., Parr, A. C., & Stockbauer, R. 1984, *IJMSI*, **58**, 1
- Cernicharo, J., Cabezas, C., Baillieux, S., et al. 2021b, *A&A*, **646**, L7
- Cernicharo, J., Cabezas, C., Endo, Y., et al. 2021a, *A&A*, **646**, L3
- Chabot, M., Béroff, K., Gratier, P., Jallat, A., & Wakelam, V. 2013, *ApJ*, **771**, 90
- Chase, M. W. 1996, *JPCRD*, **25**, 551
- Cheung, A., Rank, D. M., Townes, C., Thornton, D. D., & Welch, W. 1968, *PhRvL*, **21**, 1701
- Cheung, A., Rank, D. M., Townes, C., Thornton, D. D., & Welch, W. 1969, *Natur*, **221**, 626
- Clauberg, H., Minsek, D. W., & Chen, P. 1992, *JChS*, **114**, 99
- Clemmer, D., & Armentrout, P. 1992, *JChPh*, **97**, 2451
- Cockett, M. C., Dyke, J. M., Morris, A., & Niavarani, M. H. Z. 1989, *J. Chem. Soc., Faraday Trans.*, **85**, 75
- Dibeler, V. H., & Liston, S. K. 1968, *JChPh*, **48**, 4765
- Douglas, J. E., Rabinovitch, B. t., & Looney, F. 1955, *JChPh*, **23**, 315
- Drowart, J., Smets, J., Reynaert, J., & Coppens, P. 1978, *Adv. Mass Spectrom.* **A**, **7**, 647
- Dyke, J., Dunlavey, S., Jonathan, N., & Morris, A. 1980, *MolPh*, **39**, 1121
- Dyke, J. M. 1987, *J. Chem. Soc., Faraday Trans.*, **83**, 69
- Dyke, J. M., Morris, A., & Ridha, A. 1982, *J. Chem. Soc., Faraday Trans.*, **78**, 2077
- Dyke, J. M., Morris, A., & Trickle, I. R. 1977, *J. Chem. Soc., Faraday Trans.*, **73**, 147
- Erman, P., Karawajczyk, A., Rachlew-Källne, E., et al. 1993, *CPL*, **215**, 173
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., et al. 2016, *Gaussian Revision B.01* (Wallingford, CT: Gaussian, Inc.)
- Frost, D., Lee, S., & McDowell, C. 1973, *CPL*, **23**, 472
- Gochel-Dupuis, M., Delwiche, J., Hubin-Franskin, M.-J., & Collin, J. 1992, *CPL*, **193**, 41
- Hagberg, A. A., Schult, D. A., & Swart, P. J. 2008, *Proc. 7th Python in Science Conf.*, ed. G. Varoquaux, T. Vaught, & J. Millman, **11**, [http://conference.scipy.org/proceedings/scipy2008/paper\\_2/](http://conference.scipy.org/proceedings/scipy2008/paper_2/)
- Hepburn, J., Trevor, D., Pollard, J., Shirley, D., & Lee, Y. 1982, *JChPh*, **76**, 4287
- Herbst, E., & Klemperer, W. 1973, *ApJ*, **185**, 505
- Herzberg, G. 1966, *Electronic Spectra and Electronic Structure of Polyatomic Molecules*, Vol. 3 (New York: Van Nostrand)
- Holmes, J., Lossing, F., & Mayer, P. 1993, *CPL*, **212**, 134
- Huber, K.-P. 2013, *Molecular Spectra and Molecular Structure: IV. Constants of Diatomic Molecules* (Berlin: Springer Science & Business Media)
- Humphrey, W., Dalke, A., & Schulten, K. 1996, *J. Mol. Graph.*, **14**, 33
- Kendall, T., Jr 1992, *JChPh*, **96**, 6796
- Klemperer, W. 1970, *Natur*, **227**, 1230
- Knowles, P. J., Hampel, C., & Werner, H.-J. 1993, *JChPh*, **99**, 5219
- Landrum, G. 2016, *RDKit: Open-Source Cheminformatics Software*, <https://www.rdkit.org/>
- Larsen, A. H., Mortensen, J. J., Blomqvist, J., et al. 2017, *JPCM*, **29**, 273002
- Lattalais, M., Pauzat, F., Ellinger, Y., & Ceccarelli, C. 2009, *ApJL*, **696**, L133
- Lattalais, M., Pauzat, F., Ellinger, Y., & Ceccarelli, C. 2010, *A&A*, **519**, A30
- Lias, S. G., Bartmess, J. E., Liebman, J., Holmes, J., & Levin, R. D. 1988, *JPCRD*, **17**, Suppl. 1, <https://srdb.nist.gov/JPCRD/jpcrdS1Vol17.pdf>
- McGuire, B. A., Asvany, O., Brünken, S., & Schlemmer, S. 2020, *NatRP*, **2**, 402
- Milan, J., Buma, W., & De Lange, C. 1996, *JChPh*, **104**, 521
- Nagano, Y., Murthy, S., & Beauchamp, J. 1993, *JChS*, **115**, 10805
- Nakasgawa, H., Asano, M., & Kubo, K. 1981, *JNuM*, **102**, 292
- Norwood, K., & Ng, C. 1989a, *CPL*, **156**, 145
- Norwood, K., & Ng, C. 1989b, *JChPh*, **91**, 2898
- Ohno, K., Okamura, K., Yamakado, H., et al. 1995, *JPhCh*, **99**, 14247
- Plessis, P., & Marmet, P. 1986, *IJMSI*, **70**, 23
- Plessis, P., & Marmet, P. 1987, *CajCh*, **65**, 2004
- Podio, L., Codella, C., Lefloch, B., et al. 2017, *MNRAS*, **470**, L16
- Qi, F., Sheng, L., Zhang, Y., Yu, S., & Li, W.-K. 1995, *CPL*, **234**, 450
- Ramanathan, R., Zimmerman, J. A., & Eyer, J. R. 1993, *JChPh*, **98**, 7838
- Rappé, A. K., Casewit, C. J., Colwell, K., Goddard, W. A., III, & Skiff, W. M. 1992, *JChS*, **114**, 10024
- Reineke, W., & Strein, K. 1976, *Berichte der Bunsengesellschaft für physikalische Chemie*, **80**, 343
- Reiser, G., Habenicht, W., Müller-Dethlefs, K., & Schlag, E. W. 1988, *CPL*, **152**, 119
- Rohlfing, E. A., Cox, D. M., & Kaldor, A. 1984, *JChPh*, **81**, 3322
- Rosi, M., Mancini, L., Skouteris, D., et al. 2018, *CPL*, **695**, 87
- Rosi, M., Skouteris, D., Balucani, N., et al. 2019, *International Conference on Computational Science and Its Applications* (Berlin: Springer), 306
- Ruscic, B., & Berkowitz, J. 1993, *JChPh*, **98**, 2568
- Ruscic, B., & Berkowitz, J. 1994, *JChPh*, **100**, 4498
- Ruscic, B., Berkowitz, J., Curtiss, L., & Pople, J. 1989, *JChPh*, **91**, 114
- Shavitt, I. 1985, *Tetrahedron*, **41**, 1531
- Shin, S. K., Corderman, R. R., & Beauchamp, J. 1990, *IJMSI*, **101**, 257
- Shiner, D., Gilligan, J., Cook, B., & Lichten, W. 1993, *PhRvA*, **47**, 4042
- Smoes, S., Myers, C., & Drowart, J. 1971, *CPL*, **8**, 10
- Snow, K. B., & Thomas, T. F. 1990, *IJMSI*, **96**, 49
- Snyder, L., Le, S., & Jm, H. 1976, *Reprints: Series A, Vol. 212* (Green Bank, WV: National Radio Astronomy Observatory), 383
- Snyder, L. E., Buhl, D., Zuckerman, B., & Palmer, P. 1969, *PhRvL*, **22**, 679
- Tang, X., Lin, X., Garcia, G. A., et al. 2020, *JChPh*, **153**, 124306
- Tao, W., Klemm, R., Nesbitt, F., & Stief, L. 1992, *JPhCh*, **96**, 104
- Tercero, B., Vincent, L., Cernicharo, J., Viti, S., & Marcelino, N. 2011, *A&A*, **528**, A26
- Tiedemann, P., Anderson, S., Ceyer, S., et al. 1979, *JChPh*, **71**, 605
- Tonkyn, R. G., Winniczek, J. W., & White, M. G. 1989, *CPL*, **164**, 137
- Traeger, J. C. 1985, *IJMSI*, **66**, 271
- Traeger, J. C., McLoughlin, R. G., & Nicholson, A. 1982, *JChS*, **104**, 5318
- Trickl, T., Cromwell, E., Lee, Y., & Kung, A. 1989, *JChPh*, **91**, 6006
- Trinajstić, N. 2018, *Chemical Graph Theory* (London: Routledge)
- Verhaegen, G., Stafford, F., & Drowart, J. 1964, *JChPh*, **40**, 1622
- Vogt, J., Williamson, A. D., & Beauchamp, J. 1978, *JChS*, **100**, 3478
- Wakelam, V., Herbst, E., Loison, J.-C., et al. 2012, *ApJS*, **199**, 21
- Walters, E., & Blais, N. C. 1984, *JChPh*, **80**, 3501
- Wang, L.-S., Reutt, J., Lee, Y., & Shirley, D. 1988, *JESRP*, **47**, 167
- Wang, R.-G., Dillon, M., & Spence, D. 1984, *JChPh*, **80**, 63
- Watson, W. D. 1973, *ApJL*, **183**, L17
- Watts, J. D., Gauss, J., & Bartlett, R. J. 1993, *JChPh*, **98**, 8718
- Wiedmann, R., Tonkyn, R., White, M., Wang, K., & McKoy, V. 1992, *JChPh*, **97**, 768
- Wlodek, S., & Bohme, D. K. 1989, *J. Chem. Soc., Faraday Trans.*, **85**, 1643
- Woon, D. E., & Dunning, T. H., Jr 1993, *JChPh*, **98**, 1358
- Woon, D. E., & Herbst, E. 2009, *ApJS*, **185**, 273
- Zanchet, A., Roncero, O., Agúndez, M., & Cernicharo, J. 2018, *ApJ*, **862**, 38
- Zhao, Y., & Truhlar, D. G. 2008, *Theor. Chem. Acc.*, **120**, 215

# The GRETOBAPE gas-phase reaction network: the importance of being exothermic

The hereafter article is currently under review and accepted for publication in The Astrophysical Journal Supplement Series.





## The GRETOBAPE gas-phase reaction network: the importance of being exothermic

LORENZO TINACCI <sup>1,2</sup> SIMÓN FERRADA-CHAMORRO <sup>1</sup> CECILIA CECCARELLI <sup>1</sup> STEFANO PANTALEONE <sup>2,3</sup>  
DANIELA ASCENZI <sup>4</sup> ANDREA MARAZZANA <sup>2</sup> NADIA BALUCANI <sup>1,3,5</sup> AND PIERO UGLIENGO <sup>2</sup>

<sup>1</sup> *Université Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France*

<sup>2</sup> *Dipartimento di Chimica and Nanostructured Interfaces and Surfaces (NIS) Centre,  
Università degli Studi di Torino, via P. Giuria 7, 10125 Torino, Italy*

<sup>3</sup> *Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, 06123 Perugia, Italy*

<sup>4</sup> *Dipartimento di Fisica, Università di Trento, 38123 Povo, Italy*

<sup>5</sup> *Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy*

Submitted to ApJS

### ABSTRACT

The gas-phase reaction networks are the backbone of astrochemical models. However, due to their complexity and non-linear impact on the astrochemical modeling, they can be the first source of error in the simulations if incorrect reactions are present. Over time, following the increasing number of species detected, astrochemists have added new reactions, based on laboratory experiments and quantum mechanics (QM) computations as well as reactions inferred by chemical intuition and similarity principle. However, sometimes no verification of their feasibility in the interstellar conditions, namely their exothermicity, was performed. In this work, we present a new gas-phase reaction network, GRETOBAPE, based on the KIDA2014 network and updated with several reactions, cleaned from endothermic reactions not explicitly recognized as such. To this end, we characterized all the species in the GRETOBAPE network with accurate QM calculations. We found that  $\sim 5\%$  of the reactions in the original network are endothermic although most of them are reported as barrierless. The reaction network of Si-bearing species is the most impacted by the endothermicity cleaning process. We also produced a cleaned reduced network, GRETOBAPE-RED, to be used to simulate astrochemical situations where only C-, O-, N- and S- bearing species with less than 6 atoms are needed. Finally, the new GRETOBAPE network, its reduced version, as well as the database with all the molecular properties are made publicly available. The species properties database can be used in the future to test the feasibility of possibly new reactions.

*Keywords:* astrochemistry — ISM: Reaction network; astrochemistry; cold molecular cloud; molecules

### 1. INTRODUCTION

Since the discovery of the first molecules in the interstellar medium (ISM) in the visible spectrum (Swings & Rosenfeld 1937; Dunham Jr 1937; McKellar 1940), the astrochemical community has been trying to understand their origin and how they can survive the harsh interstellar conditions. With the evolution of astronomical facilities and, more specifically, with the advent of radio

telescopes, new and more complex molecules have been detected in various astronomical objects with an almost constant discovery rate of  $\sim 4 - 6$  new molecules per year (McGuire 2022). To date, more than 270 molecules, composed of 19 different elements, have been detected in the interstellar and circumstellar medium, and this number is steadily increasing. The list of detected interstellar species can be found in the Cologne Database for Molecular Spectroscopy (CDMS) (Endres et al. 2016)<sup>1</sup>,

Corresponding author: Lorenzo Tinacci  
Lorenzo.Tinacci@univ-grenoble-alpes.fr

<sup>1</sup> <https://cdms.astro.uni-koeln.de/classic/molecules>

in the [McGuire \(2022\)](#) census (and the related Python package<sup>2</sup>), and in The Astrochymist website<sup>3</sup>.

Understanding how interstellar molecules are formed, destroyed and connected, in addition to be interesting per se, can help inferring the properties, history, and evolution of the astronomical objects in which they have been detected. Thus, in parallel with the identification of molecules and the measurements of their abundances in different astronomical objects, soon a community dedicated to developing astrochemical models began to play an important role ([Herbst & Klemperer 1973](#); [Prasad & Huntress Jr 1980](#)). A crucial element of astrochemical models is the gas-phase reaction network, namely the lists of the gas-phase chemical reactions with their rate constants, that describe how efficient the reactions are as a function of the temperature, and product branching ratios.

The increasing number of detected species has prompted modelers and chemists to add new reactions to the existing networks in order to reproduce the observed species and their abundances. In this way, the number of reactions, reactants and products have been increasing each time a new molecule was detected, which also led to an increase in the network size and complexity. Nowadays, the publicly available astrochemical gas-phase networks, KIDA<sup>4</sup> and UMIST<sup>5</sup>, contains about 8000 reactions involving more than 500 species.

Unfortunately, the vast majority ( $\geq 80\%$ ) of the rates and branching ratios of the reported gas-phase reactions has not been measured or computed and is often based on approximate estimates (*e.g.* using the Capture Theory: [Su & Chesnavich 1982](#); [Herbst 2006](#); [Woon & Herbst 2009](#); [Loison et al. 2013](#)), educated guesses on similarity principles or simple chemical intuition. In addition, even when some experimental measurements are available, the estimated rate constants may have substantial uncertainties as they are often based on experiments at room temperature (*e.g.* [Anicich 2003](#)). The uncertainty on the reaction rates is a well-known problem in the astrochemistry community. In parallel with the efforts of studying reactions via experiments and theoretical computations, sensitivity analysis studies have provided the impact of these uncertainties on the abundances predicted by astrochemical models (*e.g.* [Wakelam et al. 2006](#); [Penteado et al. 2017](#)).

One possible and mandatory first step to improve the reliability of the gas-phase reaction networks, that has never been performed systematically on the astrochemical networks is to estimate the exo/endothemicity of the gas-phase reactions, using accurate physico-chemical quantum mechanical (QM) data for each species. Indeed, given the ISM conditions (temperatures usually less than  $\sim 100$ - $200$  K), strongly endothermic reactions are inhibited and, therefore, can be excluded from the astrochemical networks without complex and time-consuming reaction transition states studies or experiments ([Smith 2006, 2011](#)). This verification is now possible, thanks to the availability of two theoretical studies that together provide reliable QM data (electronic state, electron spin multiplicity, geometry, harmonic frequencies, absolute electronic energy, and dipole moment) for the totality of the species in the KIDA network. The first one by [Woon & Herbst \(2009\)](#) computed the QM data for a large number of neutral molecules involved in the KIDA reaction network. The second study by [Tinacci et al. \(2021\)](#) complemented the first one and provided QM data for all the ions and the remaining neutral species of the KIDA network. Please note that the two studies have been carried out at the same electronic level of theory, so that they can be reliably used for computing the exo/endothemicity of all the reactions. Also, [Tinacci et al. \(2021\)](#) created a database<sup>6</sup> with the above QM data for all the cations.

The KIDA and UMIST networks contain such a large number of reactions because their goal is to reproduce the observed abundances of large molecules, which have usually low ( $\leq 10^{-12}$ ) abundances. However, in some cases, such as astrochemical models coupled with hydrodynamical simulations of cloud collapse or protoplanetary disks formation and evolution, the goal is to reliably reproduce only the most abundant species. In these cases, the use of large reactions networks is not only useless but also detrimental, because it demands spareable computing time. Indeed, several authors have proposed reaction network reduction techniques in the last few years, each using different criteria and focusing on specific goals ([Oppenheimer & Dalgarno 1974](#); [Nelson & Langer 1999](#); [Glover et al. 2010](#); [Grassi et al. 2013, 2021](#)).

The goal of the present article is twofold. The *first goal* is to produce a gas-phase reaction network "cleaned" from endothermic reactions not reported as such. To this purpose, we perform a systematic study of the

<sup>2</sup> <https://github.com/bmcguir2/astromol>

<sup>3</sup> [http://www.astrochymist.org/astrochymist\\_ism.html](http://www.astrochymist.org/astrochymist_ism.html)

<sup>4</sup> <http://kida.obs.u-bordeaux1.fr/>

<sup>5</sup> <http://udfa.ajmarkwick.net/>

<sup>6</sup> <https://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations>

**Table 1.** ZPE-corrected reaction energy difference, *i.e.*,  $\Delta(\Delta H(0))$ , between previous literature and our data.

Reactants	Reaction Products	$\Delta(\Delta H(0))$ [kJ/mol]	Their Level of Theory	Reference
C <sub>2</sub> H <sub>5</sub> + O	C <sub>2</sub> H <sub>4</sub> + OH	1.26		
C <sub>2</sub> H <sub>5</sub> + O	CH <sub>3</sub> CHO + H	1.67	CCSD(T)/aug-cc-pVTZ//B2PYLYP/aug-cc-pVTZ	Vazart et al. (2020)
C <sub>2</sub> H <sub>5</sub> + O	H <sub>2</sub> CO + CH <sub>3</sub>	1.23		
CH + CH <sub>3</sub> OH	C <sub>2</sub> H <sub>4</sub> + OH	0.36		
CH + CH <sub>3</sub> OH	CH <sub>3</sub> CHO + H	0.77	CCSD(T)/aug-cc-pVTZ//B2PYLYP/aug-cc-pVTZ	Vazart et al. (2020)
CH + CH <sub>3</sub> OH	H <sub>2</sub> CO + CH <sub>3</sub>	0.34		
SiH + S	SiS + H	4.29	CCSD(T)/aug-cc-pVTZ//B3LYP/aug-cc-pVTZ	Rosi et al. (2018)
SiH + S <sub>2</sub>	SiS + HS	-1.85		
Si + SH	SiH + S	-3.92	MRCI/CBS	Mota et al. (2021)

NOTE— The minor differences with respect to Vazart et al. (2020) and Rosi et al. (2018) can be ascribed to the different optimization levels and also the spin contamination for all open-shell species, which was corrected in our case by using the Restricted-Open formalism.

endo/exo thermicity of all reactions present in a new network, GRETOBAPE, which is based on the KIDA2014 one and updated through the years by our group. In the process, we also verify that each species present in the network is not a sink and that the obvious reactions with the most abundant molecular cations are present. The *second goal* is to produce a reduced network from the cleaned one, which only includes H-, C-, N-, O-, and S-bearing species with not more than six atoms, to be employed when the use of a complete network is unnecessary.

The article is organized as follows. Section 2 details the adopted tools to deal with the reaction network, the cleaning procedure, and the computational methodology used to characterize all the chemical species present in the network. In Section 3, we describe the new reaction networks, GRETOBAPE and GRETOBAPE-RED, respectively resulting from the cleaning and reduction processes, as well as a list of the obvious reactions possibly missing in the network. In Section 4, we discuss the obtained results and their impact on astrochemical models of cold molecular clouds and warm molecular outflow shocks simulations. Finally, in Section 5, we present our conclusions, and in Section 6 we provide the hyperlinks to the publicly available online databases and reaction networks produced in this work.

## 2. METHODOLOGY

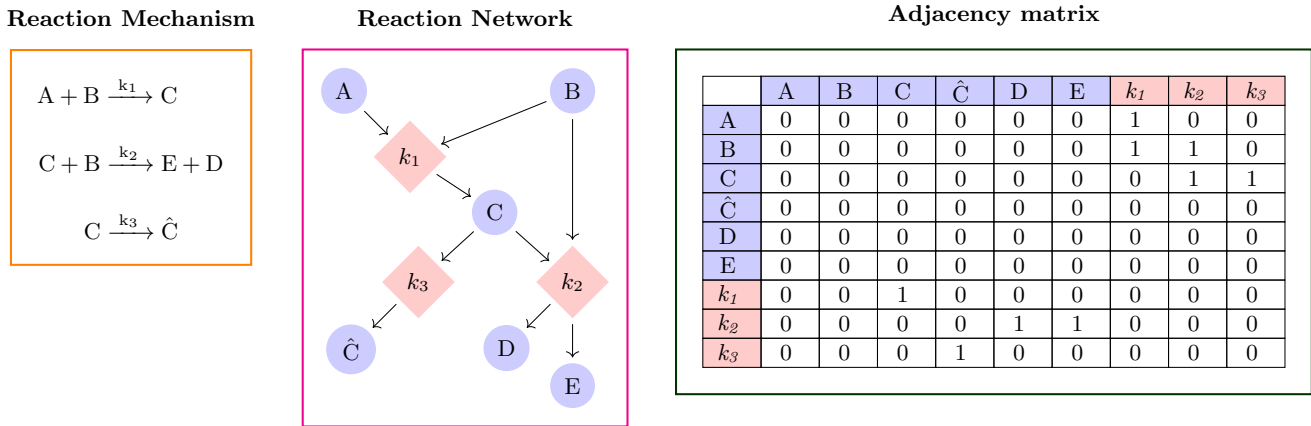
### 2.1. Computational details for QM calculation

All the species present in the astrochemical reaction network were characterized via QM calculations to obtain their optimized structure, absolute electronic energy, and Zero Point Energy (ZPE) correction. Most of

the neutral species structures and their corresponding electronic spin multiplicity ( $m_s$ ) in the ground state, a fundamental piece of information for the electronic QM calculations, are taken from Woon & Herbst (2009) while for the cations, the information is taken from our previous work (Tinacci et al. 2021). For the chemical species that have not been theoretically characterized yet, we inferred here the structure and  $m_s$  using the same approach introduced in Tinacci et al. (2021).

In order to have all species characterized at the same level of theory, *i.e.*, so to have a consistent method, we used the DFT M06-2X (Zhao & Truhlar 2008) coupled with the triple- $\zeta$  Dunning’s correlation consistent basis set (cc-pVTZ) (Kendall Jr 1992; Woon & Dunning Jr 1993) for geometry optimization and harmonic frequency calculations. The electronic absolute energy was refined at a CCSD(T) (Knowles et al. 1993, for closed shell species) or RO-CCSD(T) (Watts et al. 1993, for open shell species) level of theory, in conjunction with an augmented triple- $\zeta$  correlation consistent basis set (aug-cc-pVTZ; Kendall Jr 1992). All the molecular structures and properties computed with the adopted methodology were benchmarked in our previous work (Tinacci et al. 2021). All calculations were carried out with the Gaussian16 program (Frisch et al. 2016) and we kept the default values set up in the program.

Due to the importance to have accurate reaction energies for our reactions network cleaning methodology, we report a benchmark of our resulting ZPE-corrected reaction energy with respect to other published studies in Table 1. All differences are smaller than 4 kJ/mol, which corroborates the accuracy of present results. Please note that the differences are mostly due to a different treat-



**Figure 1.** Three different ways to visualize and work with a reaction network: reactions list (*left panel*), colored directed graph (or colored di-graph), where circles correspond to chemical species, while squares indicate rate constants (*middle panel*), and non symmetric adjacency (or connectivity) matrix representation of the colored di-graph (*right panel*).

ment of the spin contamination, which we improved in the present study.

We decided not to compute the reaction energies using the standard enthalpy of formation ( $\Delta H_f^\circ(T)$ , usually given at 298.15 K or 0 K), but instead we used the absolute electronic energy. This choice minimizes the propagation error in the QM computations, which would be larger if carried out using the formation enthalpy. One way to overcome this limitation would be to compute the  $\Delta H_f^\circ(T)$  by the Isodesmic and Homodesmic cycles (Cramer 2013; Wheeler et al. 2009), which compensates the errors. The problem with this technique is the difficulty of automatize it. In the past, some attempts were carried out but they cannot be generalized to all chemical species (e.g. Cavallotti 2022).

It is worth noticing that many astrochemical species, radicals and exotic species with respect to the terrestrial chemistry, lack laboratory measured  $\Delta H_f^\circ(T)$ . Therefore, limiting the control on the exo/endo thermicity of the reaction to only those having laboratory experimental data would greatly reduce our capacity to "clean" the astrochemical networks.

Overall, we characterized 542 species. The whole list is publicly available at the ACO (Astro-Chemical Origin) project website<sup>7</sup>. In addition, a data frame (in .csv format) with all the properties of the studied species is available in the Supplementary Material of this article.

## 2.2. Graph theory for reaction networks

In the astrochemical community, the reaction network is usually encoded and manipulated with a matrix approach. In this work, we adopt an approach based on

the *graph theory*, so that the network manipulation can be easily obtained via the NETWORKX (Hagberg et al. 2008) Python package (e.g., node connectivity, centrality, sinks, and other characteristics). Instead of using a multi-di-graph approach as done by Grassi et al. (2013), the mathematical object we used to describe the reaction network is the colored directed graph (also known as colored di-graph). Here, each node has a color attribute and the connections are directed (Barabási 2013). A colored node can have different attributes depending on the color: in our case, it can be a species-node or a rate constant-node. In addition, in a directed graph, the edges have a direction, i.e. the adjacency (or connectivity) matrix is not symmetrical, so that the connection between nodes is orientated (e.g. Barabási 2013; Newman 2018). In Figure 1, we report the codification of a generic reaction network given in the above-mentioned three ways: the simple reaction list, the reaction network visualization and, finally, the corresponding adjacency matrix of the colored directed graph. The Python scripts used to encode and characterize the network properties are publicly available in a GitHub repository<sup>8</sup>.

## 2.3. Network cleaning procedure

### 2.3.1. Identification of endothermic reactions

As explained in the Introduction, the first goal of the present work is to produce a network "cleaned" from endothermic reactions that have not been reported as such (see below), since they cannot occur in the cold ISM. To this end, we computed the reaction enthalpy  $\Delta H$  of the neutral-neutral, ion-neutral, and ion-ion reactions

<sup>7</sup> <https://aco-itn.oapd.inaf.it/home>

<sup>8</sup> [https://github.com/TinacciL/GreToBaPe\\_Cleaning](https://github.com/TinacciL/GreToBaPe_Cleaning)

present in the reaction network. Note that we did not consider photo- and electron- induced reactions (involving cosmic-ray particles and UV radiation) because they entail strongly energetic processes, whose products are usually not in their fundamental electronic state.

In the ISM, molecules primarily reside in their ground electronic state and lowest vibrational level, with only the rotational levels populated, depending on the gas conditions. Since the rotational energy has a negligible contribution to the total reaction energy, we used the electronic energy and the vibrational energy at 0 K (also called Zero-point Energy correction, *i.e.*, ZPE) to evaluate the reaction enthalpy  $\Delta H(0)$  of each reaction present in the reaction network. Then, an endothermic reaction has  $\Delta H(0)$  larger than zero.

In addition, in the astrochemical networks the rate constants of the neutral-neutral and ion-neutral (but not ion-polar) reactions are reported following the modified Arrhenius equation (Kooij 1893; Laidler 1996):

$$k(T) = \alpha \left( \frac{T}{300 \text{ K}} \right)^\beta \exp \left( -\frac{\gamma}{T} \right) \quad (1)$$

where  $T$  is the gas temperature in K, and  $\alpha$ ,  $\beta$ , and  $\gamma$  are often derived as fitted parameters of either experimental or theoretical values of  $k(T)$ . It should be emphasized that  $\gamma$  does not always have an actual physical meaning. In other words,  $\gamma$  is often not a real activation barrier. That said, it can be safely assumed that, if the reaction has a barrier, it is lower or equal to  $\gamma$ . Therefore, in order to consider that a reaction has been incorrectly inserted in the reaction network, we apply the additional criterion that  $\Delta H(0) > \gamma R$ , where  $\gamma$  is multiplied by the constant  $R$  to have it in kJ/mol.

Finally, our computations have an intrinsic uncertainty, *i.e.* the accuracy of the electronic structure QM calculations, which we conservatively evaluate to be 10 kJ/mol. Therefore, when retaining or excluding a reaction from the network we consider a 10 kJ/mol threshold regardless the actual meaning of  $\gamma$  either a real activation barrier or the results of an analytical fit.

In summary, a reaction is considered *correctly* included in the network if it satisfies the following criteria:

- 1) if the reaction is not encoded with the modified Arrhenius equation (*i.e.* the ion-pol formula for barrierless reactions),  $\Delta H(0) \leq 10$  kJ/mol;
- 2) if the reaction is encoded with the modified Arrhenius equation,  $\Delta H(0) \leq \gamma R + 10$  kJ/mol.

Reactions that do not satisfy the two above criteria are excluded.

### 2.3.2. The domino effect of removing endothermic reactions

In order for the network to be meaningful, each species must have at least one destruction route *and* one formation route. By destruction route, we mean here that the specified species acts as a reactant, no matter the product(s), in opposition to the meaning given in general chemistry, where the term usually indicates reactions leading to molecules with a smaller number of atoms. Obviously, atomic species are not subject to this constraint.

Removing reactions from the network can affect other species if they are not formed or destroyed through other reactions than those removed. In other words, if a species becomes a source (species with no formation routes) or a sink (species with no destruction routes) it is also removed from the network. As a consequence, the reactions involving the removed species are also removed.

Therefore, the removal of one or more initial reactions triggers a domino effect, causing the removal of other reactions and species. Our graph theory approach, described in § 2.2, allows us to easily follow the domino effect and identify the reactions and species to be removed.

### 2.4. Network reduction procedure

As stated in the Introduction, the second goal of this work is to produce a reduced network from the cleaned one, which only includes H-, C-, N-, O- and S- bearing species with at most six atoms, to be used when the larger network is unnecessary. To this end, we used the following criteria and steps:

1. We only consider C-, N-, O- and S-bearing species, plus Fe and Mg atoms, because their ions can be important positive charge carriers.
2. We only consider molecules with six or less atoms (either reactant or product).
3. The removal of larger molecules may cause the domino effect, explained in § 2.3.2, which can lead to the removal of other reactions and species also with less than six atoms.
4. Methanol is the only exception to rules 2 and 3. In the gas phase, it is formed by reactions involving molecules with more than 6 atoms, so it should be removed, based on rule 3. However, methanol is an important component of the interstellar grain mantles, a major carbon carrier, believed to be mainly formed on the grain surfaces and released into the gas phase by thermal and non-thermal



effects. Astrochemical applications may need to consider this species and, therefore, we keep it in the reduced network, for example used in conjunction with a grain-surface chemical network.

5. We remove all anions. We had two reasons for that. The first one is that, usually, they are not crucial species in studies that necessitate a reduced network. The second one is that there is still a debate on their rate of formation (*e.g.* Lara-Moreno et al. 2019b,a).

### 3. RESULTS

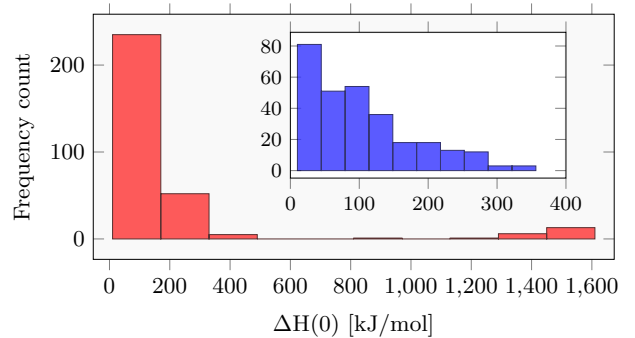
#### 3.1. Original astrochemical network

The reaction network from which we start is based on the KIDA 2014 network (Wakelam et al. 2015), updated following studies from our and other groups, as follows: Hoyermann et al. (1981); DeFrees et al. (1985); Meot-Ner et al. (1986); Dóbé et al. (1991); Anicich (2003); Hamberg, M. et al. (2010); Cheikh (2012); Lawson et al. (2012); Fournier (2014); Loison et al. (2014); Balucani et al. (2015); Neufeld et al. (2015); Vazart et al. (2016); Loison et al. (2016); Urso et al. (2016); Codella et al. (2017); Skouteris et al. (2017); Fontani et al. (2017); Balucani et al. (2018); Gao et al. (2018); Rosi et al. (2018); Skouteris et al. (2018); Sleiman et al. (2018); Ascenzi et al. (2019); Ayouz et al. (2019); Lara-Moreno et al. (2019a); Skouteris et al. (2019); Urso et al. (2019); Balucani (2020); Codella et al. (2020); Vazart et al. (2020); Blázquez et al. (2020); Mancini & Aragão (2021), and data, not yet published, courteously provided by Luca Mancini on some P-bearing reactions. We also add the recombination reactions of  $F^+$  and  $P^+$  with free electrons, which are not present in the KIDA and UMIST astrochemical reaction networks. These data are taken from the database of Badnell (2006) and given in the usual Arrhenius-Kooij format (Eq. 1).

In summary, the pre-cleaned original network comprises 500 species and 7763 reactions.

**Table 2.** List of the species removed from the original network after the cleaning process.

Species		
SiC <sub>4</sub> H	SiC <sub>6</sub> H	SiC <sub>8</sub> H
SiC <sub>2</sub> CH <sub>3</sub>	SiC <sub>3</sub> H <sub>5</sub>	H <sub>2</sub> CSiCH
H <sub>2</sub> C <sub>6</sub> N <sup>+</sup>	H <sub>2</sub> C <sub>8</sub> N <sup>+</sup>	H <sub>2</sub> C <sub>10</sub> N <sup>+</sup>
PNH <sub>3</sub> <sup>+</sup>	C <sub>4</sub> H <sub>5</sub>	



**Figure 2.** Histogram of the enthalpy  $\Delta H(0)$  (red) of the endothermic reactions removed from the original network after the cleaning process. The inset (blue) shows the zoom in the 10–400 kJ/mol interval. The bin width is computed following the Freedman Diaconis estimator (Freedman & Diaconis 1981).

#### 3.2. New cleaned network: GRETOBAPE

##### 3.2.1. Overview of the removed species and reactions

Among the 6275 reactions that have thermodynamic data<sup>9</sup> (§ 3.1) and for which we could carry out the endothermicity verification, 319 do not satisfy the two criteria described in § 2.3, namely  $\sim 5\%$  of the studied reactions. Figure 2 shows the distribution of their endothermic energy. The majority of these reactions have endothermicities lower than about 150 kJ/mol while about 140 reactions even larger.

The full list of the removed reactions is reported in the on-line Supporting Material. In addition, we supply the list of 51 reactions which are endothermic but with a reaction enthalpy lower than 10 kJ/mol and for which more accurate calculations should be carried out.

Finally, because of the domino effect described in § 2.3.2, 11 species and 23 additional reactions are removed, leading to a total of 342 removed reactions. The full list of deleted species is reported in Tab. 2. Table 3 provides an overview of the removed reactions, grouped according to each reaction class. The ion-neutral reactions are the most affected by the cleaning process in number (153), while the cation-anion reactions are the most affected in percentage (7.6%). Table 4 lists the number of species in the final cleaned network classified with their elements.

In summary, the new cleaned network GRETOBAPE contains 7421 reactions and 489 species.

<sup>9</sup> Please note that 272 reactions do not have thermodynamic data, namely they contain species for which the electronic structures are not available.

**Table 3.** Summary of the number of reactions in the original and final GRETOBAPE networks.

	Neutral-Neutral	Ion-Neutral	Cation-Anion	Recombination	Unimolecular	Total
<b>Pre-cleaning</b>	1035	4091	1421	642	574	7763
<b>Endothermic</b>	58	153	108	-	-	319
<b>Domino effect</b>	2	-	-	10	11	23
<b>Post-cleaning</b>	975	3938	1313	632	563	7421

NOTE—The reactions are listed according to their class. Each row presents the number of reactions after each step of the cleaning procedure (§ 2.3). *Raw "Pre-cleaning"*: original network (§ 3.1); *Raw "Endothermic"*: reactions removed because they are endothermic with  $\Delta H(0) > 10$  kJ/mol; *Raw "Domino effect"*: reactions removed because of the domino effect (§ 2.3.2); *Raw "Post-cleaning"*: final cleaned network.

**Table 4.** Summary of the 487 species contained in the cleaned network GRETOBAPE, together with the elements composing them.

Element	Number of species containing n times the element											Total
	1	2	3	4	5	6	7	8	9	10	11	
H	123	81	53	38	21	12	7	1	0	0	0	336
He	3	0	0	0	0	0	0	0	0	0	0	336
C	100	66	57	34	21	19	20	17	19	12	2	367
N	115	8	0	0	0	0	0	0	0	0	0	123
O	93	24	0	0	0	0	0	0	0	0	0	117
F	7	0	0	0	0	0	0	0	0	0	0	7
Na	6	0	0	0	0	0	0	0	0	0	0	336
Mg	3	0	0	0	0	0	0	0	0	0	0	336
Si	53	0	0	0	0	0	0	0	0	0	0	53
S	38	7	0	0	0	0	0	0	0	0	0	45
P	38	0	0	0	0	0	0	0	0	0	0	38
Cl	10	0	0	0	0	0	0	0	0	0	0	10
Fe	2	0	0	0	0	0	0	0	0	0	0	336

### 3.2.2. Description of the removed reactions

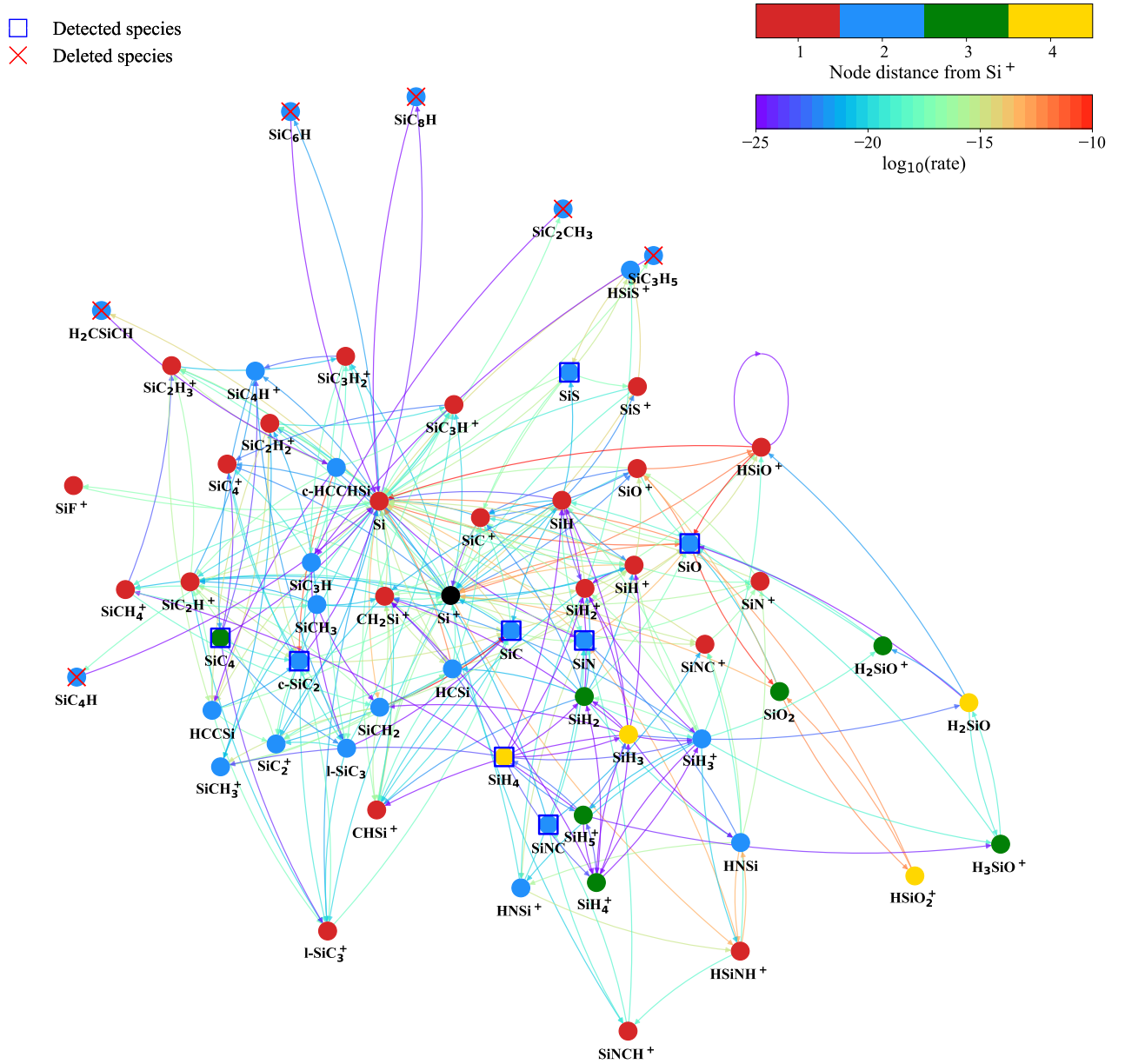
The 342 removed reactions have the following properties:

- 67 reactions are not encoded with the modified Arrhenius equation: they either do not satisfy the criterion 1 listed in § 2.3.1, namely  $\Delta H(0) \leq 10$  kJ/mol, or they are removed because of the domino effect.
- 228 reactions are encoded via the Arrhenius-Kooji equation and have  $\gamma = 0$ : they either do not satisfy the criterion 2 listed in § 2.3.1, namely  $\Delta H(0) \leq 10$  kJ/mol, or they are removed because of the domino effect.
- 47 reactions are encoded via the Arrhenius-Kooji equation and have  $\gamma \neq 0$ : they either do not satisfy

the criterion 2 listed in § 2.3.1, namely  $\Delta H(0) \leq \gamma R + 10$  kJ/mol, or they are removed because of the domino effect.

Overall, the cleaning process strongly affects the Silicon chemistry, as it will be shown by the modeling simulations of § 4.1. Indeed, out of a total number of 342 removed reactions, 51 involve Si-bearing species. In practice, about 15% of the total removed reactions regard Si-bearing species, and about 9% of the original reactions involving Si-bearing species (593) are removed. Likewise, 6 out of 11 removed species are Si-bearing species (see Tab. 2) and represents the  $\sim 10\%$  of the total original Si-bearing species (61).

Figure 3 shows a graph representation of the reactions involving Si-bearing species and present in the original network, together with the removed and detected



**Figure 3.** Original Si-bearing reaction network, visualized as directed colored graph. The reactions are represented as solid lines and the species as nodes (differently from the representation described in § 2.2, where reactions and species are both represented by nodes). The lines are colored based on the  $\log_{10}$  of the reaction rate constants calculated at 90 K and multiplied by the reactants densities at  $5 \times 10^3 \text{ yr}$  in the “warm molecular outflow shock” model described in § 4.1.2, using the original (pre-cleaning) network. All rates larger than  $10^{-10}$  are red and those less than  $10^{-25}$  are violet. The nodes are colored based on the graph distance of that species from  $\text{Si}^+$  (highlighted in black). The species removed from the original network are marked by red crosses, while those detected in the ISM are marked by blue squares. Please note that the digital format of the figure is vectorial and can be zoomed to better look at the details.

species. Figure 4 presents the same reaction network showing the deleted reactions only. First, the 6 removed Si-bearing species (all containing more than 6 atoms)

are only formed and deleted by one reaction, so that their removal did not introduce a domino effect on other species. As can be seen from Figs. 3 and 4, these species





reactions that increase the complexity of the molecules, namely they increase their number of atoms. This is a general characteristic of the cleaning process: the most affected class of reactions are those that increase the number of atoms of at least one reactant rather than those that reduce it. This result can be explained by the fact that the excess of energy of "destructive" reactions is generally much easier to guess than that of formation reactions.

Another source of endothermic reactions comes from the cation-anion reactions which involve carbon chain anions, whose data were predominantly taken from Harada & Herbst (2008).

### 3.2.3. Missing destruction reactions

In cold molecular clouds, with few exceptions, neutral species are mainly destroyed by the most abundant cations, namely  $\text{H}_3^+$ ,  $\text{HCO}^+$ ,  $\text{H}_3\text{O}^+$ ,  $\text{He}^+$  and  $\text{H}^+$ . Therefore, every neutral species should have at least one destruction channel for each of the above cations. Table 5 lists all the neutral species having no destruction channels with the aforementioned cations in the cleaned reaction network. Please note that some reaction channels were removed during the cleaning process. We mark them with a "•" symbol. The Tab. 5 list is meant to identify the reactions that need experiments and/or QM computations. In addition, our database of absolute electronic energies, ZPE and  $m_s$  (Tinacci et al. 2021) can be used to test the possible reaction channels of formation and/or destruction routes.

As an illustrative example, it is worth mentioning the destruction reactions involving SiO. The reaction of SiO with  $\text{H}_3\text{O}^+ \rightarrow \text{H}_2\text{O} + \text{HSiO}^+$  has an enthalpy of +152.1 kJ/mol. A similar endothermicity is found for the reaction  $\text{SiO} + \text{HCO}^+ \rightarrow \text{CO} + \text{HSiO}^+$ . However, if the product of both reactions is not  $\text{HSiO}^+$ , as reported in the KIDA database, but its constitutional isomer  $\text{SiOH}^+$ , the reactions become respectively exothermic of -112.2 and -208.7 kJ/mol. A more detailed study on the reliability of the  $\text{SiOH}^+$  isomer as a product is postponed to a future work by our group focusing on the improvement of the Silicon reaction network (clearly necessary, as highlighted in § 3.2.2).

### 3.2.4. Missing formation reactions

The aim of the present work is to provide a reaction network as reliable as possible. Identifying important missing reactions is, therefore, an aspect of this goal. In addition to the missing destruction reactions described in the previous subsection, here we list obvious missing formation reactions, namely those forming detected species. To this scope, we used the recent McGuire (2022) census of the detected species in the ISM and extragalactic medium. Table 6 reports the detected species which do not have a formation route in our cleaned network GRETOBAPE. Most of these molecules are also missing from commonly used databases (*e.g.*, KIDA). Hopefully, Tab. 6 can be used to motivate future works and strengthen the interdisciplinary collaboration between chemists and astronomers.

### 3.3. Reduced network: GRETOBAPE-RED

Following the criteria described in Section 2.4, the reduced network GRETOBAPE-RED contains 205 chemical species from the 489 existent in the (cleaned) network GRETOBAPE. In addition, the removal of reactions associated with pruned species reduces the size of the network from 7418 reactions to 3145. Compared to GRETOBAPE, GRETOBAPE-RED has less than half the amount of reactions and species, implying a significant decrease in computation time when used in any chemical code. This performance boost is seen in the tests we performed on the reduced network using the KROME package (see next section, § 4), where the computing time of a typical gas-phase run decreases from 142 to 21 s using the complete and the reduced networks, respectively.

Due to the criteria imposed on the creation of this network, the main discrepancy we expect between the complete and the reduced networks concerns the chemistry of medium-sized (between 3-6 atoms) carbon chains. This is due to removing potentially important destruction routes, *i.e.* reactions between these medium-sized carbon chains to form larger ones when pruning out C-bearing species with more than 6 atoms. Nonetheless, a complete assessment of the reliability of this reduced network is discussed in the following section, following detailed modeling of a cold and warm case, respectively.

**Table 5.** List of neutral species with missing reactions with the most abundant cations in the cleaned newtork GRETOBAPE.

Species	H <sup>+</sup>	He <sup>+</sup>	H <sub>3</sub> <sup>+</sup>	H <sub>3</sub> O <sup>+</sup>	HCO <sup>+</sup>	Species	H <sup>+</sup>	He <sup>+</sup>	H <sub>3</sub> <sup>+</sup>	H <sub>3</sub> O <sup>+</sup>	HCO <sup>+</sup>	Species	H <sup>+</sup>	He <sup>+</sup>	H <sub>3</sub> <sup>+</sup>	H <sub>3</sub> O <sup>+</sup>	HCO <sup>+</sup>
NH				*		HPO				*		CH <sub>2</sub> CHCN					*
OH				*		OCS				*		C <sub>6</sub> H					*
HF	*			*	*	SiO <sub>2</sub>	*			*		HC <sub>5</sub> N					*
NaH				*		SO <sub>2</sub>	*			*	*	C <sub>7</sub>					*
C <sub>2</sub>				*		CH <sub>3</sub>				*		C <sub>6</sub> N	*	*			*
MgH				*		H <sub>2</sub> CN	*	*	*	*	*	C <sub>2</sub> H <sub>6</sub>					*
CN					*	SiH <sub>3</sub>				*	*	C <sub>3</sub> H <sub>5</sub>	*	*	*	*	*
CO	*			*	*	HOOH	*	*	*	*	*	CH <sub>3</sub> OCH <sub>2</sub>	*		*		
N <sub>2</sub>	*			*		SiCH <sub>2</sub>				*		CH <sub>2</sub> CHC <sub>2</sub> H	*	*	*	*	*
NO				*	*	H <sub>2</sub> CS				*		CH <sub>3</sub> C <sub>3</sub> N					*
PH				*		C <sub>3</sub> N	*					C <sub>7</sub> H					*
O <sub>2</sub>				*		l-SiC <sub>3</sub>				*		HC <sub>6</sub> N	*		*	*	*
HS				*		C <sub>3</sub> P	*					C <sub>8</sub>					*
HCl				*	*	CH <sub>4</sub>				*		C <sub>7</sub> N	*				*
SiN				*	*	CH <sub>2</sub> NH				*		C <sub>7</sub> O	*	*			*
SiO				●	●	CH <sub>3</sub> O	*			*		CH <sub>3</sub> CHCH <sub>2</sub>	*	*	*	*	*
NS				*		SiH <sub>4</sub>				*		CH <sub>3</sub> C <sub>4</sub> H					*
PO				*		H <sub>2</sub> CCN				*	*	C <sub>8</sub> H					*
CCl	*		*	*	*	NH <sub>2</sub> CN				*		HC <sub>7</sub> N					*
ClO	*		*	*	*	CH <sub>2</sub> PH				●	●	C <sub>9</sub>					*
SO				*		HCOOH				*		C <sub>8</sub> N	*	*			*
SiS				*		c-HCCHSi				*		C <sub>3</sub> H <sub>7</sub>	*	*	*	*	*
S <sub>2</sub>				●		C <sub>5</sub>				*		CH <sub>2</sub> CHCHCH <sub>2</sub>	*	*	*	*	*
H <sub>2</sub> O				*		C <sub>4</sub> N	*	*		*		CH <sub>3</sub> C <sub>5</sub> N					*
HCO				*		SiC <sub>3</sub> H				*		C <sub>9</sub> H					*
HNO				*		SiC <sub>4</sub>				*		HC <sub>8</sub> N	*		*	*	*
PH <sub>2</sub>				*		HCCCHO	*	*				C <sub>10</sub>					*
O <sub>2</sub> H	*	*	*	*	*	c-C <sub>3</sub> H <sub>2</sub> O	*	*				C <sub>9</sub> N	*				*
CCN				*		H <sub>2</sub> CCCO	*	*		●		C <sub>9</sub> O	*	*			*
NaOH				*		C <sub>5</sub> H				*		C <sub>3</sub> H <sub>8</sub>	*	*	*	*	*
CCO				●		HC <sub>4</sub> N	*		*	*	*	CH <sub>3</sub> C <sub>6</sub> H					*
OCN	*		*	*	*	C <sub>6</sub>				*		C <sub>10</sub> H					*
HNSi				●		C <sub>5</sub> N	*			*		HC <sub>9</sub> N					*
CO <sub>2</sub>				*		C <sub>5</sub> O	*	*		*		C <sub>11</sub>	*		*	*	*
N <sub>2</sub> O				*	●	C <sub>2</sub> H <sub>5</sub>				*		C <sub>10</sub> N	*	*		*	
HCS				*	*	CH <sub>3</sub> NH <sub>2</sub>				*		c-C <sub>6</sub> H <sub>6</sub>	*				*
NO <sub>2</sub>	*	*		*	*	CH <sub>2</sub> CCH <sub>2</sub>	*	*	*	*	*	CH <sub>3</sub> C <sub>7</sub> N					*

NOTE—With the “●” and “\*” symbols we identify the neutral species reaction with the corresponding cation that was present in our network but was deleted in cleaning process, and the reaction that was originally absent, respectively.

#### 4. ASTROCHEMICAL IMPLICATIONS

This section illustrates the impact of the new reaction networks on the predicted species abundances,

obtained using the original reaction network versus the new cleaned GRETOBAPE and the reduced GRETOBAPE-RED networks, respectively. To this

**Table 6.** Species with reported observations in the ISM (McGuire 2022) which are not present in the final GRETOBAPE network.

Carbon, nitrogen and oxygen-bearing species						
<i>4-atoms</i>						
HCCO	HONO	HNCN	CNCN	HC <sub>2</sub> N		
<i>5-atoms</i>						
HNCNH	NH <sub>2</sub> OH	H <sub>2</sub> NCO <sup>+</sup>	NCCNH <sup>+</sup>			
<i>6-atoms</i>						
HNCHCN	CH <sub>3</sub> NC	C <sub>5</sub> N <sup>-</sup>				
<i>7-atoms</i>						
c-C <sub>2</sub> H <sub>4</sub> O	c-C <sub>3</sub> HCCH	CH <sub>2</sub> CHOH	HC <sub>4</sub> NC	CH <sub>3</sub> NCO	HOCH <sub>2</sub> CN	HC <sub>5</sub> O
<i>8-atoms</i>						
CH <sub>2</sub> CHCHO	CH <sub>3</sub> CHNH	NH <sub>2</sub> CH <sub>2</sub> CN	NH <sub>2</sub> CONH <sub>2</sub>			
<i>9-atoms</i>						
HC <sub>7</sub> O	H <sub>2</sub> CCCHCCH	HCCCHCHCN	H <sub>2</sub> CCHC <sub>3</sub> N	CH <sub>3</sub> NHCHO	CH <sub>3</sub> CONH <sub>2</sub>	
<i>10-atoms</i>						
HOCH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>3</sub> CH <sub>2</sub> CHO	CH <sub>3</sub> CHCH <sub>2</sub> O	CH <sub>3</sub> OCH <sub>2</sub> OH			
<i>11-atoms</i>						
C <sub>2</sub> H <sub>5</sub> OCHO	CH <sub>3</sub> COOCH <sub>3</sub>	CH <sub>3</sub> COCH <sub>2</sub> OH	NH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> OH			
<i>12-atoms</i>						
n-C <sub>3</sub> H <sub>7</sub> CN	i-C <sub>3</sub> H <sub>7</sub> CN	1-C <sub>5</sub> H <sub>5</sub> CN	2-C <sub>5</sub> H <sub>5</sub> CN			
<i>&gt;13-atoms</i>						
HC <sub>11</sub> N	c-C <sub>6</sub> H <sub>5</sub> CN	C <sub>9</sub> H <sub>8</sub>	1-C <sub>10</sub> H <sub>7</sub> CN	2-C <sub>10</sub> H <sub>7</sub> CN		
Sulfur-bearing species						
HSC	NCS	HNCS	HCCS	CHOSH	HCSCN	HCSCCH
H <sub>2</sub> CCS	H <sub>2</sub> CCCS	CH <sub>3</sub> SH	C <sub>5</sub> S	CH <sub>3</sub> CH <sub>2</sub> SH		
Silicon-bearing species						
SiCN	SiCSi	SiC <sub>3</sub>	SiH <sub>3</sub> CN	CH <sub>3</sub> SiH <sub>3</sub>		

NOTE—Please note that different isomers of the listed species may be present in KIDA and UfA.

end, we run two time-dependent gas-only astrochemical codes: MYNAHOON, a modified version of the publicly available code NAHOON (Wakelam et al. 2012)<sup>10</sup>, and KROME, a publicly available package (Grassi et al. 2014). MYNAHOON is used to evaluate the difference in the species abundances predicted with the original and the new GRETOBAPE reaction network in a typical cold molecular cloud (§ 4.1.1) and a typical warm molecular outflow shock (§ 4.1.2), respectively. KROME is used to test the reliability of the reduced network GRETOBAPE-RED (§ 4.2) in a large parameter space.

#### 4.1. Predicted abundances with the original versus the cleaned reaction network GRETOBAPE

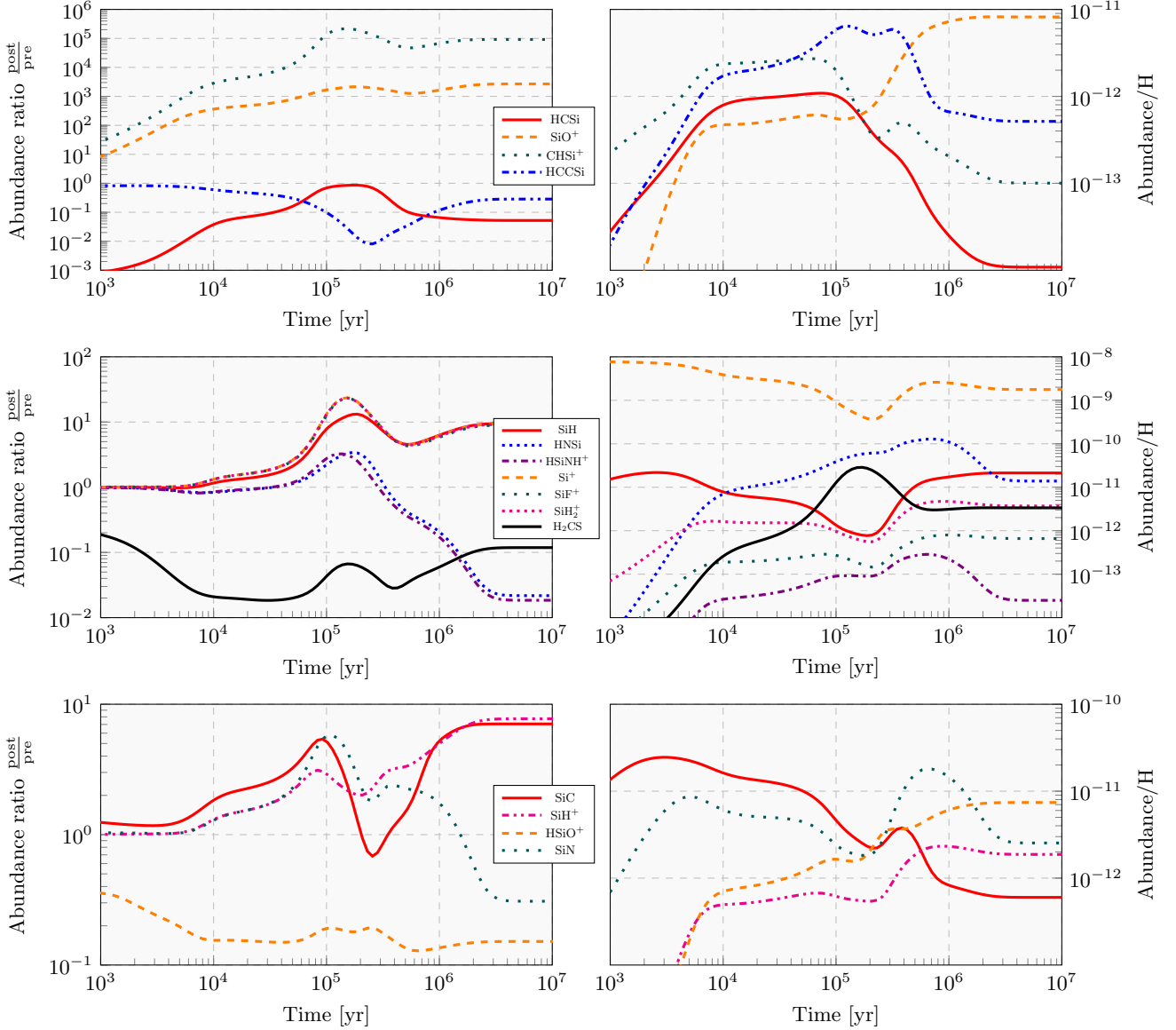
<sup>10</sup> Briefly, the modifications concern a more friendly usage of the code, in input and output, and not the core of the code itself.

In order to assess the difference between the obtained clean reaction network GRETOBAPE and the original one, we performed two simulations: the first one simulates a typical cold molecular cloud (§ 4.1.1) and the second one a typical warm molecular outflow shock (§ 4.1.2).

##### 4.1.1. Cold molecular cloud

We adopted the parameters of a typical molecular cloud, namely: temperature (gas and dust) equal to 10 K, H nuclei number density equal to  $2 \times 10^4 \text{ cm}^{-3}$ , visual extinction  $A_v$  equal to 20 mag, cosmic-ray (CR) ionization rate for H<sub>2</sub>  $\zeta$  equal to  $3 \times 10^{-17} \text{ s}^{-1}$ . The initial elemental abundances are as follows (from Jenkins 2009, with the heavy metals depleted by a factor 100): He/H =  $9.0 \times 10^{-2}$ , C/H =  $2.0 \times 10^{-4}$ , O/H =  $2.6 \times 10^{-4}$ , N/H =  $6.2 \times 10^{-5}$ , S/H =  $8 \times 10^{-8}$ , Si/H =  $8 \times 10^{-9}$ , Fe/H =  $3 \times 10^{-9}$ , Mg/H =  $7 \times 10^{-9}$ . Finally, we started with the conditions of a pseudo-translucent

**Figure 5.** Comparison of the network post- (GRETOBAPE) over pre-cleaning predicted abundance ratios (left panels) and post-cleaning abundance (right panels) as a function of time for the cold molecular cloud model (§ 4.1.1). The three rows show three groups of species where the abundance ratios during the evolution are  $\geq 10^3$  or  $\leq 10^{-3}$  (upper panels),  $10^{-3}$  or  $0.1-10^{-3}$  (middle panels) and  $3-10$  or  $0.1-0.3$  (bottom panels). Only species with abundances larger than  $10^{-13}$  are considered.



cloud, namely with hydrogen in the molecular form and the other elements in atomic (ionic or neutral) form, and we let evolve the chemistry up to  $1 \times 10^7$  years.

Figure 5 shows the ratio of the predicted abundances obtained with the cleaned GRETOBAPE and original reaction networks as a function of time, for the species where this ratio is larger than 3 or smaller than 0.3 at any time  $\geq 10^3$  yr of the simulation and whose abundances are larger than  $10^{-13}$  at steady-state. The figure also shows the abundances of these species, to further

appreciate the impact. Table 7 reports the minimum and maximum ratios in the time intervals  $10^4-10^5$  yr,  $10^5-10^6$  yr and at steady-state, respectively.

Overall, 26 species are significantly affected by the reaction network cleaning process, meaning that their abundances have changed by more than a factor three during the evolution using the two reaction networks. The affected species represent only 5% of the species in the clean network GRETOBAPE, which is a reassuring result.

**Table 7.** Results of the modeling of the cold molecular cloud (§ 4.1.1). For each species, listed in the first column, columns 2-5 list the minimum (**min**) and maximum (**max**) values of the network post- (GRETOBAPE) over pre-cleaning predicted abundance ratios in the time range  $1 \times 10^4$ – $1 \times 10^5$  and  $1 \times 10^5$ – $1 \times 10^6$  yr, respectively. Column 6-7 report the ratio and the abundance at steady-state (SS). The last two columns list the species largest reached abundance and at what time. The species are ordered following the maximum ratio variation during the entire simulation. Finally, boldfaced species have been detected in the ISM.

Interval	$1 \times 10^4$ – $1 \times 10^5$ yr		$1 \times 10^5$ – $1 \times 10^6$ yr		SS		max(Abd/H)	
	min(ratio)	max(ratio)	min(ratio)	max(ratio)	ratio	Abd/H	Abd/H	time[yr]
HCSi	$4.2 \times 10^{-2}$	$6.2 \times 10^{-1}$	$6.7 \times 10^{-2}$	$8.6 \times 10^{-1}$	$5.2 \times 10^{-2}$	$1.1 \times 10^{-14}$	$1.1 \times 10^{-12}$	$7.5 \times 10^4$
SiO <sup>+</sup>	$3.8 \times 10^2$	$1.6 \times 10^3$	$1.3 \times 10^3$	$2.1 \times 10^3$	$2.7 \times 10^3$	$8.2 \times 10^{-12}$	$8.2 \times 10^{-12}$	$2.4 \times 10^6$
CHSi <sup>+</sup>	$3.1 \times 10^3$	$1.0 \times 10^5$	$4.7 \times 10^4$	$2.1 \times 10^5$	$9.2 \times 10^4$	$1.0 \times 10^{-13}$	$2.7 \times 10^{-12}$	$6.1 \times 10^4$
HCCSi	$1.2 \times 10^{-1}$	$5.8 \times 10^{-1}$	$8.1 \times 10^{-3}$	$1.1 \times 10^{-1}$	$2.9 \times 10^{-1}$	$5.2 \times 10^{-13}$	$6.5 \times 10^{-12}$	$1.3 \times 10^5$
SiH	1.2	6.7	4.5	$1.3 \times 10^1$	9.5	$2.1 \times 10^{-11}$	$2.2 \times 10^{-11}$	$2.6 \times 10^3$
HNSi	$8.4 \times 10^{-1}$	2.0	$2.2 \times 10^{-1}$	3.4	$2.2 \times 10^{-2}$	$1.4 \times 10^{-11}$	$1.3 \times 10^{-10}$	$7.1 \times 10^5$
HSiNH <sup>+</sup>	$8.6 \times 10^{-1}$	2.5	$1.8 \times 10^{-1}$	3.2	$1.8 \times 10^{-2}$	$2.5 \times 10^{-14}$	$2.9 \times 10^{-13}$	$6.4 \times 10^5$
<b>Si<sup>+</sup></b>	1.4	$1.0 \times 10^1$	4.5	$2.4 \times 10^1$	9.4	$1.8 \times 10^{-9}$	$7.8 \times 10^{-09}$	$1.0 \times 10^2$
SiF <sup>+</sup>	1.4	$1.0 \times 10^1$	4.3	$2.3 \times 10^1$	8.9	$6.6 \times 10^{-13}$	$8.0 \times 10^{-13}$	$9.6 \times 10^5$
SiH <sub>2</sub> <sup>+</sup>	1.4	$1.0 \times 10^1$	4.4	$2.3 \times 10^1$	9.2	$3.7 \times 10^{-12}$	$4.7 \times 10^{-12}$	$8.7 \times 10^5$
<b>H<sub>2</sub>CS</b>	$1.8 \times 10^{-2}$	$3.7 \times 10^{-2}$	$2.8 \times 10^{-2}$	$6.7 \times 10^{-2}$	$1.2 \times 10^{-1}$	$3.4 \times 10^{-12}$	$2.9 \times 10^{-11}$	$1.7 \times 10^5$
<b>SiC</b>	1.9	5.4	$6.8 \times 10^{-1}$	5.1	7.1	$6.0 \times 10^{-13}$	$2.5 \times 10^{-11}$	$2.9 \times 10^3$
SiH <sup>+</sup>	1.3	3.1	2.0	4.9	7.7	$1.9 \times 10^{-12}$	$2.3 \times 10^{-12}$	$8.7 \times 10^5$
HSiO <sup>+</sup>	$1.5 \times 10^{-1}$	$1.9 \times 10^{-1}$	$1.3 \times 10^{-1}$	$1.9 \times 10^{-1}$	$1.5 \times 10^{-1}$	$7.4 \times 10^{-12}$	$7.4 \times 10^{-12}$	$1.0 \times 10^7$
<b>SiN</b>	1.3	5.4	1.8	5.8	3.1	$2.5 \times 10^{-12}$	$1.8 \times 10^{-11}$	$6.4 \times 10^5$
<b>SiNC</b>	$8.8 \times 10^{-1}$	2.8	1.6	5.8	3.4	$2.5 \times 10^{-15}$	$1.2 \times 10^{-13}$	$6.8 \times 10^4$
l-SiC <sub>3</sub>	1.1	5.5	3.6	7.7	6.5	$1.0 \times 10^{-14}$	$9.0 \times 10^{-12}$	$3.5 \times 10^5$
SiC <sub>3</sub> H	$5.8 \times 10^{-1}$	$8.8 \times 10^{-1}$	$5.8 \times 10^{-1}$	1.1	4.5	$1.8 \times 10^{-15}$	$1.5 \times 10^{-11}$	$2.1 \times 10^5$
SiCH <sub>3</sub>	$8.8 \times 10^{-1}$	$9.6 \times 10^{-1}$	$9.9 \times 10^{-1}$	3.3	6.4	$5.9 \times 10^{-14}$	$7.6 \times 10^{-13}$	$1.5 \times 10^5$
l-C <sub>3</sub> H <sub>2</sub> <sup>+</sup>	$5.3 \times 10^{-1}$	$9.0 \times 10^{-1}$	$1.0 \times 10^{-1}$	$4.7 \times 10^{-1}$	$2.7 \times 10^{-1}$	$2.5 \times 10^{-15}$	$3.7 \times 10^{-11}$	$1.0 \times 10^5$

NOTE—The detected molecules are bold-faced (McGuire 2022). The Si<sup>+</sup> atom was first detected by Haas et al. (1986).

The species impacted by the reaction network cleaning process can be divided into three groups, where we only consider those whose abundance is larger than  $10^{-13}$  during the evolution:

1. The first group (upper panels of Fig. 5) contains species where the post- over pre- cleaning ratio is larger than  $10^3$  or lower than  $10^{-3}$  during the evolution: HCSi, SiO<sup>+</sup>, CHSi<sup>+</sup> and HCCSi. None of these species have been detected in space. Note that five of these six species contain Si and all of them have an abundance lower than  $2 \times 10^{-11}$ . The most abundant one is SiO<sup>+</sup>, with a predicted steady-state abundance of  $8 \times 10^{-12}$ ,  $3 \times 10^3$  larger with respect to the original reaction network. The species whose ratio is the largest at steady-state is CHSi<sup>+</sup>, when its abundances is lower than  $10^{-13}$ . HCCSi presents a ratio of 0.3 and an abundance of  $5 \times 10^{-13}$  at steady-state, while at  $\sim 2 \times 10^5$  yr the

minimum ratio value is 0.008. We will discuss in § 4.1.3 the Si chemistry and why it is so impacted by the reaction network cleaning process.

2. The second group (middle panels of Fig. 5) contains species where the ratio is within 10 and  $10^3$  or  $10^{-3}$  and 0.1 during the evolution: SiH, HNSi, HSiNH<sup>+</sup>, Si<sup>+</sup>, SiF<sup>+</sup>, SiH<sub>2</sub><sup>+</sup> and H<sub>2</sub>CS. H<sub>2</sub>CS and Si<sup>+</sup> are the only species detected in space belonging to this group. H<sub>2</sub>CS has a ratio ranging from 0.1 to 0.03 during the evolution, namely its abundance is lower in the cleaned network simulation by about a factor 10 at steady-state ( $3.4 \times 10^{-12}$  instead of  $4 \times 10^{-13}$ ). The cleaning process affects identically, with the same ratio profile and increasing the abundances by one order of magnitude at steady-state, the following Si-bearing species: SiH, Si<sup>+</sup>, SiF<sup>+</sup> and SiH<sub>2</sub><sup>+</sup>. HNSi and HSiNH<sup>+</sup> share the same behavior, but with a steady-state reduced

abundance by about a factor 100. The predicted  $\text{SO}_2^+$  abundance is a factor  $\sim 10$  larger with the new network.

3. The third group (bottom panels of Fig. 5) contains species where the ratio is within 3 and 10 or 0.1 and 0.3: SiC, SiH<sup>+</sup>, HSiO<sup>+</sup> and SiN. Three species of this group have been detected in the ISM: SiC, SiN and SiNC. At steady-state, the abundance of SiC is about a factor 10 larger with the new reaction network, while that of SiN is a factor  $\sim 3$  larger.

Overall, 5 out of 26 species affected by the cleaning process have been detected in the ISM: Si<sup>+</sup>, SiC, SiN, SiNC and H<sub>2</sub>CS. Of those, only thioformaldehyde (H<sub>2</sub>CS) has been detected so far in cold molecular clouds (Irvine et al. 1989; Minowa et al. 1997; Marcelino et al. 2005). We will discuss the case of H<sub>2</sub>CS in Sec. 4.1.3. Finally, 18 out of 20 species affected by the cleaning process contain Si. We will discuss in some detail this case in Sec. 4.1.3.

#### 4.1.2. Warm molecular outflow shock

Young protostars are known to have spectacular ejections of matter which cause shocks when they hit the surrounding quiet environment. From the chemical point of view, the passage of the shock has two major effects: (i) it heats and compresses the gas and (ii) species frozen on the icy grain mantles are sputtered and injected into the gas phase. As a result, molecular shocks are warm ( $\sim 80$ – $100$  K) and dense ( $\geq 10^5 \text{ cm}^{-3}$ ) regions with much larger abundances of some species with respect to cold molecular clouds.

In order to simulate the chemical composition of the gas after the passage of a shock, we adopt a two-step procedure where we first compute the chemical abundances of the cold molecular cloud (previous subsection) and then we suddenly (i) increase the gas temperature and the H nuclei number density and (ii) inject into the gas large abundances of species that are known to be present on the grain icy mantles. Although the physics is slightly different, a similar modeling can be also considered a fair enough simulation of what happens in hot cores and hot corinos, which are regions heated by the central forming star and where the grain icy mantles sublimate (e.g. see the recent review by Ceccarelli et al. 2022).

In practice, our model has two steps, as follows.

Step 1: We first compute the steady-state chemical composition of a cold molecular cloud at 10 K with a H nuclei number density of  $2 \times 10^4 \text{ cm}^{-3}$  and with the parameters described in Sec. 4.1.1.

Step 2: We increase the gas temperature to 90 K, H nuclei number density to  $8 \times 10^5 \text{ cm}^{-3}$  and the gaseous abundance of grain-mantle species as follows:  $\text{CO}_2/\text{H} = 3 \times 10^{-5}$ ,  $\text{H}_2\text{O} = 2 \times 10^{-4}$ ,  $\text{NH}_3 = 2 \times 10^{-5}$ ,  $\text{SiO} = 1 \times 10^{-6}$ ,  $\text{Si} = 1 \times 10^{-6}$  and  $\text{OCS} = 2 \times 10^{-6}$ . We then leave the chemical composition evolve for  $10^4$  yr.

The values adopted for the physical parameters and abundances of the injected species are based on the astronomical observations of the prototypical molecular outflow shock L1155-B1, for which our group carried the same model described here (e.g. Codella et al. 2017; Podio et al. 2017; Codella et al. 2020).

The results of the modeling are reported in Tab. 8, which reports the ratio of the predicted abundances obtained with the cleaned GRETOBAPE and original reaction networks in the time intervals  $1 \times 10^2$ – $1 \times 10^3$  yr and  $1 \times 10^3$ – $5 \times 10^3$  yr for species with a ratio larger than 3 or less than 0.3 and abundances larger than  $10^{-13}$ .

Overall, 22 species are significantly affected by the reaction network cleaning process, namely their abundances change by more than a factor 3 during the evolution from the passage of the shock and  $5 \times 10^3$  yr. The species impacted by the reaction network cleaning process can be divided into three groups with respect to their ratio at  $5 \times 10^3$  yr (e.g. the approximate L1155-B1 age):

1. The first group contains species whose ratio is larger than  $10^3$  or lower  $10^{-3}$ : SiS<sup>+</sup>, CHSi<sup>+</sup>, SiO<sup>+</sup> and SiN. SiN is the only species in this group that has been detected in outflows and will be discussed in detail in a dedicated paragraph of § 4.1.3. The species with the largest variation due to the cleaning process is SiS<sup>+</sup>, with an almost constant ratio of  $4.2 \times 10^7$  and reaching an abundance of  $\sim 4 \times 10^{-11}$  at  $5 \times 10^3$  yr.
2. The second group contains species where the ratio is within  $10^2$  and  $10^3$  or  $10^{-3}$  and  $10^{-2}$ : HSiO<sup>+</sup>, CH<sub>3</sub>CHCH<sub>2</sub>, SiH<sub>2</sub><sup>+</sup>, Si<sup>+</sup>, SiH, HNSi, SiNC<sup>+</sup>. Two species of this group have been detected in the ISM: CH<sub>3</sub>CHCH<sub>2</sub> and Si<sup>+</sup>. Only the latter has been detected in outflows and we discuss its case in § 4.1.3. CH<sub>3</sub>CHCH<sub>2</sub> has an almost constant ratio of  $\sim 5 \times 10^{-2}$  and a maximum abundance of  $\sim 5 \times 10^{-13}$  at the beginning of the simulation. On the other hand Si<sup>+</sup> has a ratio of almost  $4.5 \times 10^2$  that increases the abundance in the cleaned simulation with respect to the original.
3. The third group contains species where the ratio is within 3 and  $10^2$  or  $10^{-2}$  and 0.3 during the evolution: HSS, c-HCCHSi, c-SiC<sub>2</sub>, SiS, SO<sub>2</sub><sup>+</sup>, HCCSi,



**Table 8.** Results of the modeling of the warm molecular outflow shock (§ 4.1.2). For each species, listed in the first column, columns 2-5 list the minimum (**min**) and maximum (**max**) values of the network post- (GRETOBAPE) over pre-cleaning predicted abundance ratios in the time range  $1 \times 10^2$ – $1 \times 10^3$  and  $1 \times 10^3$ – $5 \times 10^3$  yr, respectively. Column 6-7 report the ratio and the abundance at  $5 \times 10^3$  yr. The last two column list the species largest reached abundance and at what time. The species are ordered following the value of the ratio at  $5 \times 10^3$  yr, which is believed to approximately be the L1155-B1 age (§ 4.1.2). Finally, boldfaced species have been detected in the ISM.

Interval	$1 \times 10^2$ – $1 \times 10^3$ yr		$1 \times 10^3$ – $5 \times 10^3$ yr		$5 \times 10^3$ yr		max(Abd/H)	
Species	min(ratio)	max(ratio)	min(ratio)	max(ratio)	ratio	Abd/H	Abd/H	time[yr]
SiS <sup>+</sup>	$3.9 \times 10^7$	$4.2 \times 10^7$	$4.2 \times 10^7$	$4.5 \times 10^7$	$4.2 \times 10^7$	$3.9 \times 10^{-11}$	$6.3 \times 10^{-11}$	$1.0 \times 10^2$
CHSi <sup>+</sup>	$1.5 \times 10^6$	$1.6 \times 10^6$	$1.4 \times 10^6$	$1.9 \times 10^6$	$1.4 \times 10^6$	$4.3 \times 10^{-13}$	$4.3 \times 10^{-13}$	$4.9 \times 10^3$
SiO <sup>+</sup>	$1.3 \times 10^3$	$1.5 \times 10^3$	$1.6 \times 10^3$	$7.9 \times 10^3$	$7.9 \times 10^3$	$1.1 \times 10^{-11}$	$1.1 \times 10^{-11}$	$5 \times 10^3$
<b>SiN</b>	$1.2 \times 10^{-2}$	$1.9 \times 10^{-2}$	$2.8 \times 10^{-3}$	$7.7 \times 10^{-3}$	$5.5 \times 10^{-3}$	$3.6 \times 10^{-13}$	$3.6 \times 10^{-13}$	$5 \times 10^3$
HSiO <sup>+</sup>	$1.1 \times 10^{-1}$	$1.1 \times 10^{-1}$	$8.3 \times 10^{-2}$	$1.0 \times 10^{-1}$	$8.2 \times 10^{-2}$	$3.5 \times 10^{-10}$	$3.5 \times 10^{-10}$	$3.9 \times 10^3$
<b>CH<sub>3</sub>CHCH<sub>2</sub></b>	$5.0 \times 10^{-2}$	$5.2 \times 10^{-2}$	$5.3 \times 10^{-2}$	$6.5 \times 10^{-2}$	$6.5 \times 10^{-2}$	$3.6 \times 10^{-13}$	$4.2 \times 10^{-13}$	$1.0 \times 10^2$
SiH <sub>2</sub> <sup>+</sup>	$4.0 \times 10^2$	$4.3 \times 10^2$	$4.4 \times 10^2$	$5.0 \times 10^2$	$4.9 \times 10^2$	$1.5 \times 10^{-12}$	$1.5 \times 10^{-12}$	$5 \times 10^3$
<b>Si<sup>+</sup></b>	$4.0 \times 10^2$	$4.3 \times 10^2$	$4.4 \times 10^2$	$4.9 \times 10^2$	$4.8 \times 10^2$	$1.5 \times 10^{-10}$	$1.6 \times 10^{-10}$	$3.9 \times 10^3$
SiH	$7.1 \times 10^1$	$1.4 \times 10^2$	$1.6 \times 10^2$	$3.7 \times 10^2$	$3.7 \times 10^2$	$3.9 \times 10^{-12}$	$3.1 \times 10^{-12}$	$4.3 \times 10^3$
HNSi	$1.6 \times 10^{-2}$	$1.9 \times 10^{-2}$	$1.1 \times 10^{-2}$	$1.5 \times 10^{-2}$	$1.1 \times 10^{-2}$	$1.3 \times 10^{-11}$	$1.4 \times 10^{-11}$	$8.0 \times 10^2$
SiNC <sup>+</sup>	$5.5 \times 10^1$	$7.7 \times 10^1$	$8.4 \times 10^1$	$1.2 \times 10^2$	$1.1 \times 10^2$	$2.5 \times 10^{-12}$	$2.4 \times 10^{-12}$	$4.7 \times 10^3$
<b>HSS</b>	$9.6 \times 10^{-1}$	1.0	$2.3 \times 10^{-1}$	$9.2 \times 10^{-1}$	$2.3 \times 10^{-1}$	$4.7 \times 10^{-13}$	$4.7 \times 10^{-13}$	$5 \times 10^3$
c-HCCHSi	$4.3 \times 10^{-1}$	1.2	$2.3 \times 10^{-1}$	$3.7 \times 10^{-1}$	$2.3 \times 10^{-1}$	$6.2 \times 10^{-12}$	$6.1 \times 10^{-12}$	$4.9 \times 10^3$
<b>c-SiC<sub>2</sub></b>	$4.3 \times 10^{-2}$	$5.5 \times 10^{-2}$	$6.3 \times 10^{-2}$	$2.1 \times 10^{-1}$	$2.1 \times 10^{-1}$	$7.2 \times 10^{-13}$	$7.3 \times 10^{-13}$	$4.7 \times 10^3$
<b>SiS</b>	$1.7 \times 10^{-1}$	$2.8 \times 10^{-1}$	$1.5 \times 10^{-1}$	$1.7 \times 10^{-1}$	$1.5 \times 10^{-1}$	$8.4 \times 10^{-12}$	$8.5 \times 10^{-12}$	$4.9 \times 10^3$
SO <sub>2</sub> <sup>+</sup>	$1.3 \times 10^1$	$1.4 \times 10^1$	$1.4 \times 10^1$	$1.5 \times 10^1$	$1.5 \times 10^1$	$1.4 \times 10^{-13}$	$1.4 \times 10^{-13}$	$5 \times 10^3$
HCCSi	$2.4 \times 10^{-2}$	$3.8 \times 10^{-2}$	$4.4 \times 10^{-2}$	$1.3 \times 10^{-1}$	$1.3 \times 10^{-1}$	$4.8 \times 10^{-12}$	$4.7 \times 10^{-12}$	$4.9 \times 10^3$
Si	$1.6 \times 10^{-1}$	$1.7 \times 10^{-1}$	$1.2 \times 10^{-1}$	$1.6 \times 10^{-1}$	$1.2 \times 10^{-1}$	$3.3 \times 10^{-10}$	$2.1 \times 10^{-09}$	$1.0 \times 10^2$
HSSH	$8.0 \times 10^{-1}$	1.0	$1.0 \times 10^{-1}$	$6.7 \times 10^{-1}$	$1.0 \times 10^{-1}$	$1.6 \times 10^{-13}$	$1.6 \times 10^{-13}$	$5 \times 10^3$
SiH <sup>+</sup>	2.4	3.3	3.7	5.7	5.1	$5.3 \times 10^{-13}$	$5.4 \times 10^{-13}$	$4.9 \times 10^3$
<b>SiC</b>	1.2	1.2	1.2	3.0	3.0	$2.3 \times 10^{-11}$	$2.3 \times 10^{-11}$	$5 \times 10^3$
<b>C<sub>3</sub>O</b>	9.8	$2.1 \times 10^1$	2.0	$1.6 \times 10^1$	2.0	$2.4 \times 10^{-12}$	$7.9 \times 10^{-12}$	$1.0 \times 10^2$

NOTE—The detected species are bold-faced (McGuire 2022). The Si<sup>+</sup> atom was first detected by Haas et al. (1986).

Si, HSSH, SiH<sup>+</sup>, SiC and C<sub>3</sub>O. Five species belonging to this group have been detected in the ISM: HSS, c-SiC<sub>2</sub>, SiS, SiC and C<sub>3</sub>O. Only SiS has been detected in outflows and we discuss its case in § 4.1.3. HSS has an abundance ratio between 0.2 and 0.9 in the  $1 \times 10^3$ – $5 \times 10^3$  yr interval and a maximum abundance of  $4.7 \times 10^{-13}$  at the end of the simulation. c-SiC<sub>2</sub> at the L1155-B1 age has a ratio of  $2 \times 10^{-1}$  and an abundance of  $7.2 \times 10^{-13}$ . SiS has an almost constant ratio of 0.15 meanwhile C<sub>3</sub>O ratio oscillates between  $\sim 20$  and 2. SiC is the less affected species by the cleaning process with a factor 3 of difference.

As in the cold molecular cloud simulation (§ 4.1.1), the cleaning process mostly affected the Si reaction network. Indeed, 17 of the 22 species present in Tab. 8 are Si-

bearing species. Four of them, SiN (Schilke et al. 2003), SiS (Podio et al. 2017) and Si<sup>+</sup> (Haas et al. 1986) are detected in outflows. They will be discussed in some detail in § 4.1.3.

#### 4.1.3. Discussion of specific cases

In this section, we discuss in some details the species that are more significantly affected by the reaction network cleaning process.

*Thioformaldehyde*—In cold molecular gas (§ 4.1.1), the thioformaldehyde (H<sub>2</sub>CS) abundance predicted after the cleaning process at the steady-state is about 10 times smaller than in pre-cleaning. Here we first summarize the observations and then we comment on what causes the difference in the predicted abundances.

Thioformaldehyde has been detected in cold molecular clouds, notably TMC-1 and L134, with an abundance



(wrt to H nuclei) of about  $\sim 10^{-9}$  (e.g. Ohishi & Kaifu 1998). Vastel et al. (2018) and Spezzano et al. (2022) carried out a detailed study of thioformaldehyde toward the prototypical cold prestellar core L1544, reporting a column density of  $\sim 6 \times 10^{12} \text{ cm}^{-2}$ , corresponding to an average abundance of  $\leq 10^{-10}$ , depending on where the detected  $\text{H}_2\text{CS}$  emission comes from. These authors attribute the relatively low abundance of all the S-bearing species observed in these objects to a general depletion of the gaseous S elemental abundance. Likewise, Esplugues et al. (2022) have carried out a detailed study of  $\text{H}_2\text{CS}$  towards several cold starless cores in the Taurus, Perseus, and Orion regions and found  $\text{H}_2\text{CS}$  abundances ranging from 0.8 to  $14 \times 10^{-11}$ .

As discussed in detail in Esplugues et al. (2022), in cold gas,  $\text{H}_2\text{CS}$  is mainly formed by two reactions:  $\text{S} + \text{CH}_3$  and  $\text{H}_3\text{CS}^+ + \text{e}^-$ . While  $\text{H}_2\text{CS}$  is not directly involved in the cleaning process (Tab. 2), the  $\text{H}_3\text{CS}^+$  post-cleaning abundance at steady-state is lower by a factor  $\sim 18$  than the pre-cleaned one because one dominant reaction in the original network forming  $\text{H}_3\text{CS}^+$ ,  $\text{CH}_4 + \text{S}^+ \rightarrow \text{H} + \text{H}_3\text{CS}^+$ , is endothermic by 48 kJ/mol. Please note that the steady-state post-cleaning abundance of  $\text{H}_3\text{CS}^+$  is  $\leq 10^{-13}$  so that this species does not appear in Fig. 5 and Tab. 7. However, the low abundance ( $6 \times 10^{-15}$ ) is counterbalanced by the large recombination rate of  $\text{H}_3\text{CS}^+$ , whose product is  $\text{H}_2\text{CS}$ .

*Si chemistry and  $\text{Si}^+$* —As already anticipated, the reaction network cleaning process significantly affects the Si chemistry, with the removal of 51 reactions and 6 species that represent the  $\sim 9\%$  and  $\sim 10\%$  of total Si-bearing reactions and species, respectively. In addition, 18 out of 20 most affected species of the cold molecular cloud simulation (§ 4.1.1) contain Si, of which the vast majority (12) have less than four atoms. Similarly, 17 of the 22 most affected species in the the warm molecular outflow shock simulation (§ 4.1.2) are Si-bearing species.

The six removed Si-bearing species (Tab. 2) are (neutral) chains and are removed because their formation or destruction routes in the original network are endothermic (§ 3.2.2). Their removal very weakly affects two other Si-bearing species, c-HCCHSi and  $\text{SC}_3\text{H}$ , respectively (Figs. 3 and 4) and, consequently, the rest of the Si-bearing species.

Therefore, the large impact of the cleaning process on the Si chemical network is due to the removal of the 51 reactions involving Si-bearing species. Figure 4 shows that those reactions connect 36 Si-bearing species out of 61, namely more than half of them. It is, therefore, not surprising that 17 Si-bearing species are among the most affected ones in the warm molecular outflow shock model. For example,  $\text{SiS}^+$ , the most affected species, has

three removed reactions connecting it to  $\text{Si}^+$ ,  $\text{SiC}_3\text{H}_5$  and  $\text{SiS}$ , respectively. The first two are destruction routes of  $\text{SiS}^+$ ; their removal is, therefore, responsible of the augmented predicted  $\text{SiS}^+$  abundance. A similar argument applies to  $\text{CHSi}^+$ , the second most affected species in the modeling: a reaction of destruction is removed, causing the increase in the predicted abundance.

Probably most important,  $\text{Si}^+$  has 10 removed reactions linking it to other Si-bearing species. Because of that, the predicted abundance of  $\text{Si}^+$  is larger more than 400 times the predicted abundance before the cleaning process. This causes a domino effect, propagating as a wave: 11 of the other 16 impacted Si-bearing species are one step away from  $\text{Si}^+$  (*i.e.* 1 node distance in Fig. 3), while the other 5 are two steps away. The species one step away are also those whose predicted abundances are the most impacted.

*Silicon nitride*—Silicon nitride,  $\text{SiN}$ , is the detected Si-bearing species most affected by the cleaning process (Tab. 8), so we will analyze its case in some detail here.  $\text{SiN}$  was first detected in the circumstellar envelope of the evolved star IRC+10216 (Turner 1992) and, later towards the Galactic Center cloud SgrB2(M) (Schilke et al. 2003), probably in the warm shocked gas of the region. IRC+10216 and SgrB2(M) remain the only two objects where  $\text{SiN}$  has so far been detected, despite sensitive spectral surveys towards the high-mass star forming region Orion (Tercero et al. 2011) and several low-mass star forming regions, including the prototype of the warm molecular shocks L1157-B1 (Ceccarelli et al. 2017).

The predicted  $\text{SiN}$  abundance is more than 100 times lower when using the post-cleaning reaction network than the pre-cleaning one. Carefully looking at Figs. 3 and 4, one can notice that the  $\text{SiN}$  major formation route in the pre-cleaning network is the reaction involving  $\text{HSiNH}^+$ , followed by those involving  $\text{SiC}$  and  $\text{HNSi}^+$  at a lesser extent. Although this reaction is not affected by the cleaning process, two reactions forming  $\text{HSiNH}^+$  are removed after it (from  $\text{Si}^+$  and  $\text{HNSi}$ , respectively). Indeed, the abundance of  $\text{HSiNH}^+$  changes by three orders of magnitude (but it is not listed in Tab. 8 because of its low abundance,  $\leq 10^{-13}$ ). About these reactions, as discussed for the  $\text{HSiO}^+$  case (see § 3.2.3), it is probable that KIDA erroneously considered the product  $\text{HSiNH}^+$  instead of its more stable isomer  $\text{SiNH}_2^+$  (Parisel et al. 1996; Glosik et al. 1995).

As said, the new reaction network predicts a much lower  $\text{SiN}$  abundance than before, which may explain the paucity of the  $\text{SiN}$  detections. For example, Tercero et al. (2011) found a lower limit of  $[\text{SiO}]/[\text{SiN}] \geq 121$ . In our post-cleaning modeling, the  $\text{SiO}$  abun-

dance at  $5 \times 10^3$  yr is  $\sim 2 \times 10^{-6}$ , implying a predicted  $[\text{SiO}]/[\text{SiN}] \sim 5 \times 10^{-6}$ , therefore largely compatible with the non detection of [Tercero et al. \(2011\)](#).

*SiS*—Silicon monosulfide, SiS, is another detected Si-bearing molecules affected by the cleaning process. Specifically, SiS has been so far detected towards two warm molecular shocks, Orion KL ([Tercero et al. 2011](#)) and L1157-B1 ([Podio et al. 2017](#)). The factor between the pre- and post- cleaning predicted abundances is  $\sim 0.15$ , again probably explaining the paucity of SiS detections. Peering into Fig. 3 and 4, in the pre- and post-cleaning networks, SiS is mainly formed from  $\text{HSiS}^+$ , whose abundance ( $\leq 10^{-13}$ , so the species not present in Tab. 8) is affected by the cleaning process, with a decrease factor of  $\sim 0.06$ .

#### 4.2. Predicted abundances with the reaction network GRETOBAPE-RED versus GRETOBAPE

To explore the reliability of our reduced network, GRETOBAPE-RED, we performed an exhaustive parameter-space exploration and compared the steady-state abundances of all species with those obtained using the complete network, GRETOBAPE.

To this end, we initialized different combinations of parameters that can affect the resulting species abundances in the cold and warm conditions present during the star-formation process. Namely, the varied parameters are: the CR ionization rate for  $\text{H}_2$ , with values of  $3 \times 10^{-16}$ ,  $3 \times 10^{-17}$ , and  $3 \times 10^{-18} \text{ s}^{-1}$ ; the gas temperature, with values of 10, 50, 100, and 500 K; and the H nuclei number density, with values of  $10^2$ ,  $10^4$ ,  $10^6$ , and  $10^8 \text{ cm}^{-3}$ . All possible combinations of these parameters yield 48 different runs, the results of which we will explain in what follows. We used the KROME chemistry package ([Grassi et al. 2014](#)) to perform the parameter space exploration, considering only gas-phase chemistry. We keep the rest of the input parameters (visual extinction, dust-to-gas mass ratio, grain radius and grain density), equal to those used to perform the run described in Section 4.1.1, and left the chemistry evolve up to  $1 \times 10^7$  years.

In all runs, the reduced network GRETOBAPE-RED reproduces almost perfectly the steady-state abundances of GRETOBAPE, for all species with abundances  $> 10^{-10}$ , with differences smaller than a factor 3 in general. However, the differences progressively worsen for trace species. A few species with abundances  $\leq 10^{-10}$  have differences larger than a factor 3 when comparing the results of GRETOBAPE-RED with those of GRETOBAPE. The main species affected are the carbon chains  $\text{C}_6$ ,  $\text{C}_3$ ,  $\text{H}_2\text{CCCO}$ ,  $c\text{-C}_3\text{H}_2\text{O}$ ,  $\text{C}_3\text{O}$ ,  $\text{H}_2\text{CCN}$ ,  $\text{H}_2\text{CCO}$ , in addition to  $\text{CH}_2\text{NH}$

and  $\text{O}_2\text{H}$ . Larger differences might be present for other trace species with abundances  $< 10^{-15}$  but, as these abundances are already very low, these molecules may not be of interest in the astrochemical simulations where the reduced network GRETOBAPE-RED is suitable. As expected, the effect on carbon chains can be attributed to the removal of larger species that can either be part of their destruction or formation routes in the case of dissociative recombination of large cations that could produce the species listed here.

These results overall indicate that the reduced network GRETOBAPE-RED is reliable to be used when following the chemistry of species with fractional abundances larger than  $\sim 10^{-10}$ , while the complete network GRETOBAPE is needed when exploring the chemistry of trace species.

## 5. CONCLUSIONS

Gas-phase reaction networks are a crucial element of any astrochemical modeling. Present day publicly available networks, e.g. KIDA and UMIST, are made up of more than 7000 reactions, of which only a tiny fraction (10–20 %) has been studied in laboratory or theoretical works.

In this work, we present a new gas-phase reaction network, GRETOBAPE, built from the publicly available KIDA2014 ([Wakelam et al. 2015](#)) with the addition of several new reactions from more recent studies. The most important novelty is that GRETOBAPE is cleaned by the most obvious source of error, the presence of endothermic reactions not recognized as such. We also present a reduced network, GRETOBAPE-RED, to be used in problems that do not require following the chemical evolution of trace species.

To this end, we performed an extended and systematic theoretical characterization of more than 500 species at CCSD(T)/aug-cc-pVTZ//M06-2X/cc-pVTZ computational level, providing for each of them: electronic state, electronic spin multiplicity, geometry, harmonic frequency, absolute electronic energy and dipole moment.

We then computed the enthalpy at 0 K of each neutral-neutral, neutral-ion and cation-anion reaction in the network. Finally, we identified the endothermic reactions not recognized as such, i.e. originally reported as barrierless or having an exponential factor  $\gamma$  not taking into account the level of endothermicity.

Following these two criteria, we deleted about 5% of the studied reactions in the original pre-cleaned network leading to the removal of 11 species. The final cleaned network GRETOBAPE consists of 7421 reactions and

489 species, while the reduced network GRETOBAPE-RED contains 3145 reaction and 205 species.

We also reported a list of probably missing reactions, namely: (i) neutral species lacking reactions with the most abundant interstellar cations ( $\text{H}_3^+$ ,  $\text{HCO}^+$ ,  $\text{H}_3\text{O}^+$ ,  $\text{He}^+$  and  $\text{H}^+$ ), and (ii) detected species absent in the network.

Using astrochemical model simulations of a typical cold molecular cloud and a warm molecular outflow shock, respectively, we measured the impact of the new cleaned GRETOBAPE on the predicted abundances. Overall, only about 5% of the species abundances are affected by more than a factor 3 with respect to the original pre-cleaned network predictions. Thioformaldehyde ( $\text{H}_2\text{CS}$ ),  $\text{Si}^+$ ,  $\text{SiN}$  and  $\text{SiS}$  are the detected species whose predicted abundances are affected by the cleaning process. Of the 5% affected abundances, the immense majority concerns Si-bearing species. We discuss in detail the origin of this large impact on the Si chemistry and conclude that a systematic review of this chemistry is needed.

We also verified that the abundances predicted by GRETOBAPE-RED of the species with abundances larger than  $10^{-10}$  are not affected by the reduction of the GRETOBAPE network, on a large parameter space. Therefore, GRETOBAPE-RED can be used when the goal is to reliably reproduce only the most abundant species and make the simulations less computational demanding.

In conclusion, we carried out the first ever systematic study of the exo/endothemicity of the reactions included in an astrochemical network. The process removing the endothermic reactions not recognized as such and the sink/source species led to the new network GRETOBAPE and its reduced version GRETOBAPE-RED. The control of the exo/endothemicity and sink/source species is only the first and obvious step to make the astrochemical gas-phase reaction networks reliable. Inclusion of the reactions of neutrals with abundant interstellar cations is a second obvious but still not complete step. The next step would be to carefully review all the reactions, rate constants and products, which is obviously unfeasible given the too large number of reactions present in the networks. A compromise could be to start with the control/verification of sub-networks that form commonly detected species, like the work done by Vazart et al. (2020) on acetaldehyde. In

addition, the lack of gas-phase reactions forming interstellar detected species does not necessarily mean that they do not exist and a lot of work here is still necessary (e.g. Balucani et al. 2015; Skouteris et al. 2018; Sleiman et al. 2018; Cernicharo et al. 2022, just to mention some recent works). Last, but not least, in addition to devote studies on the formation routes, attention should also be dedicated to the destruction routes. For example, the study of the destruction of methyl formate and dimethyl ether by  $\text{He}^+$  has shown that the rate constants included in the KIDA and UMIST networks are wrong by at least a factor 10 (Ascenzi et al. 2019).

Finally, the database created in this work, and available on-line (§ 6), can be used to rapidly test or search existing or new possible formation and destruction reaction pathways and further refined via experiments or theoretical studies to computed rate constants.

## 6. ONLINE DATABASE

All scripts and data used in this work are publicly available in the following GitHub link: [https://github.com/TinacciL/GreToBaPe\\_Cleaning](https://github.com/TinacciL/GreToBaPe_Cleaning). The networks and the database of all molecular species properties are available on the ACO (AstroChemical Origins) Database.

## ACKNOWLEDGMENTS

This project has received funding within the European Union’s Horizon 2020 research and innovation programme from the European Research Council (ERC) for the project “The Dawn of Organic Chemistry” (DOC), grant agreement No 741002, and from the Marie Skłodowska-Curie for the project “Astro-Chemical Origins” (ACO), grant agreement No 811312. SP, NB, PU acknowledge the Italian Space Agency for co-funding the Life in Space Project (ASI N. 2019-3-U.O). CINES-OCCIGEN HPC is kindly acknowledged for the generous allowance of super-computing time through the A0060810797 project. LT is grateful to Jacopo Lupi and Stefano Ferrero for insightful discussions and to the L<sup>A</sup>T<sub>E</sub>X community for the insights on TikZ and PGF-Plots packages. Finally, we wish to acknowledge the extremely useful discussions with Prof. Gretobape.

*Software:* NETWORKX (Hagberg et al. 2008), Gaussian16 (Frisch et al. 2016), KROME (Grassi et al. 2014).

## REFERENCES

Anicich, V. G. 2003, JPL-Publication-03-19

Ascenzi, D., Cernuto, A., Balucani, N., et al. 2019, A&A, 625, A72, doi: 10.1051/0004-6361/201834585

- AYOUZ, M. A., YUEN, C. H., BALUCANI, N., et al. 2019, MNRAS, 490, 1325, doi: [10.1093/mnras/stz2658](https://doi.org/10.1093/mnras/stz2658)
- BADNELL, N. 2006, The Astrophysical Journal Supplement Series, 167, 334
- BALUCANI, N. 2020, Physics of Life Reviews, 34, 136, doi: [10.1016/j.plev.2019.03.006](https://doi.org/10.1016/j.plev.2019.03.006)
- BALUCANI, N., CECCARELLI, C., & TAQUET, V. 2015, Monthly Notices of the Royal Astronomical Society, 449, L16, doi: [10.1093/mnras/slv009](https://doi.org/10.1093/mnras/slv009)
- BALUCANI, N., SKOUTERIS, D., CECCARELLI, C., et al. 2018, Molecular Astrophysics, 13, 30, doi: [10.1016/j.molap.2018.10.001](https://doi.org/10.1016/j.molap.2018.10.001)
- BARABÁSI, A.-L. 2013, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371, 20120375
- BLÁZQUEZ, S., GONZÁLEZ, D., NEEMAN, E. M., et al. 2020, Physical Chemistry Chemical Physics, 22, 20562
- CAVALLOTTI, C. 2022, Proceedings of the Combustion Institute
- CECCARELLI, C., CASELLI, P., FONTANI, F., et al. 2017, ApJ, 850, 176, doi: [10.3847/1538-4357/aa961d](https://doi.org/10.3847/1538-4357/aa961d)
- CECCARELLI, C., CODELLA, C., BALUCANI, N., et al. 2022, arXiv e-prints, arXiv:2206.13270. <https://arxiv.org/abs/2206.13270>
- CERNICHAO, J., FUENTETAJA, R., AGÚNDEZ, M., et al. 2022, A&A, 663, L9, doi: [10.1051/0004-6361/202244399](https://doi.org/10.1051/0004-6361/202244399)
- CHEIKH, S. E. 2012, PhD thesis, Ph. D. thesis
- CODELLA, C., CECCARELLI, C., CASELLI, P., et al. 2017, A&A, 605, L3, doi: [10.1051/0004-6361/201731249](https://doi.org/10.1051/0004-6361/201731249)
- CODELLA, C., CECCARELLI, C., BIANCHI, E., et al. 2020, A&A, 635, A17, doi: [10.1051/0004-6361/201936725](https://doi.org/10.1051/0004-6361/201936725)
- CRAMER, C. J. 2013, Essentials of computational chemistry: theories and models (John Wiley & Sons)
- DEFREES, D., MCLEAN, A., & HERBST, E. 1985, The Astrophysical Journal, 293, 236
- DÓBÉ, S., BÉRCES, T., & SZILÁGYI, I. 1991, Journal of the Chemical Society, Faraday Transactions, 87, 2331
- DUNHAM JR, T. 1937, Publications of the Astronomical Society of the Pacific, 49, 26
- ENDRES, C. P., SCHLEMMER, S., SCHILKE, P., STUTZKI, J., & MÜLLER, H. S. 2016, Journal of Molecular Spectroscopy, 327, 95
- ESPLUGUES, G., FUENTE, A., NAVARRO-ALMAIDA, D., et al. 2022, A&A, 662, A52, doi: [10.1051/0004-6361/202142936](https://doi.org/10.1051/0004-6361/202142936)
- FONTANI, F., CECCARELLI, C., FAVRE, C., et al. 2017, A&A, 605, A57, doi: [10.1051/0004-6361/201730527](https://doi.org/10.1051/0004-6361/201730527)
- FOURNIER, M. 2014, PhD thesis, Ph. D. thesis
- FREEDMAN, D., & DIACONIS, P. 1981, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 57, 453
- FRISCH, M. J., TRUCKS, G. W., SCHLEGEL, H. B., et al. 2016, Gaussian 16 Revision B.01
- GAO, L. G., ZHENG, J., FERNÁNDEZ-RAMOS, A., TRUHLAR, D. G., & XU, X. 2018, Journal of the American Chemical Society, 140, 2906
- GLOSIK, J., ZAKOURL, P., & LINDINGER, W. 1995, The Journal of chemical physics, 103, 6490
- GLOVER, S. C., FEDERRATH, C., MAC LOW, M.-M., & KLESSEN, R. S. 2010, Monthly Notices of the Royal Astronomical Society, 404, 2
- GRASSI, T., BOVINO, S., SCHLEICHER, D., & GIANTURCO, F. 2013, Monthly Notices of the Royal Astronomical Society, 431, 1659
- GRASSI, T., BOVINO, S., SCHLEICHER, D. R. G., et al. 2014, Monthly Notices of the Royal Astronomical Society, 439, 2386, doi: [10.1093/mnras/stu114](https://doi.org/10.1093/mnras/stu114)
- GRASSI, T., NAUMAN, F., RAMSEY, J., et al. 2021, arXiv preprint arXiv:2104.09516
- HAAS, M., HOLLENBACH, D., & ERICKSON, E. 1986, The Astrophysical Journal, 301, L57
- HAGBERG, A. A., SCHULT, D. A., & SWART, P. J. 2008, in Proceedings of the 7th Python in Science Conference, ed. G. Varoquaux, T. Vaught, & J. Millman, Pasadena, CA USA, 11 – 15
- HAMBERG, M., ZHAUNERCHYK, V., VIGREN, E., et al. 2010, A&A, 522, A90, doi: [10.1051/0004-6361/201014774](https://doi.org/10.1051/0004-6361/201014774)
- HARADA, N., & HERBST, E. 2008, The Astrophysical Journal, 685, 272
- HERBST, E. 2006, Gas Phase Reactions (New York, NY: Springer New York), 561–574, doi: [10.1007/978-0-387-26308-3\\_37](https://doi.org/10.1007/978-0-387-26308-3_37)
- HERBST, E., & KLEMPERER, W. 1973, The Astrophysical Journal, 185, 505
- HOYERMAN, K., SIEVERT, R., & WAGNER, H. G. 1981, Berichte der Bunsengesellschaft für physikalische Chemie, 85, 149
- IRVINE, W. M., FRIBERG, P., KAIFU, N., et al. 1989, The Astrophysical Journal, 342, 871
- JENKINS, E. B. 2009, The Astrophysical Journal, 700, 1299
- KENDALL JR, T. 1992, J. Chem. Phys., 96, 6796
- KNOWLES, P. J., HAMPSEL, C., & WERNER, H.-J. 1993, The Journal of chemical physics, 99, 5219
- KOOIJ, D. 1893, Zeitschrift für physikalische Chemie, 12, 155
- LAIDLER, K. J. 1996, Pure and applied chemistry, 68, 149
- LARA-MORENO, M., STOECKLIN, T., & HALVICK, P. 2019a, ACS Earth and Space Chemistry, 3, 1556, doi: [10.1021/acsearthspacechem.9b00098](https://doi.org/10.1021/acsearthspacechem.9b00098)
- LARA-MORENO, M., STOECKLIN, T., HALVICK, P., & LOISON, J.-C. 2019b, PhRvA, 99, 033412, doi: [10.1103/PhysRevA.99.033412](https://doi.org/10.1103/PhysRevA.99.033412)



- Lawson, P. A., Osborne Jr, D. S., & Adams, N. G. 2012, *The Journal of Physical Chemistry A*, 116, 2880
- Loison, J.-C., Wakelam, V., & Hickson, K. M. 2014, *MNRAS*, 443, 398, doi: [10.1093/mnras/stu1089](https://doi.org/10.1093/mnras/stu1089)
- Loison, J.-C., Wakelam, V., Hickson, K. M., Bergeat, A., & Mereau, R. 2013, *Monthly Notices of the Royal Astronomical Society*, 437, 930
- Loison, J.-C., Agúndez, M., Marcelino, N., et al. 2016, *MNRAS*, 456, 4101, doi: [10.1093/mnras/stv2866](https://doi.org/10.1093/mnras/stv2866)
- Mancini, L., & Aragão, E. V. F. d. 2021, in *International Conference on Computational Science and Its Applications*, Springer, 339–350
- Marcelino, N., Cernicharo, J., Roueff, E., Gerin, M., & Mauersberger, R. 2005, *The Astrophysical Journal*, 620, 308
- McGuire, B. A. 2022, *The Astrophysical Journal Supplement Series*, 259, 30
- McKellar, A. 1940, *Publications of the Astronomical Society of the Pacific*, 52, 187
- Meot-Ner, M., Karpas, Z., & Deakayne, C. A. 1986, *Journal of the American Chemical Society*, 108, 3913
- Minowa, H., Satake, M., Hirota, T., et al. 1997, *The Astrophysical Journal*, 491, L63
- Mota, V., Varandas, A., Mendoza, E., Wakelam, V., & Galvão, B. 2021, *The Astrophysical Journal*, 920, 37
- Nelson, R. P., & Langer, W. D. 1999, *The Astrophysical Journal*, 524, 923
- Neufeld, D. A., Godard, B., Gerin, M., et al. 2015, *Astronomy & Astrophysics*, 577, A49
- Newman, M. 2018, *Networks* (Oxford university press)
- Ohishi, M., & Kaifu, N. 1998, *Faraday Discussions*, 109, 205, doi: [10.1039/a801058g](https://doi.org/10.1039/a801058g)
- Oppenheimer, M., & Dalgarno, A. 1974, *The Astrophysical Journal*, 192, 29
- Parisel, O., Hanus, M., & Ellinger, Y. 1996, *The Journal of Physical Chemistry*, 100, 2926
- Penteado, E., Walsh, C., & Cuppen, H. 2017, *The Astrophysical Journal*, 844, 71
- Podio, L., Codella, C., Lefloch, B., et al. 2017, *MNRAS*, 470, L16, doi: [10.1093/mnrasl/slx068](https://doi.org/10.1093/mnrasl/slx068)
- Prasad, S., & Huntress Jr, W. 1980, *The Astrophysical Journal Supplement Series*, 43, 1
- Rosi, M., Mancini, L., Skouteris, D., et al. 2018, *Chemical Physics Letters*, 695, 87, doi: [0.1016/j.cplett.2018.01.053](https://doi.org/10.1016/j.cplett.2018.01.053)
- Schilke, P., Leurini, S., Menten, K., & Alcolea, J. 2003, *Astronomy & Astrophysics*, 412, L15
- Skouteris, D., Balucani, N., Ceccarelli, C., et al. 2019, *MNRAS*, 482, 3567, doi: [10.1093/mnras/sty2903](https://doi.org/10.1093/mnras/sty2903)
- . 2018, *ApJ*, 854, 135, doi: [10.3847/1538-4357/aaa41e](https://doi.org/10.3847/1538-4357/aaa41e)
- Skouteris, D., Vazart, F., Ceccarelli, C., et al. 2017, *MNRAS*, 468, L1, doi: [10.1093/mnrasl/slx012](https://doi.org/10.1093/mnrasl/slx012)
- Sleiman, C., El Dib, G., Rosi, M., et al. 2018, *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)*, 20, 5478, doi: [10.1039/C7CP05746F](https://doi.org/10.1039/C7CP05746F)
- Sleiman, C., El Dib, G., Talbi, D., & Canosa, A. 2018, *ACS Earth and Space Chemistry*, 2, 1047
- Smith, I. W. 2006, *Angewandte Chemie International Edition*, 45, 2842
- . 2011, *Annual Review of Astronomy and Astrophysics*, 49, 29
- Spezzano, S., Sipilä, O., Caselli, P., et al. 2022, *A&A*, 661, A111, doi: [10.1051/0004-6361/202243073](https://doi.org/10.1051/0004-6361/202243073)
- Su, T., & Chesnavich, W. J. 1982, *The Journal of Chemical Physics*, 76, 5183
- Swings, P., & Rosenfeld, L. 1937, *Astrophysical Journal*, 86, 483
- Tercero, B., Vincent, L., Cernicharo, J., Viti, S., & Marcelino, N. 2011, *A&A*, 528, A26, doi: [10.1051/0004-6361/201015837](https://doi.org/10.1051/0004-6361/201015837)
- Tinacci, L., Pantaleone, S., Maranzana, A., et al. 2021, *The Astrophysical Journal Supplement Series*, 256, 35
- Turner, B. 1992, *The Astrophysical Journal*, 388, L35
- Urso, R. G., Scirè, C., Baratta, G. A., Compagnini, G., & Palumbo, M. E. 2016, *A&A*, 594, A80, doi: [10.1051/0004-6361/201629030](https://doi.org/10.1051/0004-6361/201629030)
- Urso, R. G., Palumbo, M. E., Ceccarelli, C., et al. 2019, *A&A*, 628, A72, doi: [10.1051/0004-6361/201834322](https://doi.org/10.1051/0004-6361/201834322)
- Vastel, C., Quénard, D., Le Gal, R., et al. 2018, *MNRAS*, 478, 5514, doi: [10.1093/mnras/sty1336](https://doi.org/10.1093/mnras/sty1336)
- Vazart, F., Calderini, D., Puzzarini, C., Skouteris, D., & Barone, V. 2016, *Journal of Chemical Theory and Computation*, 12, 5385, doi: [10.1021/acs.jctc.6b00379](https://doi.org/10.1021/acs.jctc.6b00379)
- Vazart, F., Ceccarelli, C., Balucani, N., Bianchi, E., & Skouteris, D. 2020, *MNRAS*, 499, 5547, doi: [10.1093/mnras/staa3060](https://doi.org/10.1093/mnras/staa3060)
- Wakelam, V., Herbst, E., Selsis, F., & Massacrier, G. 2006, *Astronomy & Astrophysics*, 459, 813
- Wakelam, V., Herbst, E., Loison, J.-C., et al. 2012, *The Astrophysical Journal Supplement Series*, 199, 21
- Wakelam, V., Loison, J.-C., Herbst, E., et al. 2015, *The Astrophysical Journal Supplement Series*, 217, 20
- Watts, J. D., Gauss, J., & Bartlett, R. J. 1993, *The Journal of chemical physics*, 98, 8718
- Wheeler, S. E., Houk, K. N., Schleyer, P. v. R., & Allen, W. D. 2009, *Journal of the American Chemical Society*, 131, 2547
- Woon, D. E., & Dunning Jr, T. H. 1993, *The Journal of chemical physics*, 98, 1358

Woon, D. E., & Herbst, E. 2009, *The Astrophysical Journal*  
Supplement Series, 185, 273

Zhao, Y., & Truhlar, D. G. 2008, *Theoretical Chemistry*  
*Accounts*, 120, 215

## Part III

# Thermal Desorption of Interstellar Molecules

# Introduction

**Context** Within all the processes considered inside the grain-phase reaction network in this thesis, thermal desorption is of crucial importance in astrochemical models since it rules whether a molecule is in the gas or grain-phase. The thermal desorption rate constant is normally mathematically described as an Arrhenius process in which the pre-exponential factor ( $\nu$ ) and the binding energy (BE) are the parameters that collect all the chemical information of the process, *i.e.*

$$k_{des} = \nu \exp\left(\frac{\text{BE}}{\text{RT}}\right),$$

where T is the temperature, R the ideal gas constant. The BE has a key role in astrochemical models since it is used to model not only the thermal desorption but also the thermal diffusion, since the diffusion energy barriers are approximated to be a BE fraction (see Par. 1.2.2). For this reason, many studies were focused, mainly, on the determination of BEs meanwhile the  $\nu$  was normally assumed to be a constant value of  $\sim 1 \times 10^{12} \text{ s}^{-1}$  or computed via the following formula provided by [Hasegawa and Herbst \(1993\)](#):

$$\nu = \sqrt{\frac{2 \text{BE}}{\pi^2 m A}},$$

where A is the surface area per adsorbed molecule (normally assumed to be  $10^{-19} \text{ m}^2$ ) and m is the mass of the adsorbed species. I think that the role of the latter quantity has been underestimated by astrochemists, especially when compared to the attention given by the material science/catalysis scientists.

Anyways, many uncertainties and fragmentary information regarding BE and  $\nu$  are still present in the astrochemical databases, despite over the years, a great effort from the astrochemical community has been done to provide the above-mentioned quantities via experimental or computational studies.



If Temperature-Programmed Desorption (TPD) experiments (the only technique used in astrochemistry so far) cannot provide information on an atomic scale, due to the detection methods, computational chemistry can provide a more detailed and accurate picture of the thermal desorption process and overcome some of the experimental limitations. However, in order to provide accurate data via computational chemistry, a reliable surface model and an accurate level of theory need to be used.

In my opinion, many astrochemical computational works on BEs, present important limitations. In many cases, the icy grain surface model does not receive the attention it deserves, as evidenced by the fact that its physicochemical properties are almost always not provided. Recently, few ice models with 1 to 20 water molecules have been considered to compute BEs of different species. However, first, those models cannot describe all the chemical interactions between the absorbed molecule and a “*real*” large surface of icy water and, secondly, they only have a few absorption sites.

For these reasons, it is important to provide absorption distribution of cases for important astrochemical species, without forgetting to use an accurate level of theory and a reliable surface model.

**Goals and novelty of the research** From a pure computational chemistry perspective, the main aim of this part is to provide a more accurate and complete image of the thermal desorption process of astrochemical molecules from icy surface models. This goal can be obtained with the development of: (i) a large ice grain mantle model built up with accurate calculations and its physicochemical characterization, (ii) an accurate and unbiased methodology to compute a distribution of binding sites on the icy grain and its application on some astrochemical important species and (iii) a detailed comparison on the commonly used pre-exponential factors in astrochemistry.

**Summary of the results** The first step in order to obtain our goals was to develop an amorphous water icy cluster model that can emulate the icy grain mantle. This result was principally obtained by Aurèle Germain, ITN-ACO PhD student in Turin, with whom I collaborated in this project. For obtaining an accurate model, we used the new semiempirical program GFN-xTB using mainly the GFN2 semiempirical tight-binding electronic Hamiltonian, which shows incredible accuracy and performances. Using the latter program, a python script called ACO-FROST<sup>1</sup> was developed to build up an ice grain mantle model via a step by step addition of randomly oriented water

---

<sup>1</sup>[https://github.com/aurelegermain/ACO-FROST\\_grain\\_generator](https://github.com/aurelegermain/ACO-FROST_grain_generator)

molecules. Careful checks on properties of some built models were carried out to assess their physico-chemical features.

The obtained ice models resulted in a density,  $\sim 1.2 \text{ gr/cm}^3$ , similar to that usually obtained in experiments of compact ices,  $\sim 1.1 \text{ gr/cm}^3$  (Narten et al., 1976; Jenniskens et al., 1995). Please note that, in the published article (Germain et al., 2022) integrally reported in Chapter 7, we erroneously compared our derived density with a crystalline ice, for which experiments find  $\sim 0.92 \text{ gr/cm}^3$  (Haynes et al., 2016). Please note that we attributed the difference to the GFN2 method and the van der Waals radii, but actually there is no difference to discuss. We plan to submit an errata-corrige to correct this part of the published article.

An icy grain mantle model of 200 water was finally created and I used it as starting point for the development of a new framework for computing a distribution of adsorption sites (and so their BE distribution) of any relevant interstellar species, in a reproducible and unbiased manner. This framework has a general applicability since the electronic structure calculations relies on the ONIOM method (see Subsection 2.1.5), which allows to choose the best compromise between accuracy and computational cost. All the structures and harmonic frequencies were obtained at ONIOM(B97-D3/aug-cc-pVTZ:xTB:GFN2) level of theory, while single point refinements (and BSSE corrections) were performed at ONIOM(DLPNO-CCSD(T)/aug-cc-pVTZ:xTB:GFN2) level. We applied this methodology to ammonia first and water<sup>2</sup> after. These results were compared and discussed with respect to previous experimental and computational literature results. From this comparison, I realized that computational studies do not really measure the same quantities as the TPD experiments and, maybe, the computational data are closer to the process that takes place in the ISM. Both the water and ammonia BE distributions show a low-end BE tail at low energy, which might explain the observed presence of the gas-phase of these molecules in cold regions, such as prestellar core and protoplanetary disk.

Besides the improvement in the calculation of accurate BEs, less attention was focused to  $\nu$  despite it plays a role in obtaining the thermal desorption constant. Using water, ammonia and methanol adsorption cases, as test molecules, on ice cluster and ice periodic surface I tested the impact of different pre-exponential factors on simulated TPD experiments. I compared the formula coming from Hasegawa and Herbst (1993), Tait et al. (2005) and its harmonic oscillator extension<sup>3</sup>. The difference in computing

---

<sup>2</sup>In this case, the DFT used for structure and frequency calculations was the B97-3c for its high accuracy/computational cost ratio.

<sup>3</sup>The Tait et al. (2005) and its harmonic oscillator extension for computing the pre-

the desorption temperature peak, in experimental conditions, adopting the [Tait et al. \(2005\)](#) prefactor and its harmonic oscillator extension is minimal. On the other hand, the difference between these two latter approaches and the [Hasegawa and Herbst \(1993\)](#) formula spans from 20 to 60 K, depending on the cases.

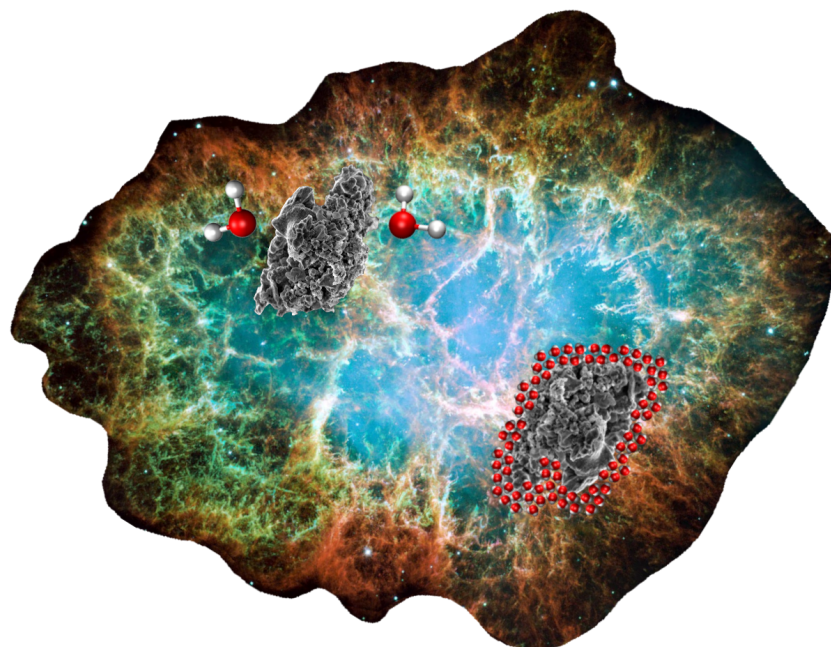
**Structure of this thesis's part** In order to achieve the above-mentioned goals two articles were published ([Germain et al., 2022](#); [Tinacci et al., 2022](#)) and other two will be submitted soon. Each of these articles constitutes the four following Chapters of this thesis's part. The first Chapter (Chap. 7) focuses on the creation and characterization of a icy grain mantle model, while using the latter model, the second Chapter (Chap. 8) on the development of an accurate BE distribution methodology and its application on the ammonia case. In the third Chapter (Chap. 9), the application of the built methodology to the water absorption case and its possible impact in protoplanetary snow-line is presented. Last but not least, the final Chapter (Chap. 10) deals with a detailed comparison on the commonly used pre-exponential factors in astrochemistry.

---

exponential factors are based on the Transition State Theory (TST). In modelling the desorption of physisorbed species with TST a gas-like transition state is assumed with a barrier and structure equal to BE and the not interacting (isolated) species and surface, respectively. We decided to keep the same nomenclature and approach used by [Tait et al. \(2005\)](#) (and in many other articles) to highlight the difference with respect to [Hasegawa and Herbst \(1993\)](#).

# Computer generated realistic interstellar icy grain models: physicochemical properties and interaction with ammonia

The hereafter article has been published in ACS Earth Space Chemistry 2022, 6, 5, 1286–1298.



# Computer Generated Realistic Interstellar Icy Grain Models: Physicochemical Properties and Interaction with NH<sub>3</sub>

Aurèle Germain, Lorenzo Tinacci, Stefano Pantaleone, Cecilia Ceccarelli, and Piero Ugliengo\*

Cite This: <https://doi.org/10.1021/acsearthspacechem.2c00004>

Read Online

ACCESS |

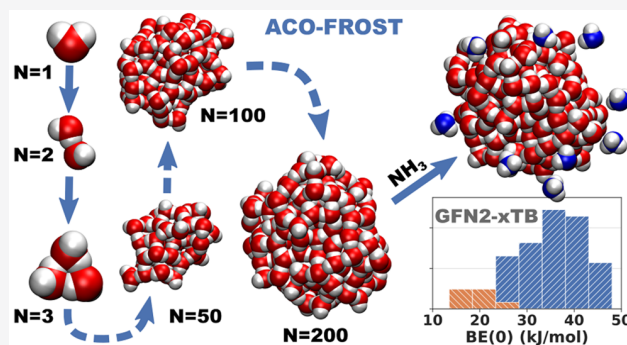
Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** Interstellar grains are composed by a rocky core (usually amorphous silicates) covered by an icy mantle, the most abundant molecule being H<sub>2</sub>O followed by CO, CO<sub>2</sub>, NH<sub>3</sub>, and also radicals in minor quantities. In dense molecular clouds, gas-phase chemical species freeze onto the grain surface, making it an important reservoir of molecular diversity/complexity whose evolution leads to interstellar complex organic molecules (iCOMs). Many different models of water clusters have appeared in the literature, but without a systematic study on the properties of the grain (such as the H-bonds features, the oxygen radial distribution function, the dangling species present on the mantle surface, the surface electrostatic potential, etc.). In this work, we present a computer procedure (ACO-FROST) grounded on the newly developed semiempirical GFN2 tight-binding quantum mechanical method and the GFN-FF force field method to build-up structures of amorphous ice of large size. These methods show a very favorable accuracy/cost ratio as they are ideally designed to take noncovalent interactions into account. ACO-FROST program can be tuned to build grains of different composition mimicking dirty icy grains. These icy grain models allow studying the adsorption features (structure, binding energy, vibrational frequencies, etc.) of relevant species on a large variety of adsorption sites so to obtain a statistically meaningful distribution of the physicochemical properties of interest to be transferred in numerical models. As a test case, we computed the binding energy of ammonia adsorbed at the different sites of the icy grain surface, showing a broad distribution not easily accounted for by other more size limited icy grain models. Our method is also the base for further refinements, adopting the present grain in a more rigorous QM:MM treatment, capable of giving binding energies within the chemical accuracy.

**KEYWORDS:** Amorphous ice, GFN-xTB, GFN2, water clusters, adsorption, binding energy



## INTRODUCTION

Micrometer grain particles made of silicate and carbonaceous materials (core)<sup>1,2</sup> are present in the interstellar medium (ISM). In dense molecular clouds, the low temperature (<20 K) and relatively high molecular density ( $\geq 10^4$  cm<sup>-3</sup>) favor the formation and subsequent freezing of molecules on top of these dust grains, forming amorphous layers of ice (aka mantles).<sup>3–5</sup> Classical laboratory experiments using N<sub>2</sub> adsorption isotherms on iced water revealed that, as a function of the formation conditions, micro porosity of the 2–3 nm size can be detected.<sup>6</sup> The mantles are made mostly of water ice, formed by hydrogenation of oxygen atoms<sup>7–9</sup> or by reactions of (OH)<sub>2</sub> as studied by Redondo et al.,<sup>10,11</sup> but molecules present in the gas phase (such as carbon monoxide) can also freeze onto the mantles of water ice.<sup>3</sup> This richness of molecules make icy grains a host for important chemical reactions, such as the synthesis of H<sub>2</sub><sup>12</sup> (the most abundant molecule in the ISM<sup>12,13</sup>), dictated by the configuration of adsorption sites on the grain surface. The paths of formation of the most astrochemically relevant iCOMs have been recently

reviewed.<sup>14</sup> Other factors such as cosmic rays have been observed to compact the ice layers of interstellar grains.<sup>15</sup> Hence, the study of adsorption, desorption, diffusion, and reaction of molecules on top of the icy grain surface is of high importance if we want to understand the evolution overtime of the chemical composition of the ISM. To study the atomistic details of these processes by computational chemistry methods we need icy grain models that embody the observed properties of interstellar icy grains.

As far as we know, no public grain models are available to the astrochemical community. Some models have been proposed in the literature, as summarized in the next few paragraphs, with the scope to satisfy requirements of specific

Received: January 4, 2022

Revised: February 28, 2022

Accepted: March 21, 2022



cases, i.e., to study a given adsorbate or particular reaction features of interest. None of them was built with the purpose to have, in a single model grain, all the features exhibited by a real interstellar amorphous ice grain, i.e., in which adsorption sites exhibit different electrostatics, hydrogen bond features, and local dispersive interactions with respect to any adsorbed molecule.

The simplest model was proposed by Wakelam et al.,<sup>16</sup> adopting a single water molecule as a model grain to compute the binding energies of dozens of chemical species at density functional level of theory (DFT), more precisely the M06-2X functional associated with the Gaussian aug-cc-pVTZ basis set. The computed BE are then corrected for the missing effects due to the adoption of a single water molecule, by linear fitting experimental BE against computed ones. The resulting global scaling factor brings the computed BEs closer to the experimental ones. Even though the procedure is clever and computationally very cheap, this approach cannot portray the richness of different binding sites at the grain and cannot provide the expected BE distribution.

Like Wakelam et al., Das et al.<sup>17</sup> computed the BEs of a similar set of molecules using clusters from 1 up to 6 water molecules with the MP2 method and an aug-cc-pVDZ basis set. They showed a decreasing deviation between experimental and theoretical BEs by increasing the size of the water cluster. As for Wakelam et al. these clusters had, however, a limited number of hydrogen bonds per water molecules, decreasing the internal hydrogen bond cooperativity and the variety of binding sites.

To the opposite situation, grain models of hundred of thousands of water molecules have been built using special Lennard-Jones (LJ) potentials by Garrod<sup>18</sup> and Christianson and Garrod.<sup>19</sup> While these models match the interstellar icy grain sizes, they are not able to represent the hydrogen bond features linking each water molecules, due to the roughness of the adopted LJ potential. Furthermore, the size is so large that it is impossible to adopt DFT to characterize their electronic features.

In our group, we have also proposed both crystalline and amorphous ice models, within the periodic boundary conditions. In Zamirri et al.,<sup>20</sup> three different periodic crystalline models were used to study the adsorption of CO on water ice using various DFT methods. However, the amorphous nature of the icy grains was missing. In Ferrero et al.,<sup>21</sup> the adsorption of different molecules was studied on periodic crystalline and amorphous water ice models using DFT (B3LYP-D3/A-VTZ\* and M06-2X/A-VTZ\*). The amorphous unit cell included 60 water molecules allowing to identify from three to eight different binding sites as a function of the considered adsorbate molecule, providing some variability in the computed BE values. In that work, however, the adsorbates were manually positioned at the ice sites to maximize the hydrogen bond strength, likely biasing the final values of the BEs.

Sameera et al.<sup>22</sup> used a crystalline ice structure to represent icy grains surfaces and computed the BE of several radical species such as CH<sub>3</sub>. Up to 16 crystalline models with approximately 160 water molecules exhibiting 16 different binding sites, were built up using the crystalline water ice structure. The clusters were treated with a QM:MM method containing at least 44 H<sub>2</sub>O for the QM (wB97XD/def2-TZVP) part and 112 for the MM (AMBER force field or the AMOEB09 polarizable force field) part. In the same work,

amorphous solid water cluster models of 162 H<sub>2</sub>O molecules exhibiting up to 10 different binding sites were also used to study CH<sub>3</sub> radicals with a QM:MM method (49 molecules for the QM part and 113 for the MM part).

Shimonishi et al.<sup>23</sup> used nine amorphous water clusters including 20 water molecules (optimized with  $\omega$ -B97XD and the Gaussian 6-311+G basis set) to represent different binding sites that could be found on the grain, even if the sizes are still rather limited to provide enough variability. Rimola et al.,<sup>24</sup> Duvernay et al.,<sup>25</sup> and Enrique-Romero et al.<sup>26</sup> all used water clusters ranging from 18 to 32 water molecules and various DFT methods to study chemical reactions bringing simple radical species to iCOMs. In that case, due to the highly demanding cost of characterizing the entire potential energy surface of the reactions, the site variability was less relevant than for computing the BEs. Bovolenta et al.<sup>27</sup> used two types of clusters made of 22 and 37 water molecules (revPBE0/def2-TZVP) to study the adsorption of HF and found minimal changes in the BE distributions between the two cluster sizes.

Last, Song and Kästner<sup>28</sup> proposed an icy model (also used in Lamberts and Kästner,<sup>29</sup> Molpeceres et al.,<sup>30</sup> and Molpeceres and Kästner<sup>31</sup>) as an hemisphere of 499 water molecules in which 295 are free to react, cut out from a 18937 water molecules cell that was equilibrated at 300 K for 100 ps and then suddenly quenched at 10 K for 50 ps using the NAMD code (in the NVT ensemble and using a Langevin thermostat). This model is close to the one we are proposing here, but its construction does not have any resemblance with the process occurring in the ISM in which water is added step by step at a very low temperature.

The interested reader can find more about modeling interstellar ice in the recent papers by Zamirri et al.<sup>14</sup> and Cuppen et al.<sup>32</sup>

In this paper, we will describe a new approach to automatically building up an amorphous icy grain model constituted by several hundreds of water molecules. We labeled our models as “realistic”, not based on a size criterion only but requiring the following essential features: (i) the grain should not be minimal, i.e., envisaging 20–30 water molecules only, as usually found in the literature, but includes at least few hundreds water molecules; (ii) the hydrogen bond features within the icy grain should be accurately represented by adopting a proper quantum level of theory; (iii) the ice should be amorphous by construction, avoiding the usual approach of heating at high  $T$  and sudden cooling of a crystalline ice model, which has no counterpart in the grain evolution in the ISM. Indeed, the grain is not derived from crystalline ice or the like, but is built from scratch by sequentially adding water molecules in a way comparable to ice generation in terrestrial laboratories. In the ISM conditions, as already discussed, water is formed *in situ* and we will provide the reasons why this process is very hard to be simulated quantum mechanically (*vide infra*). As for the adopted computational method, we choose the newly developed xTB semiempirical quantum mechanical GFN2 method<sup>33</sup> and the force field GFN-FF,<sup>34</sup> both developed by Grimme’s group at the University of Bonn. Both methods showed excellent results with system dominated by noncovalent interactions.<sup>34–38</sup> In previous papers, we also showed that GFN2 was close in accuracy to CCSD(T)/CBS for small water clusters<sup>39</sup> and very accurate when compared to DFT methods for the computation of binding energies of several molecules adsorbed on a crystalline water ice slab,<sup>40</sup> making them well suited for our purpose. On the generated icy

grain, we showed how to compute the BEs of ammonia ( $\text{NH}_3$ ) on more than 100 different adsorbing sites of a 200 water icy grain, providing the BE distribution at very low computational cost with reliable BE values.

## COMPUTATIONAL DETAILS

**Methodology.** All the calculations were carried out with the xTB program<sup>41</sup> (version 6.3.3 for the binding energy sampling of  $\text{NH}_3$ , and 6.4.0 for the icy grain model of 1000 water molecules), using two different methods: the semi-empirical GFN2<sup>33</sup> and the classical force field GFN-FF.<sup>34</sup> For the grain building process, all geometry optimizations at GFN2 and GFN-FF levels were performed using the default settings of the code (i.e.,  $5 \times 10^{-6}$  hartree and  $1 \times 10^{-3}$  hartree bohr<sup>-1</sup> on energy and gradient, respectively) while molecular dynamics simulations were done only with GFN-FF with the temperature set to 10 K to mimic the ISM conditions, a time step of 1.0 fs, and a total time of 1 ps. For the BE distribution, all calculations were performed using GFN2 and the default settings except for the precision of geometry optimizations that was set to “extreme” (i.e.,  $5 \times 10^{-8}$  hartree and  $5 \times 10^{-5}$  hartree bohr<sup>-1</sup> on energy and gradient, respectively). Constraints on fixed atoms were used through the keyword “constrain” which apply a force constant of 0.5 hartree bohr<sup>-1</sup> to the distance among selected atoms.

The different action performed for the grain building process and the binding energy sampling explained in the following sections were automated via two Python scripts combining approximately two thousands lines of codes and using standard packages such as Numpy and Scipy, and more specialized ones such as the Atomic Simulation Environment (ASE). The set of scripts named “ACO-FROST” are freely available (see [Data and Software Availability](#)).

The BEs of  $\text{NH}_3$  at the grain were computed as follows:

$$\text{BE} = E(\text{NH}_3) + E(\text{grain}) - E(\text{NH}_3 + \text{grain}) \quad (1)$$

Here  $E(\text{NH}_3)$ ,  $E(\text{grain})$ , and  $E(\text{NH}_3 + \text{grain})$  are the GFN2 electronic energies after geometry optimization of  $\text{NH}_3$ , the water grain with the constraints applied (*vide infra*), and  $\text{NH}_3$  interacting with the grain, respectively.

To confirm that the optimized geometries are minima on the potential energy surface, we computed the harmonic vibrational frequencies and checked that all frequencies were real-valued. Structures with imaginary frequencies were discarded. Careful check of each of these structures revealed that the imaginary frequency belongs to atoms of the local region where  $\text{NH}_3$  is adsorbed.

From the set of frequencies, the zero point energy (ZPE) is computed for each component to arrive to the zero point energy correction ( $\Delta\text{ZPE}$ ):

$$\Delta\text{ZPE} = \text{ZPE}(\text{NH}_3 + \text{grain}) - \text{ZPE}(\text{NH}_3) - \text{ZPE}(\text{grain}) \quad (2)$$

to compute the zero point energy corrected binding energies  $\text{BE}(0)$ , by subtracting the  $\Delta\text{ZPE}$  to the electronic BE:

$$\text{BE}(0) = \text{BE} - \Delta\text{ZPE} \quad (3)$$

In general, the  $\Delta\text{ZPE}$  is a positive quantity; therefore, the  $\text{BE}(0)$  is always smaller than the electronic BE.

To characterize the shape of the generated icy grains we computed the gyration tensor as described in Theodorou and Suter.<sup>42</sup> For a three-dimensional Cartesian coordinate system, it reads:

$$\mathbf{S} = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \quad (4)$$

With each component  $S_{mn}$  of the  $\mathbf{S}$  matrix as

$$S_{mn} = \frac{1}{N} \sum_{i=1}^N m^{(i)} n^{(i)}, \quad m, n = x, y, z \quad (5)$$

with  $N$ , the number of atoms in the system, and  $m^{(i)}$ , the Cartesian coordinate  $m = x, y, z$  of the  $i$ th atom.

By diagonalizing the  $\mathbf{S}$  matrix the eigenvalues  $\lambda_x^2$ ,  $\lambda_y^2$ , and  $\lambda_z^2$  are ordered as  $\lambda_x^2 \leq \lambda_y^2 \leq \lambda_z^2$ . From them, we derived the gyration radius  $R$ , the asphericity  $b$ , and the relative shape anisotropy  $\kappa^2$  defined as

$$R = \sqrt{\lambda_x^2 + \lambda_y^2 + \lambda_z^2}; \quad b = \lambda_z^2 - \frac{1}{2}(\lambda_x^2 + \lambda_y^2);$$

$$\kappa^2 = \frac{3}{2} \frac{\lambda_x^4 + \lambda_y^4 + \lambda_z^4}{(\lambda_x^2 + \lambda_y^2 + \lambda_z^2)^2} \quad (6)$$

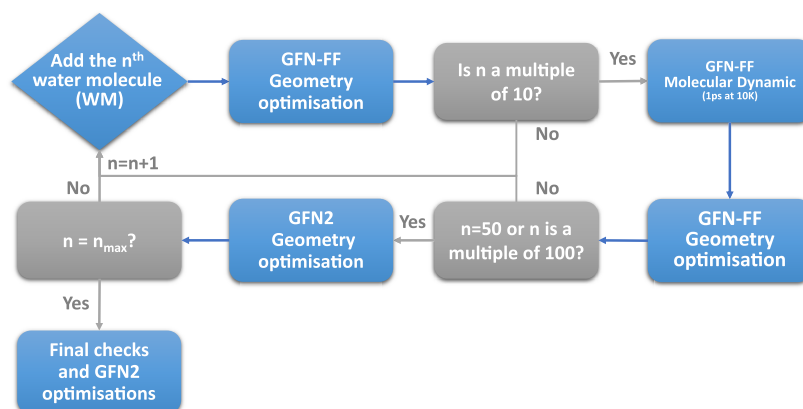
The asphericity  $b$  is greater than zero, becoming exactly zero only when  $\lambda_x = \lambda_y = \lambda_z$  (spherically symmetric system). The relative shape anisotropy  $\kappa^2$  is bounded between zero (all points are spherically symmetric) and one (all points lie on a line).

Hydrogen bonds were sampled by computing the O...H distances of the GFN-xTB optimized structures and imposing a lower/upper threshold of 1.2/2.2 Å to filter out the O...H bonds. Dangling hydrogens were sampled by counting every hydrogen without an O distance inferior to 2.2 Å; dangling oxygens are defined as oxygen atoms free from an O...H distance smaller than 2.2 Å. To classify dangling hydrogen/oxygen atoms a minimum distance from the structure center of mass was forced to limit the influence of possible dangling species present inside the structure.

We also computed the density  $\rho$  of each grain by assigning van der Waals radii to H/O of 1.30/2.05 Å and using a Monte Carlo integration scheme (using 30000 sampling points) to compute the corresponding grain volume as encoded in the MOLDRAW program.<sup>43</sup> The adopted radii were chosen so that the computed density  $\rho$  of the experimental structure of crystalline ice  $I_h$  resulted in the experimental<sup>44</sup> value  $\rho = 0.92$  gr/cm<sup>3</sup>.

We also render many of the static structures shown in the figures as hyperactive molecules using the JSmol rendering engine.<sup>45</sup> These were made accessible through specific QR codes readable through the standard mobile phone camera.

**Grain Building Process.** Water can be formed by gas-phase reactions. However, by far the largest fraction of the total amount of water in cold ( $\leq 200$  K) ISM comes from reactions between oxygen and hydrogen atoms at both the grain core and on top of an already formed icy mantle.<sup>46</sup> Simulating this process is, however, computationally very expensive and almost impossible to carry out on icy clusters of large size. To overcome this difficulty, we assume that the water molecules are already formed and that they are adsorbed on top of an initial seed (a single water molecule). It is worth noting that laboratory experiments meant to simulate the ISM ice grow the grain by addition of preformed water molecules adsorbed on specific inert supports. This implies that the energy transfer between the newly formed OH bonds toward the grain, with



**Figure 1.** Flowchart of the grain building process (a video showing the building process of the 1000 water molecules grain is available in the Supporting Information).

possible restructuring of the ice, is missing within this approach. However, recent studies from our lab (Pantaleone et al.<sup>47,48</sup>) at *ab initio* molecular dynamics level of the H<sub>2</sub> and HCO radical formation on both crystalline and amorphous ice models (extended models within the periodic boundary conditions) did not show signs of amorphization of the ice, irrespective of its initial degree of crystallinity. Nevertheless, we decided to amend for the potential missing energy transfer by running a short MD computation at 10 K at given steps during the ice accretion (*vide infra*).

The grain building process is iterated in a series of sequential steps adding water from a random direction, internal orientation and distance from the grain surface (between 2.5 and 3 Å, slightly more than the typical H-bond length), until the desired number of water molecules is reached. A schematic representation of the procedure is summarized in Figure 1. At each new added water molecule the structure is first optimized at GFN-FF level. As discussed before, to simulate the conditions of the ISM in which a fraction of the heat of the water formation is transferred to the grain increasing its local temperature,<sup>47</sup> every 10 added water molecules a NVT molecular dynamics at 10 K is performed for 1 ps to simulate the water relaxation, directly followed by a GFN-FF optimization. We counted, on average, from a sample of 20 clusters of 200 water molecules, an increase of 6 hydrogen bonds as a result of the MD step followed by a GFN-FF optimization. One drawback of GFN-FF is its tendency to produce ice structures which are too fluffy, with many water molecules with incomplete H-bond coordination behaving like dangling species. To avoid this nonphysical feature, at regular intervals during the grain building process, a GFN2 full optimization is performed. At variance with GFN-FF, GFN2 shrinks the size of the cluster, to maximize the hydrogen bond (H-bond) interactions. This allows water molecules to form a compact structure which is preserved even after the subsequent GFN-FF optimizations (see Figure S1 of the Supporting Information). This strategy ensures the best compromise between speed and quality of the intermolecular interactions in the formed grain. Table 1 shows the comparison of the computational time between the two methods for a 200 water cluster.

When the desired number of water molecules is reached a final GFN2 full relaxation is performed, followed by an automatic quality check on the grain structure aimed at correcting any geometrical oddity that could have arisen during

**Table 1.** Comparison between the GFN2 and GFN-FF Geometry Optimization Times for Different Grain Sizes<sup>a</sup>

Number of water molecules	GFN2	GFN-FF	ratio GFN2/GFN-FF
200	34 min	1 min 26 s	24
400	5 h 15 min	7 min	44
800	1 day 17 h	53 min	46

<sup>a</sup>The xTB program was run on 20 core Intel Xeon E5-2630 v4 @ 2.20 GHz.

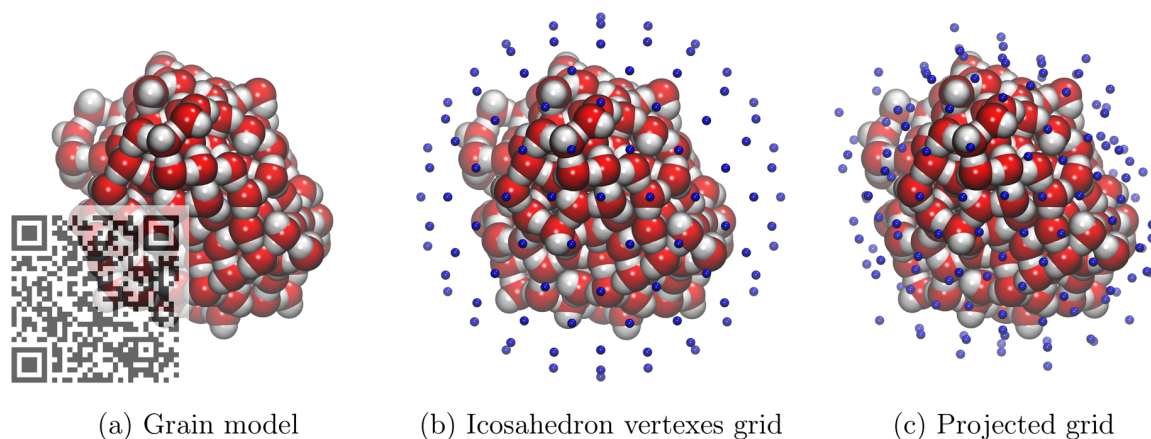
the grain building, such as water molecules interacting with a single H-bond. If found, these water molecules are removed, and the process goes back a few steps by removing the upper water layers and to reconstruct again the grain until no geometrical oddities are present and the number of desired water molecules is reached.

**Binding Energy Sampling.** To compute the binding energy distribution of NH<sub>3</sub> a grain consisting of 200 water molecules was used, as it represents the best trade-off between the number of binding sites (large enough to have a significant statistical distribution) and the computational cost.

To sample all possible adsorption sites of NH<sub>3</sub> around the grain model, a discrete grid of points taken from Teanby<sup>49</sup> was superimposed to the grain surface, where each point represents the NH<sub>3</sub> barycenter. In this way, a uniform and complete coverage of NH<sub>3</sub> molecules around the icy grain was achieved, ensuring an unbiased BE sampling of strong and weak binding energies.

The grid is initially made up of 12 vertices of an icosahedron and serves as starting position for NH<sub>3</sub>. These 12 initial vertices can then be duplicated several times to obtain additional subgrids with an increased number of vertices. This implies that the vertices of a subgrid will be also included in all subsequent tighter levels, without the need to recalculate the points of the previous level (see Figure S3 of the Supporting Information). Figure 2 shows the bare grain, and the grid level used for the NH<sub>3</sub> BE sampling. It contains 162 vertices which are used as barycenter for NH<sub>3</sub> and then projected on top of the grain structure. The projection brings a distance between 2.5 and 3 Å from the grain, used for NH<sub>3</sub> positioning. For each positioning, NH<sub>3</sub> is also randomly rotated to ensure an unbiased contact with the grain surface. The BE is computed for each position (blue dot) occupied by NH<sub>3</sub>, resulting in 162 different values (*vide infra*).



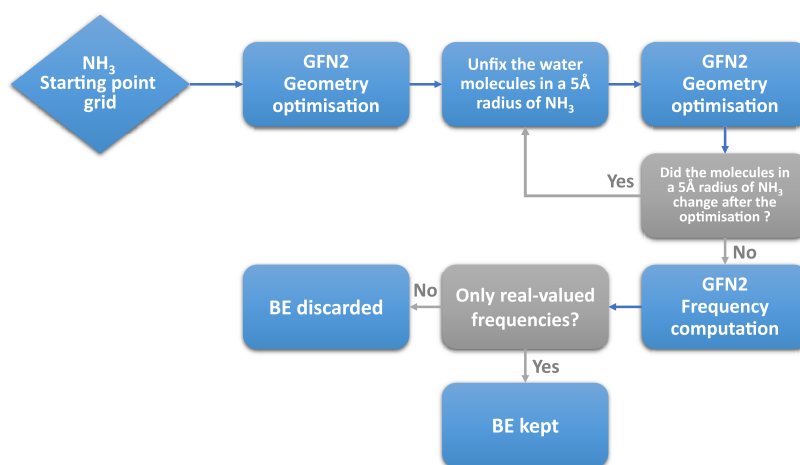


(a) Grain model

(b) Icosahedron vertexes grid

(c) Projected grid

**Figure 2.** Van der Waals representation of the icy grain (section a); starting grid positions (as blue dots, section b) and final projected grid for the location of each  $\text{NH}_3$  molecule (JSmol views can be seen by scanning the overlapping QR code).



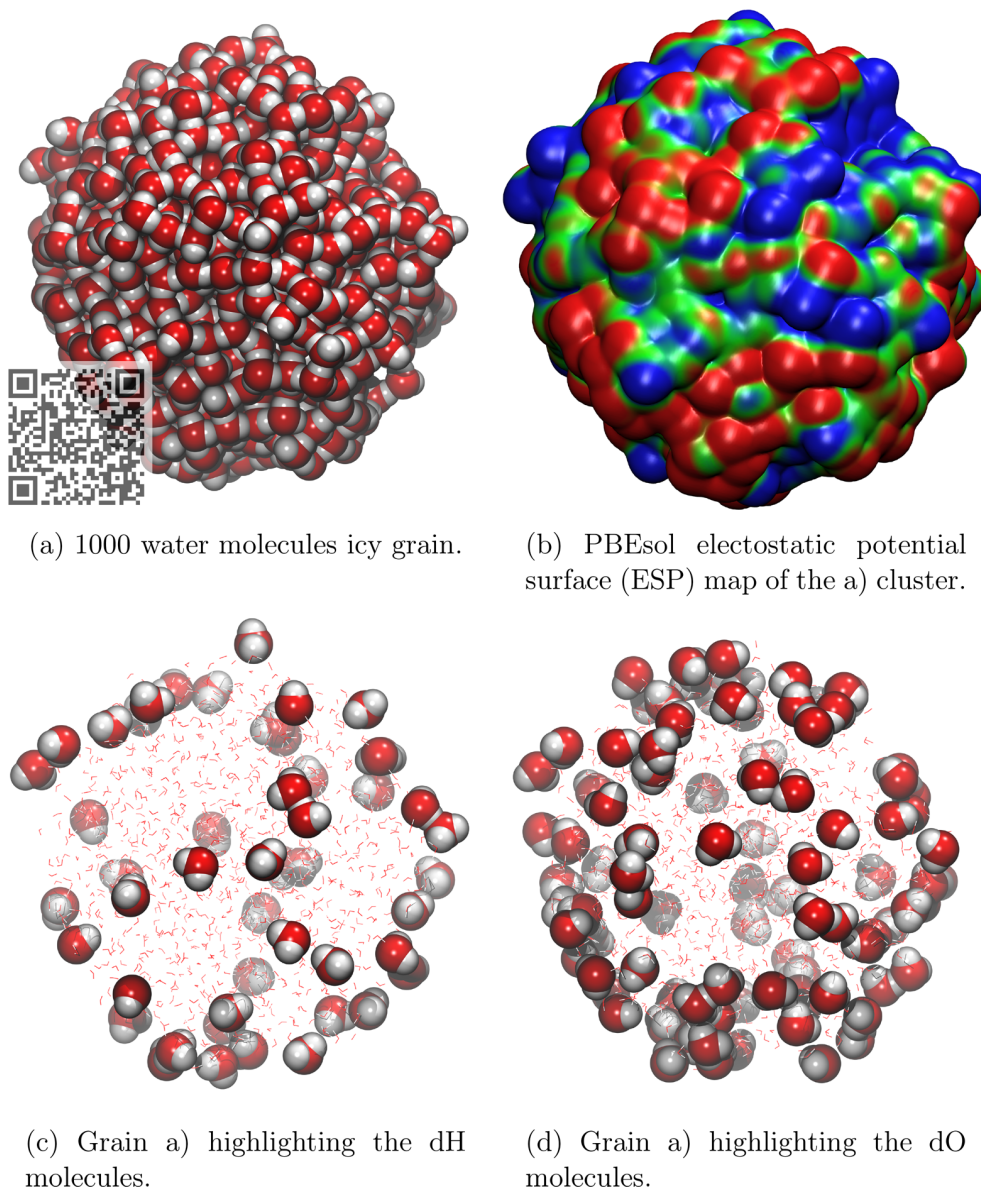
**Figure 3.** Flowchart of the  $\text{NH}_3$  BE distribution sampling.

Once the starting positions of  $\text{NH}_3$  have been defined, the method to compute the BE consists of two geometry optimization steps using GFN2 (see Figure 3). At the first step, only the coordinates of  $\text{NH}_3$  are allowed to relax, while the coordinates of all the water molecules composing the grain are held fixed at their starting positions. This strategy ensures that, despite the relatively small grain size, the long-range structural rigidity of the real grain is enforced by limiting the mobility of the model. This strategy ensures also a unique reference for the free ice cluster, in agreement with what would happen for a very large and more realistic cluster. In a second step, in line with the previous strategy and considering the interaction of  $\text{NH}_3$  as a local one, we relaxed, together with that of  $\text{NH}_3$ , the coordinates of the water molecules within a 5-Å radius from the  $\text{NH}_3$  location. This last relaxation process can, in principle, displace the  $\text{NH}_3$  molecule from the starting position and, therefore, reaching a position in which new interactions with water molecules not included in the original relaxation sphere have to be considered. To account for these cases, we redefined a new sphere of 5 Å around  $\text{NH}_3$ ; if the number of water molecules included in the new sphere changes with respect to that included in the original sphere a new GFN2 optimization is carried out. This process is repeated until the number of water molecules included in the sphere does no longer change during the optimization procedure. One

possible drawback of this procedure is that the  $\text{NH}_3$  molecule starts to travel on the grain surface at each new defined sphere, moving far apart from the original position. To mitigate this problem, we discarded situations in which the sphere of neighbors has to be redefined more than twice. Fortunately, we observed only 21 of such cases. Harmonic frequency calculations were run on the optimized structures to ensure that the final structures were indeed a minimum on the constrained potential energy surface (PES). All the structures giving one or more imaginary frequencies were rejected from the BE distribution.

## RESULTS

**Icy Grain Model.** The above-described procedure was carried out to build up a cluster up to 1000 water molecules. We extracted selected structures characterizing, as a function of the cluster size, the following features: H-bond length distribution (Figure 5), number of H-bonds per water molecule (Figure 6, left) and ratio of the number of dangling oxygens (dO) over the number of dangling hydrogens (dH) as a function of the inverse of the grain sizes in water molecules (Figure 6, right). Figure 4 shows the structure (in the van der Waals representation), the electrostatic potential (ESP) map of the 1000 water grain, and the vdW representation of the dangling species dH and dO (Figure S4 of the Supporting



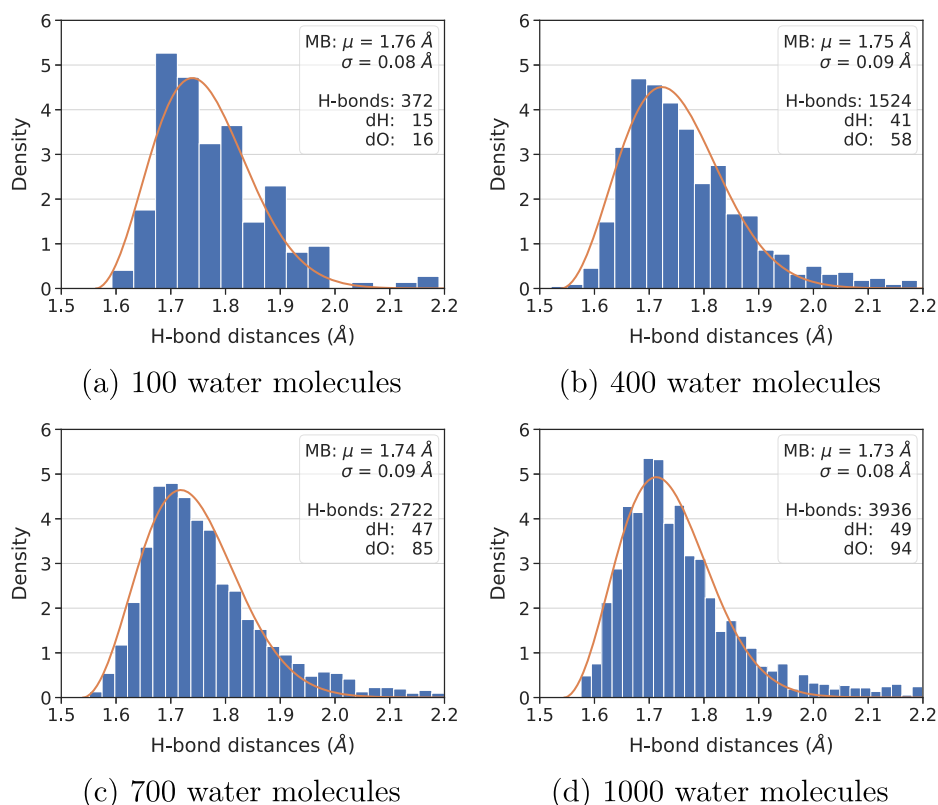
**Figure 4.** Different representation of a 1000 H<sub>2</sub>O molecules grain highlighting: (a) the surface sites structure (a JSmol view can be seen by scanning the overlapping QR code); (b) the electrostatic potential (blue (positive)/red (negative) zones); (c and d) the dH (dangling hydrogen) and dO (dangling oxygen) sites.

Information highlights multiple amorphous surface structures of the grain). The ESP has been computed as a single point calculation at PBEsol<sup>50</sup> using the parallel version of CRYSTAL17.<sup>51</sup>

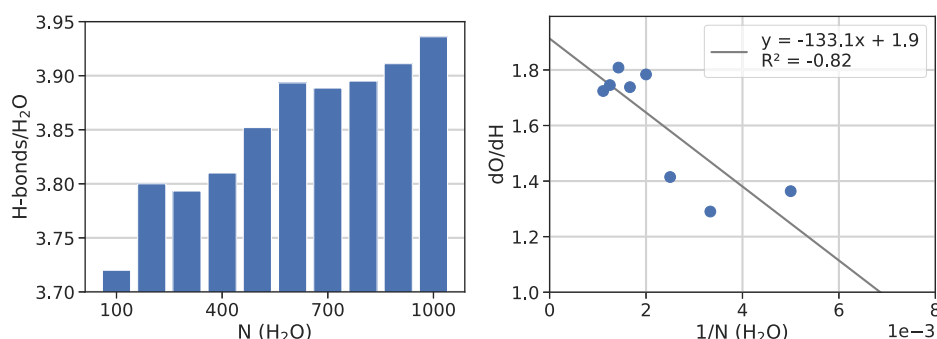
The H-bond length distribution for different grain sizes is nicely fitted by a Maxwell–Boltzmann probability density function (Figure 5). The feature of the distributions are all very similar as a function of the grain size, with the average H-bond length in the 1.73–1.77 Å interval, showing rapid convergence of the H-bond features with the cluster size.

The average of the total number of H-bonds per H<sub>2</sub>O molecule starts at 3.72 (100 water molecules) up to 3.94 (1000 water molecules) (Figure 6, left), indicating that almost all water molecules of the grain behave as a regular water in bulk ice (four H-bonds per water molecule). Obviously, water molecules at the grain surface contribute to lower the full H-bond coordination as both dangling hydrogen (dH) and

oxygen (dO) are present at the grain surface. Both dH and dO sites are extremely important for the adsorption of H-bond acceptor (like NH<sub>3</sub>) and H-bond donor molecules (like CH<sub>3</sub>OH), respectively. The ESP map of Figure 4 highlights the differences in terms of the electrostatic potential associated with the two dangling species. The total number of dangling species increases with the cluster size (see Figure 5), as well as the dO/dH ratio to reach 1.9 at infinite cluster size (see Figure 6, right). Interestingly, the number of dO is almost twice as large as that of dH, which is reflected by the ESP map, whose surface we estimated roughly to be occupied at 46% by negative (red-dO) zones, 38% by positive (blue-dH) zones, and the remaining 13% by almost neutral (green) zones. This distribution has important consequences on the capability of the ice grain to preferentially adsorb molecules with a H-bond donor character with respect to molecules with an H-bond acceptor character.



**Figure 5.** H-bond length (O...H) distribution of four icy grains of different size. Top right inset to each plot: average H-bond length  $\mu$  and its standard deviation  $\sigma$  of the Maxwell-Boltzmann distribution fit; sum of the number of H-bonds given and received for each  $\text{H}_2\text{O}$  molecule; number of dH (dangling hydrogen) and dO (dangling oxygen) sites.

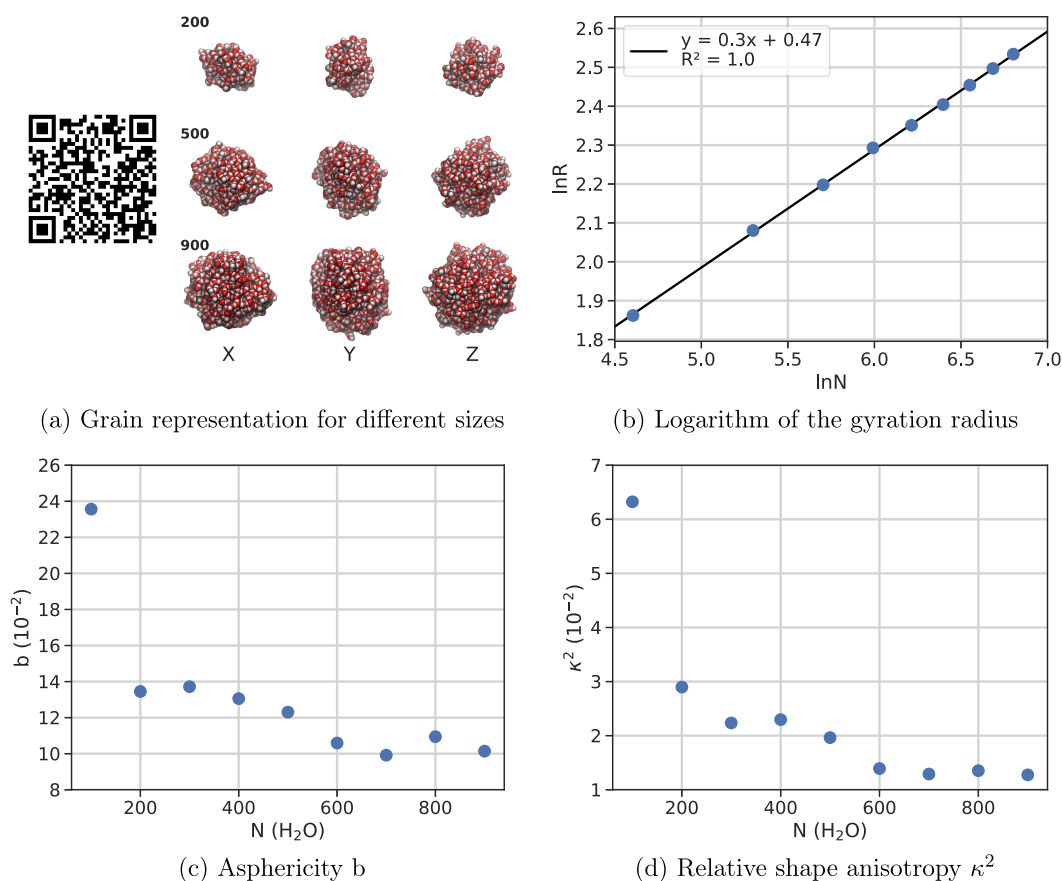


**Figure 6.** Number of H-bonds received and given per water molecule for different grain sizes (left). Ratio of dangling oxygens (dO) and hydrogens (dH) for the different grain sizes (right).

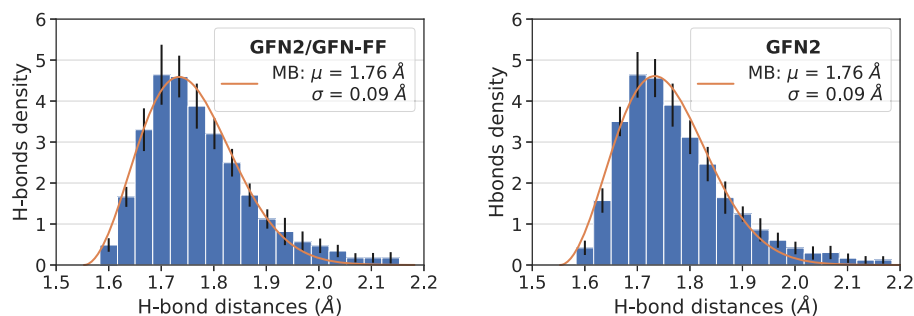
The density  $\rho$  of the  $N = 200$ , 500, and 900 grains was computed to be 1.12, 1.21, and 1.25  $\text{gr}/\text{cm}^3$ , respectively, with an error of  $\pm 0.03 \text{ gr}/\text{cm}^3$ . This shows an increased density of the grain with respect to that of the crystalline ice of  $\rho = 0.92 \text{ gr}/\text{cm}^3$ , in agreement with the amorphous (liquidlike) character of the simulated grains. However, in absence of a GFN2 calculation of a crystalline ice (at the moment periodic boundary conditions calculations are not yet implemented in the xTB code), some uncertainty in the  $\rho$  absolute values due to the systematic errors of GFN2 is expected.

The shape of the derived grains has been characterized by the gyration tensors and related quantities (see the [Computational Details](#)). [Figure 7b](#) shows a very nice linear correlation between the natural logarithms of the gyration radius  $R$  and that of the size  $N$  of the grain. The slope close to  $1/3$  can be

justified by considering that the volume of the grain associated with the gyration radius is  $V = \frac{4}{3}\pi R^3 \approx v_w N$ , in which  $v_w$  is the gyration radius associated with a single water molecule. By taking the natural logarithm of both sides, one gets the dependence of  $\frac{1}{3} \ln(N)$ , close to the best fit value. The graph warns that, if one doubles the  $R$  value of a 1000 water cluster (from 13 to 26  $\text{\AA}$ ), the number of needed water molecules increases by 1 order of magnitude, implying a steep increase of the computational resources needed to simulate grains of realistic size. Both the asphericity ([Figure 7c](#)) and the relative shape anisotropy ([Figure 7d](#)) data show a rapid convergence toward very small values, reminiscent of the increased spherical shape of the grain with the increased cluster size as visually shown in [Figure 7a](#).



**Figure 7.** Different shape descriptors from the gyration tensor for different grain sizes (JSmol views can be seen by scanning the QR code). In part a, X, Y, and Z represent the view axis facing the screen.



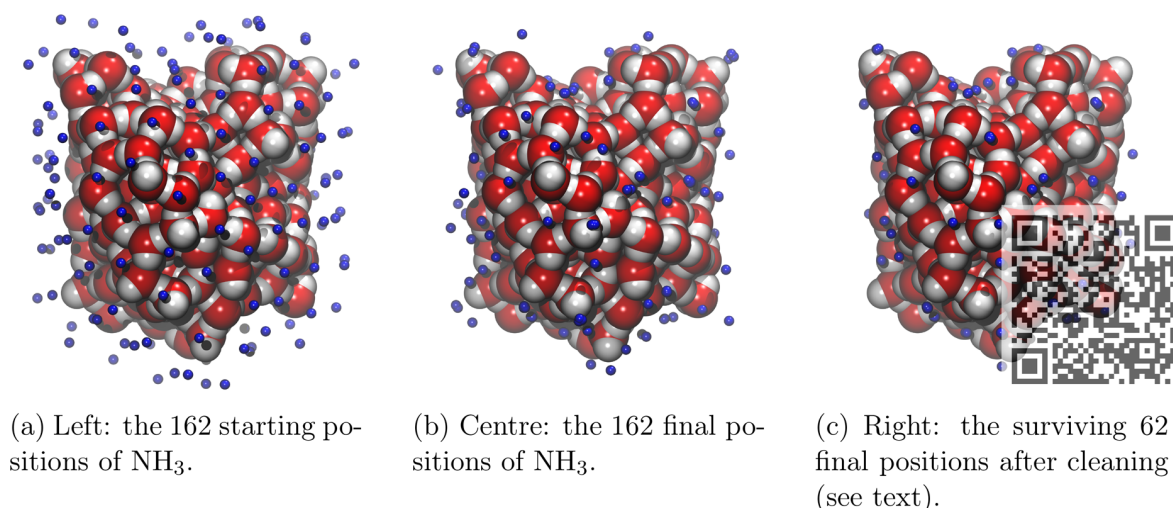
**Figure 8.** Average H-bond length for the 20<sup>th</sup> water clusters made up of 200 water molecules. The building process was carried out at both GFN2/GFN-FF (left) and at full GFN2 (right) levels. Black lines represent the standard deviation, and the orange line represents the Maxwell–Boltzmann fit.

Due to the random positioning of each added water during the grain accretion, different starting random seeds used to initiate the pseudorandom numbers will lead to different grains morphology. To explore such a dependence, we run 20 models with the same final composition of 200 water molecules using different initial random seeds. Averaging the resulting structural features for all considered grains gives  $379 \pm 4.4$  number of H-bonds,  $21.4 \pm 3.1$  dH, and  $30.6 \pm 2.8$  dO. We also computed the H-bond distances (H $\cdots$ O) and averaged them to obtain a distribution over the 20 grains, for each icy grain model (Figure 8, left). The small standard deviation of all these properties, as well as the similarity of the H-bond distribution, make us confident that the grain models should have the same characteristics. Therefore, we expect similar BE

distribution for grains obtained with different initial random seeds. As Figure 8 (right panel) shows, there is almost no difference by building up the cluster with the faster strategy of adopting a GFN-FF first optimization followed by a GFN2 final optimization step with respect to a more expensive full GFN2 optimization, giving credit to the adopted building up strategy. As already pointed out, this strategy allows a huge saving of computer resources (see Table 1).

**Binding Energy Sampling.** Binding energies of NH<sub>3</sub> molecule on a 200 water molecules grain (a vdW representation of the grain and an EPS map are shown on Figure S5 of the Supporting Information) were computed using the method depicted in Figure 3. Ideally, the largest cluster including up to 1000 water molecules would be a better

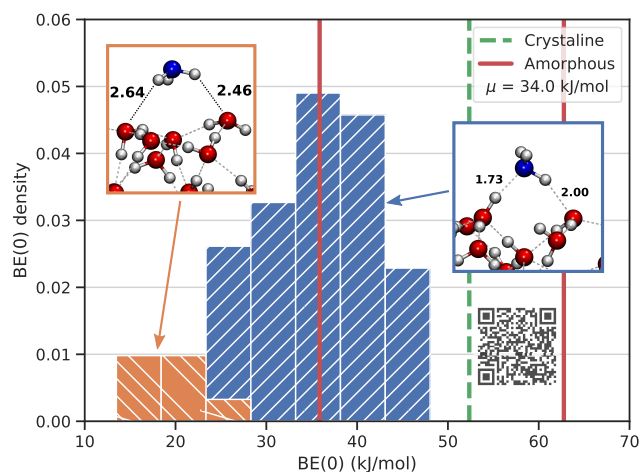




**Figure 9.** Positions of the  $\text{NH}_3$  molecules at different stages of the BE calculations (JSmol views can be seen by scanning the overlapping QR code of part c).

option than the one with 200 water molecules. For instance, Figure 6 shows that the  $\text{dO/dH}$  ratio for the 200 water molecules cluster is underestimated with respect to the largest one. Nevertheless, we will show later (see Figure 11) that clusters with the same size but different  $\text{dO/dH}$  ratio are giving  $\text{NH}_3$  BE distributions which are very close to each other. Considering that one BE evaluation on the largest cluster would take a full day long to accomplish and that more than 100 different cases should be run for the BE distribution, the adoption of the 200 water molecules was a reasonable trade-off between the computational cost and the number of available adsorption sites. This was also enough to ensure a representative statistical sampling of  $\text{NH}_3$  at the surface sites. Clearly, future work is needed to assess the variability of the BE distribution as a function of the ice cluster size.

On each sample, several checks were performed to discard those showing the following features: (i) presence of imaginary frequencies; (ii) nonconservation of the number of water molecules within the 5 Å-radius for more than two runs during the optimization procedure (see the Computational Details); (iii) duplicated structures. Of the 115 samples left after criteria i and ii, 31 land to a minimum which is present only once in the BE distribution, while the remaining 84 samples collapse to the same minimum of at least one other structure. More specifically, 19 structures present two twin minima, 7 three, 3 four, 1 six, and 1 seven (see Figure 9). Therefore, a total number of 62 binding sites were finally selected in the computed BE distribution of Figure 10. In the Supporting Information we discussed shortly the changes between the fixed and partially unfixed final structures and their respective BE distribution. It may be possible that more binding sites could emerge by starting with a tighter grid for the  $\text{NH}_3$  positioning, but it is unlikely that the number increases significantly. As one can see from Figure 10,  $\text{NH}_3$  acts as both H-bond donor and acceptor; compared to  $\text{H}_2\text{O}$ ,  $\text{NH}_3$  is a much stronger H-bond acceptor but a weaker H-bond donor. The most stable interactions are those where  $\text{NH}_3$  is both a H-bond donor and acceptor, while the least stable ones exhibit  $\text{NH}_3$  as exclusively H-bond donor, as represented by the orange histograms of Figure 10. With the approach used throughout the paper, also weakly bounded systems are taken into account, because of the starting random orientation of



**Figure 10.**  $\text{NH}_3$  zero point energy corrected BE distribution (JSmol representative structures can be seen by scanning the overlapping QR code).

$\text{NH}_3$ . These cases are particularly interesting for the potential chemical reactions occurring in the ISM, because the weaker the BE of the adsorbate is with the surface, the higher its probability to diffuse and therefore to react with other neighboring species adsorbed at the grain surfaces is.

In Figure 10 we also report the BE values of  $\text{NH}_3$  computed by Ferrero et al.<sup>21</sup> on periodic models of either crystalline (dashed line) and amorphous (full lines) ices. Crystalline ice presents only one BE because the  $\text{dO}$  and  $\text{dH}$  distribution is controlled by the imposed crystal symmetry and, accordingly, only one adsorption site is found for the  $\text{NH}_3$  case. However, crystalline ice is not the dominant water ice phase, even though it was found in non-negligible abundance in protoplanetary disks<sup>52</sup> and in protostellar molecular shocks.<sup>53</sup> Therefore, crystalline ice does not represent a realistic model of a typical grain ice mantle, usually of amorphous nature. The periodic amorphous model developed by Ferrero et al.<sup>21</sup> is large enough to identify different sites, but still too small to allow the building up of a sensible BE distribution. The BE range reported by Ferrero et al. falls at higher values with respect to ours. We suspect this is due to two reasons: (i) first, the

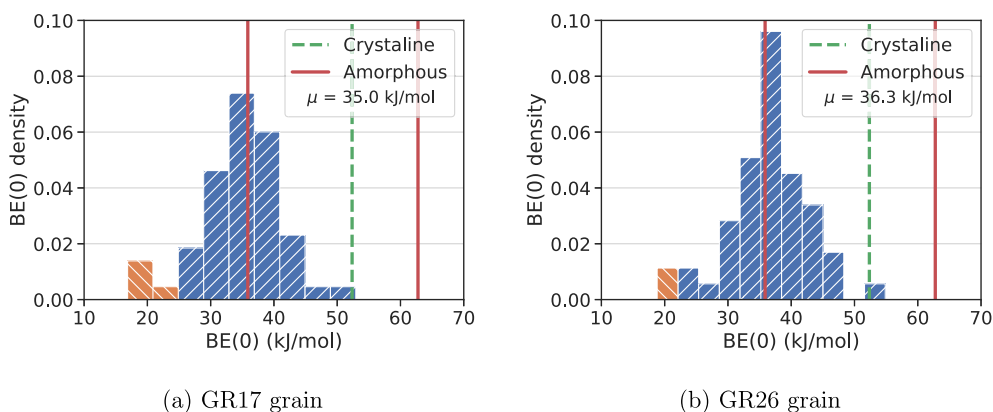


Figure 11.  $\text{NH}_3$   $\text{BE}(0)$  distributions for GR17 and GR26 clusters grains.

rigidity of our cluster compared to the full relaxation allowed by the periodic amorphous slab; (ii) second, we do not bias the positions of the  $\text{NH}_3$  molecule, while Ferrero et al. set up its position manually by maximizing the chance of a strong H-bond interactions.

To obtain a more reliable BE distribution, the adsorption of  $\text{NH}_3$  was studied also on two different grains among those generated with the different random seeds. The new grains have been chosen according to the largest difference in the number of dH, while the dO remains almost the same. The chosen GR17 and GR26 models identify grains with  $\text{dO}/\text{dH}(\text{GR17}) = 30/17 = 1.76$  and  $\text{dO}/\text{dH}(\text{GR26}) = 32/26 = 1.23$ , respectively. Figure 11 shows the  $\text{NH}_3$  BE distribution for these two grains. Both grains present an identical number of adsorption sites, which, after proper cleaning (*vide supra*), is 54, allowing for a fair comparison. We expect that GR26, with a higher dH, will give higher BEs due to the nature of  $\text{NH}_3$  as H-acceptor molecule. However, this is not the case: the average BE on the two grains is very similar (35.0/36.3 kJ/mol for GR17/GR26), with corresponding BE minima/maxima at 16.9/18.8 kJ/mol and 52.9/54.9 kJ/mol. Despite the very modest increase in the BE for the GR26, all values are almost grain model independent, thus showing convergence of data on the final shape of the grain at a fixed size. A careful analysis of the BE distributions on Figure 11 reveals a difference in the shape at high BE values with an overrepresentation of values around 35 kJ/mol for GR26 compared to GR17, while the tail at low BE values is rather similar (orange part of the chart), in agreement with the excess of dH of GR26 over GR17.

One important issue is to correct the electronic BE for the zero point energy (ZPE) contribution (see Computational Details), the most important correction to the BE at the very low temperature of the interstellar medium. However, the calculation of the Hessian matrix (second derivative of the GFN2 energy with respect to each nuclear Cartesian coordinate) is very expensive and it would be useful to find a way to skip this step. For instance, while the Hessian GFN2 calculation for a 200 water grain takes half an hour, about 6 days are needed to compute the 800 water grain (time obtained on the same hardware reported in Table 1). Here, we propose a simplified method to avoid this expensive step, by using our recent work<sup>40</sup> in which we compared the GFN2 BE with the *ab initio* ones by Ferrero et al.<sup>21</sup> for a large set of molecules adsorbed on crystalline models of ice. There, we also compared the GFN2 BE against the ZPE  $\text{BE}(0)$  corrected one, for each considered molecules, including  $\text{NH}_3$ . Figure 12

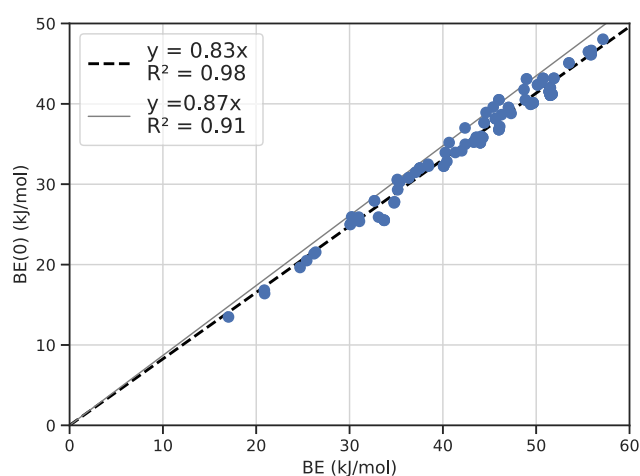


Figure 12. Linear correlation between BE and zero point energy corrected  $\text{BE}(0)$  values (kJ/mol) at GFN2 level from a previous work<sup>40</sup> (continuous line) and for all considered cases of  $\text{NH}_3$  adsorption at the 200 water grain of the present work (best fitted dashed line on filled circle points).

shows the best linear correlation (continuous line) using our previous data contrasted with the present GFN2 BE and  $\text{BE}(0)$  ones (dashes) for all considered  $\text{NH}_3$  adsorption cases at the icy grain. The similarity between the two linear fit is striking, and we proposed to adopt the same scaling obtained in our previous work also for the grain, without the need to compute the Hessian matrix. This approach will also be valid for other molecules to be considered in future work.

## CONCLUSION

In the present work, we proposed a new method for building up amorphous water icy cluster models apt to mimic interstellar icy grains. We adopted the semiempirical program GFN-xTB using the GFN2 semiempirical tight-binding Hamiltonian<sup>33</sup> and the newly developed GFN-FF<sup>34</sup> force field. A fully automated procedure based on in-house developed python ACO-FROST script is proposed, capable to build up an icy grain by a step by step addition of randomly oriented water molecules. We considered grains of 200 up to 1000 water molecules (with intermediary steps every 100 water molecules). A sequence of geometry relaxations followed by short molecular dynamics steps at 10 K were implemented to resemble the physicochemical of grain formation in the interstellar regions. Careful checks on the icy grain properties

(O...H bond length, dangling hydrogen (dH) and oxygen (dO), gyration tensor quantities (gyration radius, asphericity, and relative shape anisotropy), and surface electrostatic potential maps) were carried out to assess the physicochemical features of the model. We showed that the icy grain rapidly reaches a spherical shape, as shown by the gyration tensor, with the dO/dH value showing almost twice as dO than dH sites. This is an important result, implying that the grain is more apt to selectively adsorb hydrogen donor molecules with respect to hydrogen acceptors. Future works will be focused on studying how the dO/dH evolves when heating the grain at higher temperatures (from 10 to 100 K), as experimental works suggested an increasing dO/dH at higher temperature; i.e., the grain becomes deprived of dH sites. We showed that the average hydrogen bond features rapidly reach a common value, as for example, the average number of hydrogen bonds and the above-described dO/dH ratio. As a possible application, we computed the binding energies of NH<sub>3</sub> at the surface sites of a 200 water grain. By extending the automated procedure adopted for building up the icy grain we set up about 160 different starting sites for the NH<sub>3</sub> adsorption. The procedure is capable of purging duplicated and ill-defined situations providing a physically sound set of adsorption structures. This approach allows to compute a distribution of binding energies, with a much better statistical significance than previous work by our group<sup>21</sup> and from other models proposed in the literature. The agreement between our data, even at the relatively cheap GFN2 level of theory, is encouraging when compared with the much expensive *ab initio* data from Ferrero et al.<sup>21</sup>

We believe that the present water cluster can be adopted as a general model to study not only the BE of molecules of interest but diffusion as well, at least with the purpose of rapidly characterizing the diffusion barriers to be refined by more accurate QM:MM calculation. In a forthcoming work, we have adopted and improved the present methodology to obtain a very accurate estimation of the NH<sub>3</sub> BE distribution with a QM:MM methodology capable to give results at the coupled cluster quality level. The subject and the results will be the content of a future work by our group.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsearthspacechem.2c00004>.

JSmol rendering of the building up process of a large icy grain and a file containing the GFN2 steps of the 1000 grain building process, the 40 grain structures used for the random seed study, and the 162 final structures of the NH<sub>3</sub> binding energies computation (a text file with the BE, BE(0),  $\Delta$ ZPE values, and if the binding energy values were included in the final BE(0) distribution of Figure 10 is also included) (ZIP)

Additional details on computational methods and results, including rendered images of obtained structures (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Piero Ugliengo – Dipartimento di Chimica and Nanostructured Interfaces and Surfaces (NIS) Centre,

Università degli Studi di Torino, 10125 Torino, Italy;

ORCID: [orcid.org/0000-0001-8886-9832](https://orcid.org/0000-0001-8886-9832);

Email: [piero.ugliengo@unito.it](mailto:piero.ugliengo@unito.it)

## Authors

Aurèle Germain – Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; ORCID: [orcid.org/0000-0001-7856-0516](https://orcid.org/0000-0001-7856-0516)

Lorenzo Tinacci – Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), CNRS, Université Grenoble Alpes, 38000 Grenoble, France; ORCID: [orcid.org/0000-0001-9909-9570](https://orcid.org/0000-0001-9909-9570)

Stefano Pantaleone – Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi di Perugia, 06123 Perugia, Italy; ORCID: [orcid.org/0000-0002-2457-1065](https://orcid.org/0000-0002-2457-1065)

Cecilia Ceccarelli – Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), CNRS, Université Grenoble Alpes, 38000 Grenoble, France; ORCID: [orcid.org/0000-0001-9664-6292](https://orcid.org/0000-0001-9664-6292)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsearthspacechem.2c00004>

## Notes

The authors declare no competing financial interest.

**Data and Software Availability.** The “ACO-FROST” set of scripts is open source and freely available on two github repositories named “ACO-FROST\_grain\_generator” ([https://github.com/aurelegermain/ACO-FROST\\_grain\\_generator/](https://github.com/aurelegermain/ACO-FROST_grain_generator/)) for the grain building process, and “ACO-FROST\_BE-grain” ([https://github.com/aurelegermain/ACO-FROST\\_BE-grain/](https://github.com/aurelegermain/ACO-FROST_BE-grain/)) for the binding energies computation. The GFN-xTB sets of computational methods are available on the github repository “xtb” (<https://github.com/grimme-lab/xtb>). All data described in the paper (structure coordinates and alike) are described and made available in the Supporting Information (see specific section for details).

## ■ ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 811312 for the project “Astro-Chemical Origins” (ACO). S.P. and P.U. acknowledge the Italian Space Agency for cofunding the Life in Space Project (ASI N. 2019-3-U.O). Support from the Italian MUR (PRIN 2020, Astrochemistry beyond the second period elements, Prot. 2020AFB3FX) is gratefully acknowledged. C.C. acknowledges the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, for the Project “The Dawn of Organic Chemistry” (DOC), Grant Agreement No. 741002.

## ■ REFERENCES

- (1) Jones, A. P. Heteroatom-doped hydrogenated amorphous carbons, a-C:H: X. *Astronomy & Astrophysics* **2013**, 555, A39.
- (2) Jones, A. P.; Köhler, M.; Ysard, N.; Bocchio, M.; Verstraete, L. The global dust modelling framework THEMIS. *Astron. Astrophys.* **2017**, 602, A46.
- (3) Boogert, A. C.; Gerakines, P. A.; Whittet, D. C. Observations of the icy universe. *Annual Review of Astronomy and Astrophysics* **2015**, 53, 541–581.



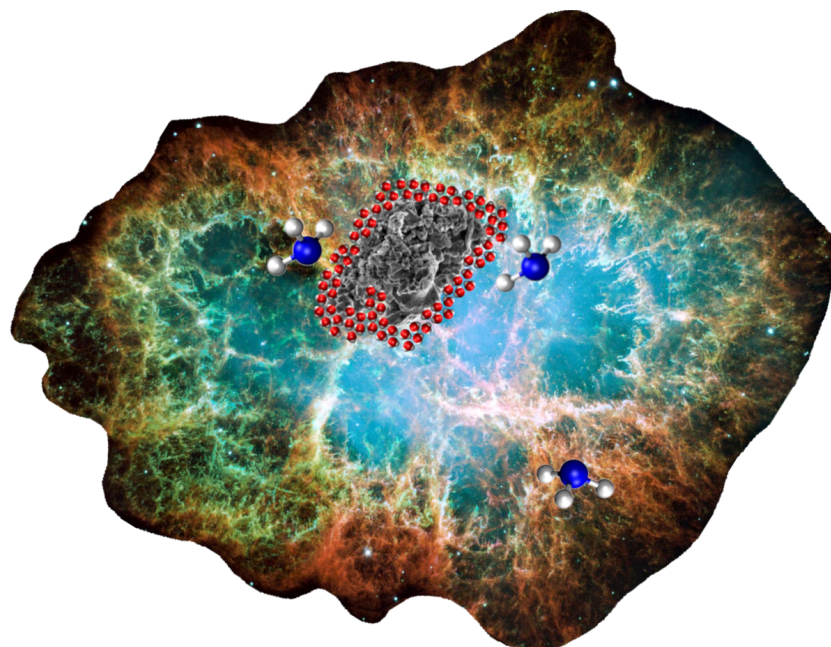
- (4) Oba, Y.; Miyauchi, N.; Hidaka, H.; Chigai, T.; Watanabe, N.; Kouchi, A. Formation of compact amorphous H<sub>2</sub>O ice by codeposition of hydrogen atoms with oxygen molecules on grain surfaces. *Astrophys. J.* **2009**, *701*, 464–470.
- (5) Watanabe, N.; Kouchi, A. Ice surface reactions: A key to chemical evolution in space. *Prog. Surf. Sci.* **2008**, *83*, 439–489.
- (6) Mayer, E.; Pletzer, R. Astrophysical implications of amorphous ice—a microporous solid. *Nature* **1986**, *319*, 298–301.
- (7) Oba, Y.; Watanabe, N.; Hama, T.; Kuwahata, K.; Hidaka, H.; Kouchi, A. Water formation through a quantum tunneling surface reaction, OH + H<sub>2</sub>, at 10 K. *Astrophys. J.* **2012**, *749*, 67.
- (8) Dulieu, F.; Amiaud, L.; Congiu, E.; Fillion, J. H.; Matar, E.; Momeni, A.; Pirronello, V.; Lemaire, J. L. Experimental evidence for water formation on interstellar dust grains by hydrogen and oxygen atoms. *Astron. Astrophys.* **2010**, *512*, A30.
- (9) Hiraoka, K.; Miyagoshi, T.; Takayama, T.; Yamamoto, K.; Kihara, Y. Gas-Grain Processes for the Formation of CH<sub>4</sub> and H<sub>2</sub>O: Reactions of H Atoms with C, O, and CO in the Solid Phase at 12 K. *Astrophysical Journal* **1998**, *498*, 710–715.
- (10) Redondo, P.; Pauzat, F.; Ellinger, Y.; Markovits, A. Reconstruction of water ice: the neglected process OH + OH → H<sub>2</sub>O + O. *Astronomy & Astrophysics* **2020**, *638*, A125.
- (11) Redondo, P.; Pauzat, F.; Markovits, A.; Ellinger, Y. Revisiting the OH + H<sub>2</sub> → H<sub>2</sub>O + H reaction at the molecular level: The plausible catalytic role of ice in its own reconstruction. *Astron. Astrophys.* **2021**, *646*, A163.
- (12) Vidali, G. H<sub>2</sub> Formation on Interstellar Grains. *Chem. Rev.* **2013**, *113*, 8762–8782.
- (13) Hollenbach, D.; Salpeter, E. E. Surface Recombination of Hydrogen Molecules. *Astrophysical Journal* **1971**, *163*, 155.
- (14) Zamirri, L.; Ugliengo, P.; Ceccarelli, C.; Rimola, A. Quantum Mechanical Investigations on the Formation of Complex Organic Molecules on Interstellar Ice Mantles. Review and Perspectives. *ACS Earth and Space Chemistry* **2019**, *3*, 1499–1523.
- (15) Rothard, H.; Domaracka, A.; Boduch, P.; Palumbo, M. E.; Strazzulla, G.; da Silveira, E. F.; Dartois, E. Modification of ices by cosmic rays and solar wind. *Journal of Physics B: Atomic, Molecular and Optical Physics* **2017**, *50*, 062001.
- (16) Wakelam, V.; Loison, J. C.; Mereau, R.; Ruaud, M. Binding energies: New values and impact on the efficiency of chemical desorption. *Molecular Astrophysics* **2017**, *6*, 22–35.
- (17) Das, A.; Sil, M.; Gorai, P.; Chakrabarti, S. K.; Loison, J. C. An Approach to Estimate the Binding Energy of Interstellar Species. *Astrophysical Journal Supplement Series* **2018**, *237*, 9.
- (18) Garrod, R. T. Three-dimensional, off-lattice Monte Carlo kinetics simulations of interstellar grain chemistry and ice structure. *Astrophys. J.* **2013**, *778*, 158.
- (19) Christianson, D. A.; Garrod, R. T. Chemical Kinetics Simulations of Ice Chemistry on Porous Versus Non-Porous Dust Grains. *Frontiers in Astronomy and Space Sciences* **2021**, *8*, 1–26.
- (20) Zamirri, L.; Casassa, S.; Rimola, A.; Segado-Centellas, M.; Ceccarelli, C.; Ugliengo, P. IR spectral fingerprint of carbon monoxide in interstellar water-ice models. *Mon. Not. R. Astron. Soc.* **2018**, *480*, 1427–1444.
- (21) Ferrero, S.; Zamirri, L.; Ceccarelli, C.; Witzel, A.; Rimola, A.; Ugliengo, P. Binding energies of interstellar molecules on crystalline and amorphous models of water ice by ab-initio calculations. *Astrophysical Journal* **2020**, *904*, 11.
- (22) Sameera, W. M. C.; Senevirathne, B.; Andersson, S.; Al-Ibadi, M.; Hidaka, H.; Kouchi, A.; Nyman, G.; Watanabe, N. CH<sub>3</sub>O Radical Binding on Hexagonal Water Ice and Amorphous Solid Water. *J. Phys. Chem. A* **2021**, *125*, 387–393.
- (23) 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.
- (24) Rimola, A.; Taquet, V.; Ugliengo, P.; Balucani, N.; Ceccarelli, C. Combined quantum chemical and modeling study of CO hydrogenation on water ice. *Astronomy & Astrophysics* **2014**, *572*, A70.
- (25) Duvernay, F.; Rimola, A.; Theule, P.; Danger, G.; Sanchez, T.; Chiavassa, T. Formaldehyde chemistry in cometary ices: the case of HOCH<sub>2</sub>OH formation. *Phys. Chem. Chem. Phys.* **2014**, *16*, 24200–24208.
- (26) Enrique-Romero, J.; Rimola, A.; Ceccarelli, C.; Balucani, N. The (impossible?) formation of acetaldehyde on the grain surfaces: Insights from quantum chemical calculations. *Monthly Notices of the Royal Astronomical Society: Letters* **2016**, *459*, L6–L10.
- (27) Bovolenta, G.; Bovino, S.; Vöhringer-Martinez, E.; Saez, D. A.; Grassi, T.; Vogt-Geisse, S. High level ab initio binding energy distribution of molecules on interstellar ices: Hydrogen fluoride. *Molecular Astrophysics* **2020**, *21*, 100095.
- (28) Song, L.; Kästner, J. Formation of the prebiotic molecule NH<sub>2</sub>CHO on astronomical amorphous solid water surfaces: accurate tunneling rate calculations. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29278–29285.
- (29) Lamberts, T.; Kästner, J. Influence of Surface and Bulk Water Ice on the Reactivity of a Water-forming Reaction. *Astrophysical Journal* **2017**, *846*, 43.
- (30) Molpeceres, G.; Zaverkin, V.; Kästner, J. Neural-network assisted study of nitrogen atom dynamics on amorphous solid water – I. adsorption and desorption. *Mon. Not. R. Astron. Soc.* **2020**, *499*, 1373–1384.
- (31) Molpeceres, G.; Kästner, J. Adsorption of H<sub>2</sub> on amorphous solid water studied with molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7552–7563.
- (32) Cuppen, H. M.; Fredon, A.; Lamberts, T.; Penteado, E. M.; Simons, M.; Walsh, C. Surface astrochemistry: a computational chemistry perspective. *Proceedings of the International Astronomical Union* **2017**, *13*, 293–304.
- (33) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (34) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angewandte Chemie - International Edition* **2020**, *59*, 15665–15673.
- (35) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355–2368.
- (36) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.* **2020**, *16*, 6305–6316.
- (37) Kříž, K.; Nováček, M.; Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.* **2021**, *17*, 1548–1561.
- (38) Spicher, S.; Caldeweyher, E.; Hansen, A.; Grimme, S. Benchmarking London dispersion corrected density functional theory for noncovalent ion-π interactions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 11635–11648.
- (39) Germain, A.; Ugliengo, P. Modeling Interstellar Amorphous Solid Water Grains by Tight-Binding Based Methods: Comparison Between GFN-XTB and CCSD(T) Results for Water Clusters. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 2020; Vol. 12253, pp 745–753;
- (40) Germain, A.; Corno, M.; Ugliengo, P. Computing Binding Energies of Interstellar Molecules by Semiempirical Quantum Methods: Comparison Between DFT and GFN2 on Crystalline Ice. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2021; Vol. 12953, pp 632–645.
- (41) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2021**, *11*, No. e1493.



- (42) Theodorou, D. N.; Suter, U. W. Shape of unperturbed linear polymers: polypropylene. *Macromolecules* **1985**, *18*, 1206–1214.
- (43) Ugliengo, P.; Viterbo, D.; Chiari, G. MOLDRAW: Molecular Graphics on a Personal Computer. *Zeitschrift Fur Kristallographie* **1993**, *207*, 9–23.
- (44) Harvey, A. H. In *CRC Handbook of Chemistry and Physics*, 97th ed.; Haynes, W. M., Lide, D. R., Bruno, T. J., Eds.; CRC Press: Boca Raton, FL, 2017.
- (45) Hanson, R. M. *Jmol* – a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* **2010**, *43*, 1250–1260.
- (46) Hama, T.; Watanabe, N. Surface Processes on Interstellar Amorphous Solid Water: Adsorption, Diffusion, Tunneling Reactions, and Nuclear-Spin Conversion. *Chem. Rev.* **2013**, *113*, 8783–8839.
- (47) Pantaleone, S.; Enrique-Romero, J.; Ceccarelli, C.; Ferrero, S.; Balucani, N.; Rimola, A.; Ugliengo, P. H<sub>2</sub> Formation on Interstellar Grains and the Fate of Reaction Energy. *Astrophysical Journal* **2021**, *917*, 49.
- (48) Pantaleone, S.; Enrique-Romero, J.; Ceccarelli, C.; Ugliengo, P.; Balucani, N.; Rimola, A. Chemical Desorption versus Energy Dissipation: Insights from Ab Initio Molecular Dynamics of HCO Formation. *Astrophysical Journal* **2020**, *897*, 56.
- (49) Teanby, N. A. An icosahedron-based method for even binning of globally distributed remote sensing data. *Computers and Geosciences* **2006**, *32*, 1442–1450.
- (50) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100*, 136406.
- (51) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-mechanical condensed matter simulations with CRYSTAL. *WIREs Computational Molecular Science* **2018**, *8*, No. e1360.
- (52) Terada, H.; Tokunaga, A. T. Discovery of crystallized water ice in a silhouette disk in the M43 region. *Astrophysical Journal* **2012**, *753*, 19.
- (53) Molinari, S.; Ceccarelli, C.; White, G. J.; Saraceno, P.; Nisini, B.; Giannini, T.; Caux, E. Detection of the 62 micron crystalline H<sub>2</sub>O ice feature in emission toward HH 7 with the Infrared Space Observatory long-wavelength spectrometer. *Astrophysical Journal Letters* **1999**, *521*, L71.

# Theoretical distribution of the ammonia binding energy at interstellar icy grains: a new computational framework

The hereafter article has been published in ACS Earth Space Chemistry 2022, 6, 6, 1514–1526. A cover image for the aforementioned article has been selected as a supplementary cover for ACS Earth and Space Chemistry in the same issue, it will be presented after the article in Section 8.1.



# Theoretical Distribution of the Ammonia Binding Energy at Interstellar Icy Grains: A New Computational Framework

Lorenzo Tinacci, Auréle Germain, Stefano Pantaleone, Stefano Ferrero, Cecilia Ceccarelli, and Piero Ugliengo\*



Cite This: <https://doi.org/10.1021/acsearthspacechem.2c00040>



Read Online

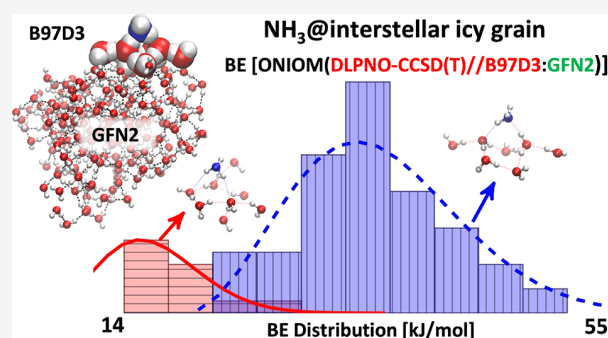
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** The binding energies (BE) of molecules on the interstellar grains are crucial in the chemical evolution of the interstellar medium (ISM). Both temperature-programmed desorption (TPD) laboratory experiments and quantum chemistry computations have often provided, so far, only single values of the BE for each molecule. This is a severe limitation, as the ices enveloping the grain mantles are structurally amorphous, giving rise to a manifold of possible adsorption sites, each with different BEs. However, the amorphous ice nature prevents the knowledge of structural details, hindering the development of a common accepted atomistic icy model. In this work, we propose a computational framework that closely mimics the formation of the interstellar grain mantle through a water by water accretion. On that grain, an unbiased random (but well reproducible) positioning of the studied molecule is then carried out. Here we present the test case of  $\text{NH}_3$ , a ubiquitous species in the molecular ISM. We provide the BE distribution computed by a hierarchy approach, using the semiempirical  $x\text{TB-GFN2}$  as a low-level method to describe the whole icy cluster in combination with the B97D3 DFT functional as a high-level method on the local zone of the  $\text{NH}_3$  interaction. The final ZPE-corrected BE is computed at the ONIOM(DLPNO-CCSD(T)//B97D3: $x\text{TB-GFN2}$ ) level, ensuring the best cost/accuracy ratio. The main peak of the predicted  $\text{NH}_3$  BE distribution is in agreement with experimental TPD and computed data in the literature. A second broad peak at very low BE values is also present, which has never been detected before. It may provide the solution to a longstanding puzzle about the presence of gaseous  $\text{NH}_3$  also observed in cold ISM objects.



**KEYWORDS:** *amorphous water ice, xTB-GFN2, ONIOM, DLPNO, B97D3, NH<sub>3</sub> adsorption, NH<sub>3</sub> binding energy*

## INTRODUCTION

Interstellar dust grains in cold ( $\sim 10$  K) molecular clouds are made up of sub-micrometer-sized refractory cores (mainly silicates and carbonaceous material), on top of which water molecules are formed *in situ* through reactions involving hydrogen and oxygen.<sup>1–5</sup> Eventually, this process leads to the accretion of a thick (made up of more than 100 layers: e.g. Taquet et al.<sup>6</sup>) amorphous icy mantle. At the same time, other atoms and molecules formed in the gas phase can condense and be adsorbed onto the grain mantle, where they may diffuse and react on the icy surface, enriching the chemical composition of the grain mantle.

The vast majority of the species frozen or trapped on the grain mantles are only observable when they are released into the gas phase either in warm ( $\geq 100$  K) regions, such as hot cores/corinos via thermal desorption,<sup>7–9</sup> or in shocked regions, via sputtering of the mantles.<sup>10–12</sup> All of the processes mentioned above, adsorption and diffusion as well as desorption, are governed by a key parameter, the so-called binding energy (BE): namely, the strength of a species to remain glued to the surface.

Since BE has an exponential dependence on the expressions of astrochemical models that describe the above processes, its estimation with a good level of accuracy is crucial for our knowledge of chemical evolution of whatever species.<sup>13</sup> This fundamental piece of information can be obtained via either theoretical or experimental approaches.

Usually, estimates of PES via theoretical methods involve computations using a molecular mechanics force-field approach and/or rigorous quantum mechanical methods. In both cases, an atomistic model of the icy grain is needed and the BE is computed via a supermolecular approach: namely, computing the difference between the energy of the adsorbate interacting with the icy grain and the energies of the free adsorbate plus the

Received: February 5, 2022

Revised: May 19, 2022

Accepted: May 19, 2022

original icy grain. Despite this simple definition, the final BE value can be affected by many factors, from both modeling and methodological points of view. To start with, the computerized icy model is usually ill-defined, as the structure of the interstellar ice is poorly known. Therefore, a variety of models to simulate the ice-species adsorption has been proposed in the literature, from just a single water molecule up to periodic models of either crystalline or amorphous water ice.<sup>14–16</sup> Due to the difficulty of simulating the icy grain accretion by *in situ* water formation, all of the models so far have been constructed by assembling a variety of already formed water molecules interacting through hydrogen bonds.<sup>17,18</sup> This may have serious consequences on the final ice structure, as the fraction of water formation energy transferred to the grain can affect its final structural features much more than the mere hydrogen-bond interaction between the water molecules.<sup>19</sup> In addition, it has been shown theoretically and experimentally that any species does not have a single BE on amorphous water surfaces (AWS) but rather a distribution of BEs, which depends on the species and the surface.<sup>16,20–23</sup> Therefore, the icy grains should be large and varied enough to allow to reconstruct the BE distribution of a species and not just a value. To overcome the aforementioned problems, we have recently proposed<sup>24</sup> an automatic and unbiased approach to construct water-ice clusters and obtain the binding energy distribution of any species (see [Methodology](#) for further details).

In addition to the icy model definition, the second important issue to compute the BE is the adopted level of theory, which always represents a compromise between the computational accuracy and the computational cost (method and system size).

Methods based on molecular mechanics may reach some accuracy when they are designed to treat very specific cases but fail for cases outside their specific parametrization. Alternatively, methods based on the best level of quantum chemistry, such as the gold standard CCSD(T),<sup>25</sup> ensure a well-balanced treatment of all the relevant interactions responsible for the adsorption on the icy grain surface, irrespective of the considered adsorbate molecule. However, the computational time required by CCSD(T) grows too steeply to be applicable to large icy grains.

Here, we propose a new method that optimizes the computation accuracy on very large icy grain models. Specifically, we implemented an automatic procedure that is based on the ACO-FROST code, recently developed by our group,<sup>24</sup> to construct a large ( $\geq 1000$  water molecules) icy grain. Briefly, only a selected portion of the icy grain, where the adsorption takes place, is treated at a very high level of theory, while the whole cluster is treated at a lower level. This procedure itself is not completely new, as it has been already adopted in the field of surface science adsorption<sup>26</sup> and for some ice models.<sup>16,27–30</sup> In a recent work, we adopted a similar scheme to improve the BEs computed for a set of molecules on periodic ice models (both crystalline and amorphous) reaching CCSD(T)-quality results.<sup>16</sup> Similarly, Duflo et al.<sup>31</sup> adopted a QM:MM approach using for the QM method the DLPNO-CCSD(T) technique,<sup>32</sup> a very accurate and computationally feasible version of the CCSD(T) standard based on localized orbitals and the PM6 semiempirical method<sup>33</sup> for the rest of the system.

Our newly proposed procedure, described in this work, possesses the following novelties with respect to the above works:

- (i) an unbiased procedure to generate a large variety of adsorbed structures, not dependent on the nature of the

adsorbate molecule and the size of the icy cluster, which allows computation of a BE distribution of the considered species

- (ii) the low-level theory adopted to treat the whole icy cluster on the basis of the accurate semiempirical tight-binding xTB-GFN2 method, very recently developed by Grimme's group<sup>34</sup>
- (iii) the high-level theory adopted to describe the ice around the adsorbing site on the basis of the DLPNO-CCSD(T) method with a selection of large Gaussian basis sets

In addition, the procedure is carried out automatically by a package of Python scripts, which allow the construction, submission, and data extraction of the needed calculations.

The BE values resulting from the above approach should in principle be comparable to experimental derivations of BEs. However, this is not straightforward for the following reasons. Binding energies are usually experimentally derived via the so-called temperature-programmed desorption (TPD) method. Strictly speaking, this method provides the desorption activation energy (DAE), which is often interpreted as BE. In practice, the DAE is derived indirectly from the TPD peaks through Redhead method<sup>35</sup> or more sophisticated numerical techniques. In most TPD experiments, a water-ice surface hosts a monolayer of the adsorbate and, therefore, the BE also depends on the surface coverage.<sup>23</sup> This renders the comparison between DAE and the computed BE to be actually not straightforward.<sup>36</sup> For example, ice restructuring processes may affect the final DAE, making it different from the BE. Also, sometimes TPD experiments only provide desorption temperature peaks  $T_{\text{des}}$ , with no numerical estimate of the DAE. For instance, Collings et al.<sup>37</sup> computed the BE of a species  $X$  as  $\text{BE}(X) = [T_{\text{des}}(X)/T_{\text{des}}(\text{H}_2\text{O})] \times \text{BE}(\text{H}_2\text{O})$ , in which  $T_{\text{des}}(X)$  is the desorption temperature of the species  $X$  in contrast with that of water,  $T_{\text{des}}(\text{H}_2\text{O})$ , by assuming  $\text{BE}(\text{H}_2\text{O}) = 4800 \text{ K}$  ( $\sim 40 \text{ kJ/mol}$ ). For the above reasons, a one by one comparison between experiment and modeling should be carried out with extreme care, particularly when a BE distribution is computed, as in the present work.

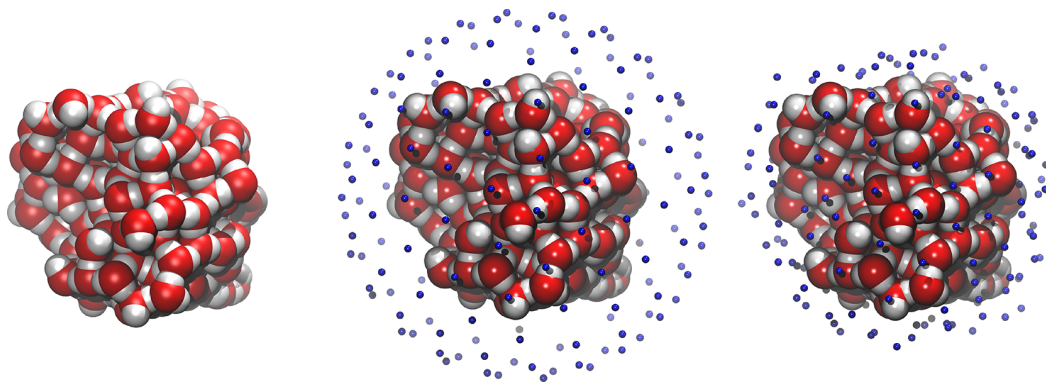
For our first application of the new method presented here, we chose the ammonia molecule, because it is a thoroughly studied and important species in the molecular ISM. It is the first detected interstellar polyatomic molecule<sup>38</sup> and one of the most observed, ubiquitous, and studied. It is found in a gaseous form toward the Galactic Center warm molecular clouds and cores,<sup>38,39</sup> diffuse clouds,<sup>40</sup> massive hot cores,<sup>41</sup> molecular outflows,<sup>42</sup> solar-type protostars,<sup>43</sup> cold molecular clouds,<sup>44</sup> prestellar cores,<sup>45</sup> and protoplanetary disks.<sup>46</sup> Ammonia is also observed to be quite abundant in the icy mantles that envelope the interstellar dust grains in cold regions.<sup>47</sup> Obviously, whether ammonia is in either gaseous or solid forms is governed by its BE. Along the same vein, understanding the ammonia chemistry requires a good knowledge of the ammonia BE and, more specifically, its BE distribution, which is the focus of this work.

## ■ METHODOLOGY

**Icy Grain Model and  $\text{NH}_3$  Binding Site Sampling.** The water-ice grain model used throughout this work, the binding energy sampling procedure, and the preliminary BE optimized structures were taken from a previous work by our group, which is summarized in this section.<sup>24</sup>

**Water-Ice Grain Model.** In order to build up the grain model, a bottom-up approach was followed: i.e., by random successive aggregations of water molecules. A geometry optimization was





**Figure 1.** (left) Icy grain model, (center) model with overlapped 162 vertices grid points shown in blue, and (right) model with the same vertices projected closer (2.5–3 Å) to the grain surface. Atom color legend: oxygen in red, hydrogen in white. Data were taken from ref 24.

performed at each addition of a water molecule, followed by a short molecular dynamics (MD) run at 10 K every 10 added H<sub>2</sub>O molecules, to mimic the induced thermal motion due to the partially transferred energy of water formation<sup>19</sup> occurring in the real grain but not taken into account here (*vide supra*).

As was already discussed in the **Introduction**, our grain model includes 200 water molecules, large enough to allow for a proper sampling of many adsorbing sites in comparison to previously adopted models. The grain construction was performed with a mixed semiempirical and molecular-mechanics level using the xTB (v.6.3.3)<sup>48</sup> code (GFN2<sup>34</sup> and the force field GFN-FF methods<sup>49</sup>) developed by Grimme's group at the University of Bonn.

**Binding Energy Sampling Site Procedure.** The NH<sub>3</sub> binding site sampling was done by placing a grid consisting of 12 vertexes (forming an icosahedron), which were tightened for a total of 162 vertexes uniformly spread around the grain.<sup>50</sup> The grid points were projected closer to the grain surface, and each point was substituted by a randomly oriented ammonia molecule with respect to the direction vector joining the N atom and the grain center of mass. The projection gives a distance between 2.5 and 3 Å from the grain, used to locate NH<sub>3</sub> (Figure 1).

**Preliminary Geometry Optimization.** After the NH<sub>3</sub> sampling, a preliminary geometry optimization via xTB-GFN2<sup>34,48</sup> was performed. Two subsequent geometry optimizations were carried out in which (i) only the NH<sub>3</sub> molecule was set free to relax on the grain, while all the water molecules were kept fixed at the optimized free grain positions, and (ii) the atomic positions of NH<sub>3</sub> and the water molecules included within a cutoff distance of 5 Å from the NH<sub>3</sub> were relaxed, while the remaining water molecules were kept fixed. This choice enforces the structural rigidity experienced by the water molecules in a real (and much larger) icy grain. During the second task, we found cases where the number of mobile water molecules changed during the optimization procedure, due to the rearrangements of both the NH<sub>3</sub> and the water molecules within the selected zone. In these cases, the described cycle was repeated by again selecting a new mobile zone and reoptimizing the structure until no changes in the number of water molecules occurred.

**Computational Methods.** After a preliminary geometry optimization with the xTB (v.6.3.3)<sup>48</sup> computational program, the refined binding energy distribution of ammonia on the amorphous ice model was obtained by combing the tools implemented in three codes: xTB (v.6.3.3), Gaussian (v.16, Revision B.01),<sup>51</sup> and ORCA (v.4.2.1).<sup>52</sup> We relied on the

multilevel ONIOM<sup>53</sup> (DFT:xTB-GFN2) approach as implemented in the Gaussian program to obtain accurate optimized geometries. As the GFN2<sup>34</sup> method has not yet been implemented in the Gaussian program, xTB (v.6.3.3)<sup>48</sup> was used as an external program to work on the low-level zone of the ONIOM method. Finally, the energies of the high-level zone were refined with ORCA (v.4.2.1)<sup>52</sup> at the DLPNO-CCSD-(T)<sup>54</sup> level of theory. Rendering of molecular images has been obtained via VMD software,<sup>55</sup> while the graphics elaboration and plots were obtained via the TikZ and PGFplots LATEX packages.

**ONIOM Method.** The ONIOM (“our own N-layered integrated molecular orbital and molecular mechanics”) method<sup>56</sup> is a hybrid approach that enables different *ab initio*, semiempirical, or classical mechanics based methods to be combined to different parts of a system to give a reliable hamornic frequencies, geometry and energy at a reduced computational cost. All of the calculations were performed with the two-layer ONIOM(QM:SQM) method. To be specific: the zone in which the quantum-mechanical method (QM) is used (also called the Model zone) consists of NH<sub>3</sub> and neighboring water molecules within 5 Å from NH<sub>3</sub>, while the whole system (Real zone) is treated at the semiempirical quantum mechanical (SQM) level. The total energy ( $E$ ), gradient vector ( $\mathcal{G}$ ) and Hessian matrix ( $\mathcal{H}$ ) for the ONIOM(QM:SQM) two-layer set up are, therefore

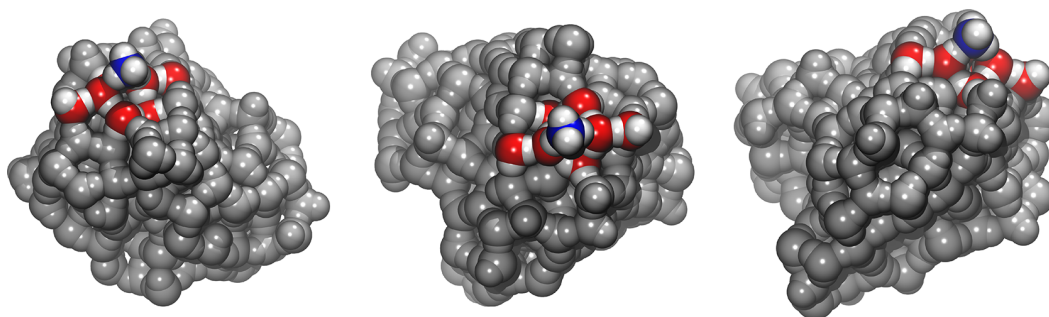
$$E = E(\text{R:SQM}) + E(\text{M:QM}) - E(\text{M:SQM}) \quad (1a)$$

$$\begin{aligned} \mathcal{G} = & \mathcal{G}(\text{R:SQM}) \times \mathcal{J}_{\text{M:R}} + \mathcal{G}(\text{M:QM}) \times \mathcal{J}_{\text{M:R}} \\ & - \mathcal{G}(\text{M:SQM}) \times \mathcal{J}_{\text{M:R}} \end{aligned} \quad (1b)$$

$$\begin{aligned} \mathcal{H} = & \mathcal{H}(\text{R:SQM}) \times \mathcal{J}_{\text{M:R}} + \mathcal{J}_{\text{M:R}}^T \times \mathcal{G}(\text{M:QM}) \times \mathcal{J}_{\text{M:R}} \\ & - \mathcal{J}_{\text{M:R}}^T \times \mathcal{G}(\text{M:SQM}) \times \mathcal{J}_{\text{M:R}} \end{aligned} \quad (1c)$$

where  $\mathcal{J}_{\text{M:R}}$  is the Jacobian matrix between the Model (M) and the Real (R) nuclei.

The binding energy (BE, positive for a bounded system), is defined as the opposite of the interaction energy, the last quantity being the difference between the energy of the complex between the grain and the adsorbate ( $E_c$ ) and the sum of the energies of the isolated adsorbate ( $E_{\text{ads}}^{\text{iso}}$ ) and the isolated grain ( $E_{\text{grn}}^{\text{iso}}$ ). The equation adopted for the calculation of the ONIOM BEs, after eq 1a, is



**Figure 2.** Three different perspectives showing the ONIOM zone: the atoms in the Model (high-level) zone are shown in colors while the low-level zone of the system is pictured in gray. Atom color legend: oxygen in red, nitrogen in blue, hydrogen in white.

$$\begin{aligned} \text{BE} &= -\Delta E \\ &= E_{\text{ads}}^{\text{iso}}(\text{QM}) + E_{\text{grn}}^{\text{iso}}(\text{QM:SQM}) - E_{\text{c}}(\text{QM:SQM}) \end{aligned} \quad (2)$$

where the energies of the isolated systems are referred to the specified level at which geometry are also optimized. BE can be decomposed in the pure electronic interaction ( $\text{BE}_e$ ) corrected for the basis set superposition error (BSSE) and the deformation energy ( $\delta E_{\text{def}}$ ) contributions.

The  $\text{BE}_e$  is given by

$$\text{BE}_e = E_{\text{ads}}^{\text{iso||c}}(\mathcal{G}(\text{grn})) + E_{\text{grn}}^{\text{iso||c}}(\mathcal{G}(\text{ads})) - E_{\text{c}}(\text{QM}) \quad (3)$$

where  $E_{\text{ads}}^{\text{iso||c}}(\mathcal{G}(\text{grn}))$  and  $E_{\text{grn}}^{\text{iso||c}}(\mathcal{G}(\text{ads}))$  are the energies of the isolated adsorbate and the grain in the geometries assumed in the complex (iso||c) in the presence of the ghost orbitals of the grain  $\mathcal{G}(\text{grn})$  and the adsorbate  $\mathcal{G}(\text{ads})$ , respectively. Obviously, as the BSSE is already taken into account by the definition in the GFN2 method, eq 3 only applies to the QM methods (*vide infra*) on the model zone.

The  $\delta E_{\text{def}}$  value is given by

$$\delta E_{\text{def}} = \frac{(E_{\text{ads}}^{\text{iso||c}} - E_{\text{ads}}^{\text{iso}})}{\delta E_{\text{def}}^{\text{ads}}} + \frac{(E_{\text{grn}}^{\text{iso||c}} - E_{\text{grn}}^{\text{iso}})}{\delta E_{\text{def}}^{\text{grn}}} \quad (4)$$

where  $\delta E_{\text{def}}^{\text{ads}}$  and  $\delta E_{\text{def}}^{\text{grn}}$  are the deformation energy of the adsorbate and the surface, respectively.  $\delta E_{\text{def}}$  is for the large majority of the cases a positive quantity; the exceptions will be discussed in a dedicated section.

Moreover, vibrational frequencies were computed on the model zone to obtain the zero-point energies (ZPE), from which the  $\Delta\text{ZPE}$  resulted as

$$\Delta\text{ZPE} = \text{ZPE}_{\text{c}} - \text{ZPE}_{\text{ads}}^{\text{iso}} - \text{ZPE}_{\text{grn}}^{\text{iso}} \quad (5)$$

When all the aforementioned contributions are included, eq 2 becomes

$$\text{BH}(0) = \frac{\text{BE}_e - (\delta E_{\text{def}}^{\text{grn}} + \delta E_{\text{def}}^{\text{ads}})}{\text{BE}} - \Delta\text{ZPE} \quad (6)$$

In our ONIOM setup, the low-level layer was treated with the xTB-GFN2 semiempirical quantum mechanical (SQM) method,<sup>34</sup> working as an external program with Gaussian16. The default xTB-GFN2 parameters were used for the SCF. On the high-level layer two different methods were used in order to compute subsequent tasks:

- Geometry optimization and frequency calculations: the B97D3<sup>57,58</sup> functional, as implemented in Gaussian16,

with the aug-cc-pVTZ basis set<sup>59</sup> and the default setup for geometry optimization, SCF, and integral grid density.

- Final energy refinement: DLPNO-CCSD(T) method,<sup>32,54</sup> as implemented in ORCA, with aug-cc-pVTZ as the primary basis set, while aug-cc-pVTZ/C<sup>60</sup> is used as the auxiliary basis set for the resolution of the identity (RI) approximation in electron repulsion integrals. All these calculations were carried out with a tight-PNO set up and the default settings for the SCF.

During the ONIOM geometry optimization all atoms outside the model zone were kept fixed; only mechanical embedding and no microiterations were used. In the frequency calculations (calculated in the harmonic approximation), only the normal modes related to the nuclei inside the Model zone were taken into account, keeping all the other nuclei fixed (Figure 2).

The treatment of the isolated icy surface required extra care, as the Model zone may change during the search for the optimum structure when the grain is adsorbing the  $\text{NH}_3$  molecule. Therefore, to ensure a proper coherence we used in this section as Model zone for evaluating the energy  $E_{\text{grn}}^{\text{iso}}$  of the free grain, the very last set of water molecules defined in the cyclic procedure described above, on an otherwise unique and fixed reference geometry of the free cluster.

**Model Zone Setup.** The definition of the Model zone, which is the core of any ONIOM-based procedure, implies not only a proper choice of the level of theory but also the number of atoms to be included in the QM description.

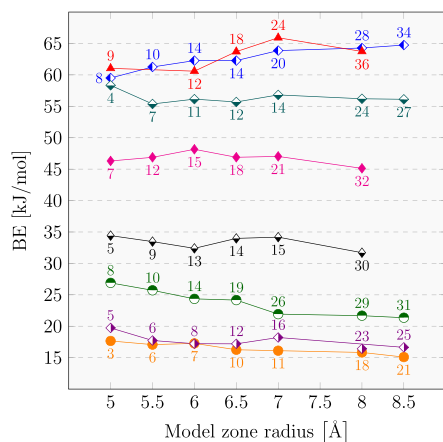
**Geometry Optimization Constraints.** We adopted the same strategy for the ONIOM calculation used for the optimizations performed with the GFN2<sup>34</sup> level. However, since the method to treat the Model zone is computationally demanding, a less tight criterion on the optimization convergence was used: when the number of water molecules of the Model zone changes by  $\geq 2$  units, we run further geometry optimizations with the redefined Model zone, until the above condition is satisfied.

**Model Zone Size Benchmark.** The Model zone defined within 5 Å from the  $\text{NH}_3$  relies on a tradeoff between two main requirements: (i) including all of the local  $\text{NH}_3$ - $\text{H}_2\text{O}$  interactions and (ii) saving computational resources.

In order to understand the influence of the Model zone size on the BE, a benchmark was performed by taking the single-point energy evaluation of eight different optimized cases with the standard Model zone definition (5 Å) and expanding its size from 5 Å up to 8.5 Å (which corresponds to including up to 21–34 water molecules) while the geometry of the whole system was kept fixed.

Single-point energy calculations were carried out at the same level of theory described in the previous section: i.e.,

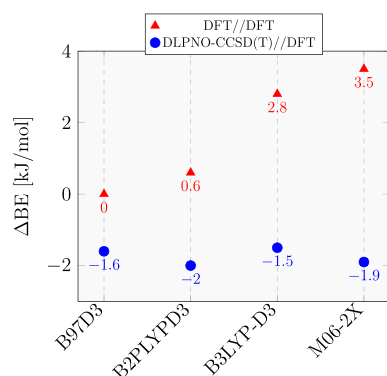
ONIOM(B97D3/aug-cc-pVTZ:xTB-GFN2). Figure 3 shows, for all but two samples, a change in the BE value well within 5 kJ/



**Figure 3.** BSSE-corrected BEs calculated at the ONIOM(B97D3/aug-cc-pVTZ:xTB-GFN2) level as a function of the Model zone size. Each symbol/color represents the same BE sample, while the number of water molecules inside the Model zone is reported close to the related symbol with the same color.

mol and a rather flat variation in the BE values. The two exceptions are at the limit of the threshold of 5 kJ/mol (i.e., within the chemical accuracy limit).

**Model Zone Method Benchmark.** The pure GGA B97D3 functional<sup>57,58</sup> used to deal with the Model zone is well adapted to deal with noncovalent interactions such as those responsible for the grain cohesion and the NH<sub>3</sub> BEs.<sup>57</sup> To assess the B97D3 performance for the present case, we compared, for one selected NH<sub>3</sub>/grain case, structures and BEs (corrected for BSSE) with (i) the B2PLYPD3 double-hybrid functional with empirical dispersion corrections,<sup>61</sup> (ii) the B3LYP<sup>62,63</sup> functional with the D3 version of Grimme's dispersion with the Becke–Johnson damping function,<sup>58</sup> and (iii) the Minnesota double-exchange M06-2X functional,<sup>64</sup> coupled with the aug-cc-pVTZ<sup>59</sup> basis set. DFT BEs were then refined at the DLPNO-CCSD(T)/(aug-cc-pVTZ and aug-cc-pVTZ/C) tight-PNO level (all the values were corrected for the BSSE) computed at each DFT optimized geometry. The results are presented in Figure 4. Among all adopted functionals, B97D3 is the one with the closest BE value with respect to the reference DLPNO-CCSD(T) value.



**Figure 4.** Differences  $\Delta$ BE between the BE B97D3 reference value (48.4 kJ/mol) and the BEs computed with the reported QM methods, all coupled with a aug-cc-pVTZ basis set quality and BSSE correction.

We also calculated the BE with the gold standard CCSD(T)/aug-cc-pVTZ on the same ONIOM(B97D3:xTB-GFN2) sample used in the previous test. The BE relative errors of B97D3 and DLPNO-CCSD(T) with respect to CCSD(T) are 0.9 and  $-0.7$  kJ/mol, respectively. These results validate the performance of both B97D3<sup>57</sup> and DLPNO-CCSD(T).<sup>65,66</sup>

**Adsorption Site Redundancy Reduction.** During the geometry optimization different NH<sub>3</sub> starting points may end up in the same minimum of the potential energy surface (PES), due to the complexity of the PES and the relatively weak interaction energy. For instance, many identical structures differ only in the permutation between the ammonia hydrogen atoms. Therefore, this redundancy in the adsorption sites was reduced by comparing the RMSD and  $\Delta$ BE values between all considered structures and discarding the cases for which  $\text{RMSD} \leq 1$  Å and  $|\Delta\text{BE}| \leq 1$  kJ/mol. After cleaning, a total of 77 unique structures from the total 162 starting points were analyzed.

**Machine Learning Binding Energy Classification.** Once the BE distribution, without site redundancy, was obtained, a clustering procedure was performed to analyze the data. Cluster analysis, or clustering, is an unsupervised machine-learning technique that involves the grouping of data points. This grouping is done in such a way that the members of the same cluster can be considered “similar” in some way (e.g., through metrics such as the  $L^2$  distance). In our case we exploited hierarchical agglomerative clustering (HAC), where a hierarchy of clusters is built with a bottom-up approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Sets of observations are linked via the so-called linkage criterion. The algorithm will merge the pairs of cluster that minimize this criterion. In our study we used the Scikit-Learn implementation of HAC,<sup>67</sup> using the minimum-distance linking criterion (namely “single”), specifying an *a priori* number of clusters of 2 (i.e., the number of clusters that we want to find). Finally, we scaled every feature to  $[0,1]$  in order to obtain the scaled invariance.

## RESULTS AND DISCUSSION

NH<sub>3</sub> usually behaves as a strong hydrogen bond acceptor, due to the negative electrostatic potential in the nitrogen lone pair region, while being a very weak hydrogen bond donor. For instance, the NH<sub>3</sub> crystal structure<sup>68</sup> shows only very weak hydrogen bonds between the NH<sub>3</sub> molecules, the N...H distance being as great as 2.35 Å. Indeed, our results basically show NH<sub>3</sub> acting as a strong H-bond acceptor of the dangling hydrogens of the icy grain and a weak H-bonding donor toward the water oxygen dangling atoms. After the harmonic frequency analysis, 16 samples show only one imaginary frequency in the  $[-50, -8]$   $\text{cm}^{-1}$  wavenumber range. Since the imaginary frequencies fall at very low wavenumbers and do not reflect nuclear motion of the NH<sub>3</sub> position, we also kept these structures to improve the statistics of the BE distribution, as their very low values do not alter the final BH(0) values.

**NH<sub>3</sub> Desorption Rate Prefactor.** In the desorption process, the desorption rate can be expressed as  $k_{\text{des}} = \nu(T)e^{-(\text{BE})/(k_{\text{B}}T)}$ , where  $\nu(T)$  is a pre-exponential factor that takes into account entropic effects, while the enthalpic contribution is inside the exponential part. In order to give reliable data to be used in astrochemical models and/or to have a connection with experiments, a pre-exponential factor must be provided together with the BE. Usually, depending on the substrate and adsorbate, a value between  $10^{12}$  and  $10^{13}$   $\text{s}^{-1}$  is assumed in experiments or as a first approximation in modeling



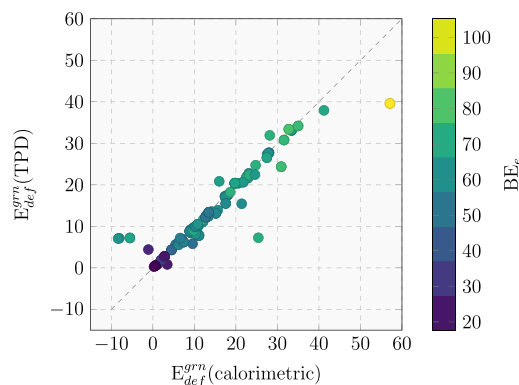
studies, as reported by Hasegawa and Herbst<sup>69</sup> (see e.g. the discussion in Minissale et al.<sup>70</sup>). We prefer to adopt the transition state theory within the immobile adsorbate approximation<sup>70,71</sup> to estimate the prefactor

$$\nu(T) = \frac{k_B T}{h} \left( \frac{2\pi m k_B T}{h} \right) A \frac{\sqrt{\pi}}{\sigma h^3} (8\pi^2 k_B T)^{3/2} \sqrt{I_x I_y I_z} \quad (7)$$

where  $k_B$  is the Boltzmann constant,  $m$  is the mass of the molecule,  $h$  is the Planck constant,  $A$  is the surface area per adsorbed molecule usually assumed to be  $10^{13} \text{ N}_a/\text{\AA}^2$ ,  $I_i$  is the  $i$ -esimal adsorbate principal moment of inertia, and  $\sigma$  is the symmetry adsorbate rotation factor. For  $\text{NH}_3$ , the principal moments of inertia are 2.76, 1.71, and 1.71  $\text{amu} \times \text{\AA}^2$ ,  $\sigma = 3$ , and  $m = 17 \text{ amu}$ . When these values and a desorption peak at  $T_{\text{des}} = 100 \text{ K}$  are used, the pre-exponential factor is  $1.94 \times 10^{15} \text{ s}^{-1}$ .<sup>70</sup> This value is recommended in association with the BE values computed with quantum mechanical approaches similar to those described in the present work.

**BE Evaluation: Calorimetric versus TPD Reference.** In the BE calculations, the definition of the “free” grain structure, from which the  $E_{\text{grn}}^{\text{iso}}$  value is computed, is crucial and may differ depending on what process one is simulating, while that for the  $\text{NH}_3/\text{grain}$  adduct ( $E_{\text{ads}}^{\text{iso}}$ ) is unambiguous. Usually, in dealing with adsorption on extended surfaces of metal or oxide materials, the reference structure is the bare isolated surface, fully optimized at the given level. In such cases, the forces keeping the metal atoms or the ions in place are much stronger than the BE with the adsorbate and, therefore, the whole structure is little affected by the interaction. In the present case, the icy grain is held by forces of the very same nature as those occurring between the adsorbate and the water molecules within the grain. Therefore, it may happen that, during the geometry optimization of the adsorbate/grain complex, the grain structure will be altered in such a way that the deformation energy  $\delta E_{\text{def}}^{\text{grn}} = E_{\text{grn}}^{\text{c}} - E_{\text{grn}}^{\text{iso}}$  becomes negative: i.e., the *deformed grain* is more stable than the isolated starting grain. In other words, the geometry relaxation induced by the adsorbate brings the icy cluster in a new local minimum, slightly deeper than the initial minimum. This only happens in a few cases, especially when the Model zone is redefined due to large movements associated with the  $\text{NH}_3$  molecule. To solve this ambiguity in the definition of the deformation energy, we chose, as a starting structure for the isolated cluster to be optimized, the structure resulting after the interaction of  $\text{NH}_3$ . In this way,  $\delta E_{\text{def}}^{\text{grn}}$  will always be positive. We defined these two approaches by considering different reference pristine grain geometries, as “calorimetric” (original initial grain geometry) and “TPD” (reference grain geometry after adsorption), respectively. The BE distributions from the two approaches will be presented and discussed in the following.

In the calorimetric approach, as in microcalorimetric measurements, it is assumed that the reference system is a clean unperturbed surface and that the heat of adsorption occurs when the adsorbate arrives on the surface from the gas phase. In the temperature-programmed desorption (TPD), the molecule is first adsorbed on the surface and then the temperature is raised up to the point at which the adsorbate leaves the surface. Clearly, when the surface is made by water ice, what is left after desorption cannot be considered equivalent to an unperturbed pristine icy surface, as in the calorimetric approach. These two approaches may lead to different BE values, as shown in Figure 5, which correlates the deformation energy  $E_{\text{def}}^{\text{grn}}$  contribution to the BE computed with both the TPD and calorimetric approaches.



**Figure 5.** Correlation of TPD vs calorimetric deformation  $E_{\text{def}}^{\text{grn}}$  energies. The color map shows the corresponding electronic interaction  $\text{BE}_e$  (see eq 6) associated with each point. All data are given in kJ/mol.

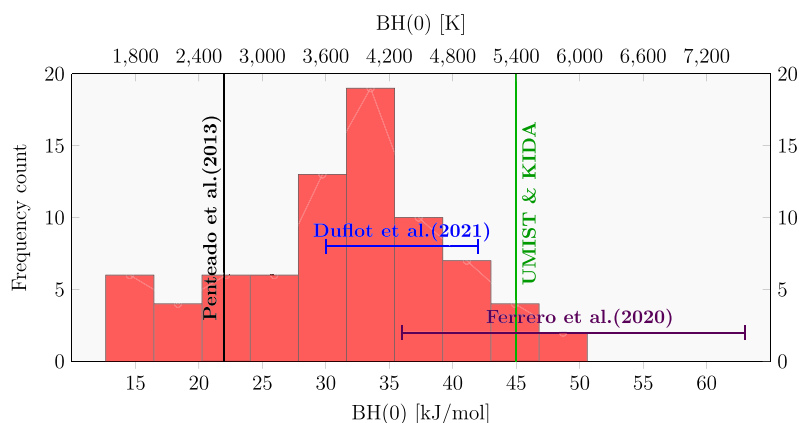
The purely electronic  $\text{BE}_e$  (which is free from the deformation energy) is also shown as a reference color bar. As expected, the two approaches lead to the same results for most cases. Nevertheless, there are some exceptions, such as some samples with low deformation energy values, in which the surface restructuring leads to a negative deformation energy in the calorimetric approach. The other two outliers ( $E_{\text{def}}^{\text{grn}}(\text{calorimetric}) \approx 25$  and  $60 \text{ kJ/mol}$ ) are due to the formation/breaking of some H bonds at the interface between high- and low-level zones, thus implying a redefinition of the Model zone itself and, therefore, the displacements of many water molecules. In the following, we only refer to the TPD method to compute the final BE distribution.

**Binding Energy Distribution. New BE Distribution versus Previous Values.** The final BH(0) values (see eq 6) have been organized in a bin width distribution following the Freedman–Diaconis estimator,<sup>72</sup> as shown in Figure 6. Due to the large number of different adsorbing sites the distribution is asymmetric, with a data dispersion ranging from 12.7 to 50.6 kJ/mol and a mean and mode (the most frequent values) of 31.1 and 33.5 kJ/mol, respectively. A fine analysis of the data shows that the deformation energy is the main source of data dispersion. The ZPE plays a minor role in the BH(0), its contribution being on the order of 10% on the total BH(0). The ZPE correction decreases the BE value by about 10 kJ/mol. A value of  $\sim 45.7 \text{ kJ/mol}$  is reported in the astrochemistry databases, which is in the same range, or higher, with respect to the BE values for water self-adsorption.<sup>14,73</sup>

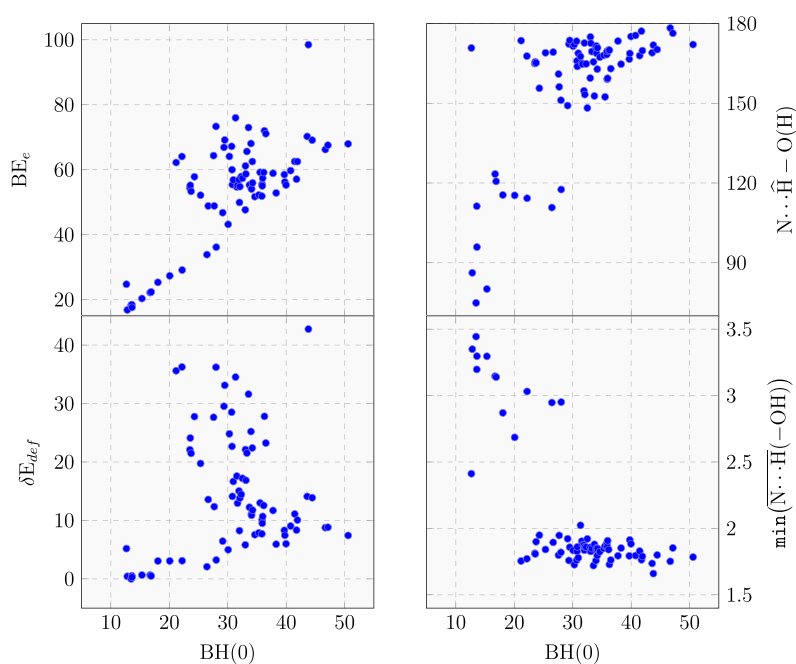
A comparison with computed literature BE values by Ferrero et al.,<sup>16</sup> computed on a periodic crystalline proton-ordered ice slab model (51.8 kJ/mol) and on an amorphous water slab model (35.9–62.8 kJ/mol), is shown in Figure 6. In that work, the sampling of binding sites on the amorphous slab included just seven cases and all the interactions found displayed  $\text{NH}_3$  as an acceptor of at least one hydrogen bond. In the work by Duflo et al.,<sup>31</sup> a procedure similar to the present one (ONIOM(CBS/DLPNO-CCSD(T):PM6)//ONIOM( $\omega$ B97X-D/6-31+G\*\*):PM6)) was adopted to compute a ZPE-corrected BE. BE values of  $35.9 \pm 11.6 \text{ kJ/mol}$  have been computed, in good agreement with our values of  $31.1 \pm 8.6 \text{ kJ/mol}$ , despite the fact that a very different methodology was adopted to build up the underneath ice.

**Clustering Analysis.** On the final data set of 77 BEs, a machine-learning (ML)-based procedure was used in order to





**Figure 6.** BSSE-corrected BH(0) distribution at the DLPNO-CCSD(T)/aug-cc-pVTZ level and ZPE calculated at the ONIOM(B97D3/aug-cc-pVTZ:xTB-GFN2) level.



**Figure 7.** Correlation plots between BH(0) and the feature vectors used in the ML clustering. BH(0),  $BE_e$  and  $\delta E_{def}$  are given in units of kJ/mol. Distances are given in Å and angles in degrees. All BH(0) and  $BE_e$  values are BSSE-corrected.

correlate BH(0) with other energetic and geometrical parameters:

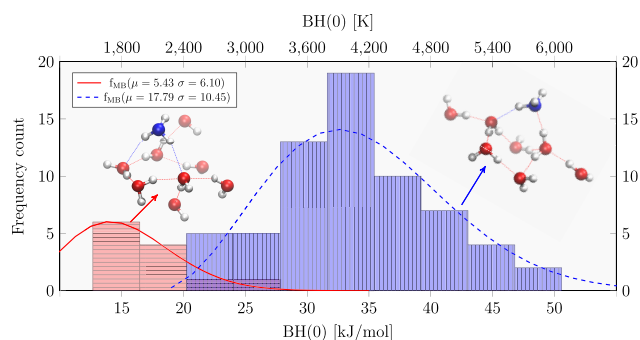
- the minimum H-bond distance,  $\min(\overline{N\cdots H(-OH)})$
- the H-bond angle  $N\cdots\hat{H}-O(H)$  referenced to the  $\min(\overline{N\cdots H(-OH)})$  H bond
- the deformation energy  $\delta E_{def}$
- the pure electronic  $BE_e$

The correlation plots are shown in Figure 7.

The plot of both  $\overline{N\cdots H(-OH)}$  and  $N\cdots\hat{H}-O(H)$  revealed a rather clean clustering, in which at high BH(0) values correspond to H-bond lengths well below 2 Å ( $NH_3$  as H-bond acceptor), while at low BH(0) values H-bond distances were over 2.5 Å ( $NH_3$  as H-bond donor). This correlates also with the  $N\cdots\hat{H}-O(H)$  angle, moving from values close to linearity for high BH(0) values to random values from linearity for the low-BH(0) range. Less trivial is the correlation between BH(0) and its different energy components. About the

deformation energy  $\delta E_{def}$  a number of points are almost aligned as a baseline in the 0–10 kJ/mol range, while in the region of intermediate BH(0) values the points are quite spread out. The same erratic trend is observed in the correlation with  $BE_e$ , revealing that the vast majority of cases exhibits a final BH(0) value that is a compromise of a large geometry deformation energy compensated by a large electronic binding energy. The few cases at very high BH(0) characterized by small  $\delta E_{def}$  values are due to favorable adsorption sites, already suitable to host the  $NH_3$  molecule and, therefore, not requiring a large structural deformation.

The geometrical clustering analysis applied to the binding energy distribution shown in Figure 6 is reported in Figure 8. The two clusters rely, as expected from chemical knowledge, on the two possible H bonds that the ammonia can form with water: the stronger  $N\cdots H(-OH)$  and the weaker  $N-H\cdots O(H_2)$ , where the ammonia is respectively a H-bond acceptor and a H-bond donor. In light of these results, the asymmetric shape of the



**Figure 8.** ML clustering analysis applied to the BH(0) distribution of Figure 6. The continuous red and dashed blue curves are the  $f_{\text{MB}}(\text{hist}(\text{BH}(0)), \sigma, \mu)$  Maxwell–Boltzmann best fit for the two histogram clusters. The insets show the Model (high-level) zones of two representative samples, with high (rightmost) and low (leftmost) BH(0) values. Atom color legend: oxygen in red, nitrogen in blue and hydrogen in white.

distribution at low BH(0) is due to the cluster distribution related to the H-bonds in which ammonia is the proton donor.

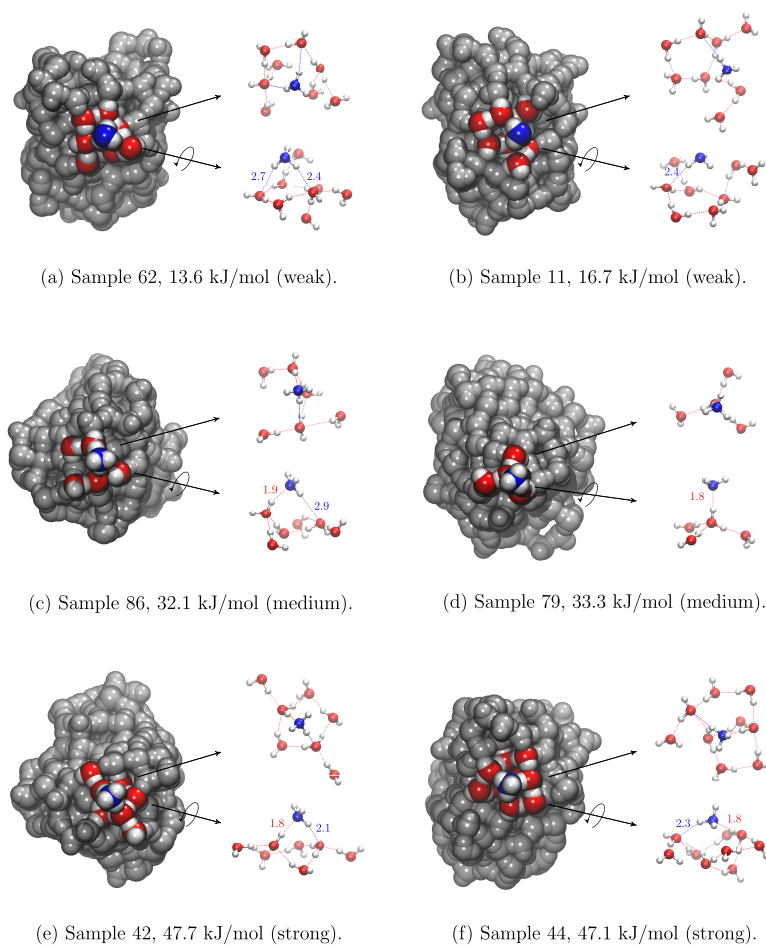
Moreover, as shown in Figure 8, the two histogram clusters were fitted with a non-normalized Maxwell–Boltzmann distribution function  $f_{\text{MB}}(x, \sigma, \mu)$ :

$$f_{\text{MB}}(x, \sigma, \mu) = \frac{(x - \mu)^2}{\sigma^3} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (8)$$

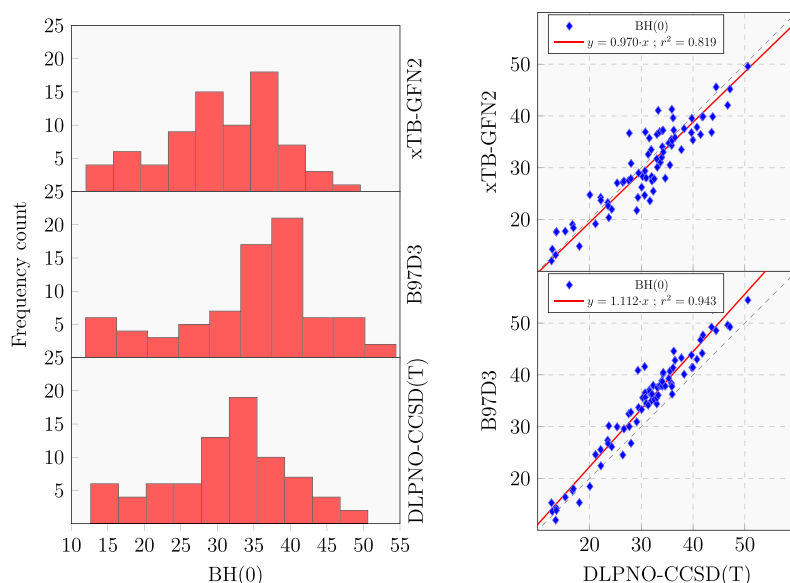
where, in our case,  $x$  values are the bin width medium of the BH(0) histogram and  $\mu$  and  $\sigma$  the distribution parameters.

Figure 9 shows a selected number of grain/NH<sub>3</sub> structures, spanning from weak to strong values of BH(0), evidencing the already mentioned features of NH<sub>3</sub> when it interacts through H bonds.

Experimental evidence of the tail distribution at very low BE(0) can be searched in the literature, as summarized by Ferrero et al.<sup>16</sup> NH<sub>3</sub> TPD experiments on amorphous and crystalline water surfaces were reported by Collings et al.<sup>37</sup> and He et al.<sup>23</sup> However, Collings et al.,<sup>37</sup> who only carried out experiments on amorphous water ice, did not explicitly derive the NH<sub>3</sub> BE. On the basis of their curve, Penteado et al.<sup>13</sup> successively estimated a BE equal to 22.5 kJ/mol = 2706 K using a pre-exponential factor equal to  $10^{12} \text{ s}^{-1}$ . The BE becomes 3460 K if a pre-exponential factor of  $1.94 \times 10^{15} \text{ s}^{-1}$  is used. In contrast, He et al.<sup>23</sup> only derived the BE for adsorption on crystalline ice, as they found that NH<sub>3</sub> desorbs at the temperature where the amorphous water ice becomes crystalline. Inverting the TPD curve for the crystalline ice adsorption using the pre-exponential factor of  $10^{-12} \text{ s}^{-1}$ , He et



**Figure 9.** Selected cases of weak, medium, and strong NH<sub>3</sub> BH(0) values. On the right of each cluster the Model zone is highlighted in ball and stick representation (top and side views). Distances are given in Å. An online database could be used to easily interact with and inspect all the samples, as described in the relative subsection.



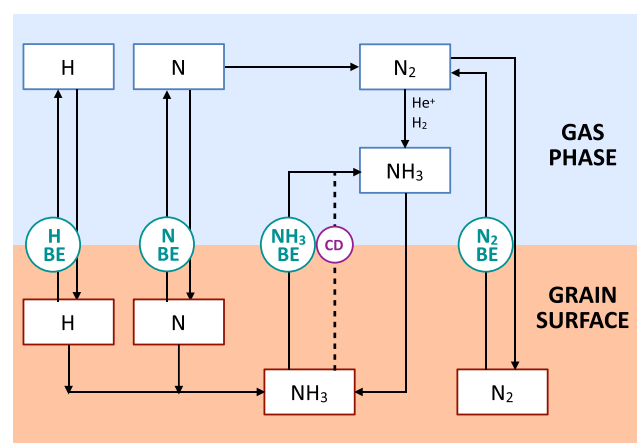
**Figure 10.** (left) BH(0) distributions for DLPNO-CCSD(T), B97D3, and xTB-GFN2 methods. (right) BH(0) correlation diagrams of B97D3 and xTB-GFN2 against DLPNO-CCSD(T). Each histogram bin width has been calculated with the proper Freedman–Diaconis estimator. All values are given in kJ/mol.

al.<sup>23</sup> derived a BE of about 4000 K for a low surface coverage ( $\leq 0.5$ ) and of about 3000 K for a full coverage (see Figure 9 of ref 23). However, this last value is almost the same as that derived by the TPD experiments of  $\text{NH}_3$  adsorbed on a gold surface,<sup>74,75</sup> suggesting that a sizable fraction of BE is due to the lateral interactions between  $\text{NH}_3$  within the adsorbed multilayers and not to the interaction with the ice surface. Moreover the 3000 K BE value (computed with a pre-exponential factor of  $10^{-12} \text{ s}^{-1}$ ) becomes 3754 K, with a pre-exponential factor of  $1.94 \times 10^{15} \text{ s}^{-1}$ , indeed larger than our lower-end BE value. Therefore, in both the experimental works presented, the low end of the ammonia BE that we computed was not detected. One possibility is that, under low  $\text{NH}_3$  coverage,  $\text{NH}_3$  exhibiting very weak BE values (such as that corresponding to our lowest BEs) will easily diffuse to empty sites characterized by higher BE values, instead of being entirely desorbed. This process is only effective at moderate  $\text{NH}_3$  coverage, where sites with high BE values are still available for occupation. This indeed happens in the TPD experiment, in which the thermal heating brings an oversampling of sites at high BE values.<sup>70</sup> While a detailed astrochemical modeling that may better elucidate this point is postponed to a dedicated work, this discussion also highlights how critical the comparison can be between experimental data extracted from the TPD and the computed data through quantum mechanical calculations if the pre-exponential factor is not treated on the same foot and similar  $\text{NH}_3$  surface coverages are considered.

**xTB-GFN2 Validation.** In our recent works,<sup>24,76,77</sup> we adopted xTB-GFN2 as the low-level semiempirical method. The ONIOM procedure requires, to be robust, a low level of theory giving structures and energies not too far from the high-level theory. Here, we compare the xTB-GFN2 BH(0) values computed as a single-point xTB-GFN2 energy evaluation on the ONIOM optimized geometries with the more accurate ONIOM structures, computed at the DLPNO-CCSD(T) level. Figure 10 shows the excellent performances of xTB-GFN2, considering its very low computational cost, also in comparison with B97D3, which gives results in better agreement with the

DLPNO-CCSD(T) data. xTB-GFN2 BH(0) values are, instead, systematically underestimated with respect to the reference. The worse GFN2 correlation may be due to the geometric distortion in the Model zone, since it is evaluated at the B97D3 level.

**Astrochemical Implications on  $\text{NH}_3$  BE Distribution.** As mentioned in the Introduction,  $\text{NH}_3$  is ubiquitous in the molecular ISM and can be either gaseous or iced. Also,  $\text{NH}_3$  can be formed both in the gas phase from molecular nitrogen<sup>78</sup> and on the grain surfaces by hydrogenation of atomic nitrogen,<sup>79</sup> as shown in Figure 11.



**Figure 11.** Scheme of the interstellar chemistry involving  $\text{NH}_3$ . Ammonia can be synthesized on the grain surfaces by hydrogenation of frozen  $\text{N}$ <sup>79</sup> (left part of the figure) or in the gas phase from reactions involving  $\text{N}_2$ <sup>78</sup> and then frozen onto the grain surfaces (right part of the figure). Once on the grain surface,  $\text{NH}_3$  can be thermally desorbed or injected into the gas phase via the so-called chemical desorption (CD) or because of the cosmic-ray desorption (CRD), as marked with the dashed line. Both thermal and CRD desorption are governed by the  $\text{NH}_3$  BE and involve the whole frozen  $\text{NH}_3$ , while CD injects a small fraction ( $\leq 1\%$ ) of the  $\text{NH}_3$  formed by the N hydrogenation on the grain surface.

The crucial parameter that governs whether  $\text{NH}_3$  is in the gaseous or solid form is its BE. The fact that the  $\text{NH}_3$  BE is not a single value but a distribution that covers a relatively large range of energies, from 1800 to 6000 K (15–50 kJ/mol), can have an important effect (see e.g. Grassi et al.<sup>80</sup>).

While gaseous ammonia in warm ( $\geq 100$  K) regions does not present any particular puzzle, its presence in cold objects might. The most extreme example is the gaseous ammonia observed in prestellar objects. In L1544, a very well studied prestellar core,<sup>81</sup> the dust temperature at the center of the condensation is only 7 K<sup>82</sup> and ammonia should be completely frozen onto the grain mantles.<sup>83,84</sup> In contrast, ammonia is observed to be gaseous.<sup>45</sup> Various reasons have been proposed, mainly that ammonia is desorbed from the grain mantles because of the chemical energy released by its formation, which is believed to be due to the hydrogenation of N (see e.g. Sipilä et al.<sup>84</sup>). These authors found that slightly less than 1% of the ammonia formed on the grain icy surfaces could be necessary to reproduce the observed values. However, these authors also modeled the possibility that the ammonia BE is smaller than the standard high value and considered the cases with BEs equal to 1000 and 3000 K (8 and 25 kJ/mol), respectively. As expected, an ammonia BE equal to 1000 K would result in an overly large gaseous ammonia abundance with respect to the observed value. However, if one considers the BE distribution of Figure 8, about 3% of the frozen ammonia would have a BE equal to 1800 K (15 kJ/mol) so that, very likely, the predictions would be in agreement with the observations.

## SUMMARY AND CONCLUSIONS

In this paper we provide a new framework to compute the binding energy (BE) distribution of any relevant interstellar species adsorbed at the surface of an icy grain mantle, in a reproducible and user-friendly automated way. Two main parameters are controlled by the user: the ONIOM high-level zone size, which should be large enough to account for all the H-bond interactions with the ice, and the DFT method for geometry optimization (and subsequent frequency analysis). The framework can be divided into four subsequent blocks:

- (1) building up of the grain model and choice of the species to be absorbed
- (2) sampling of all possible binding sites on the icy grain model by an automatic unbiased procedure and geometry optimization with a low level of theory (xTB-GFN2)
- (3) geometry optimization and zero-point energy correction using the ONIOM method (B97D3:xTB-GFN2)
- (4) final ONIOM single-point (SP) energy refinement with a higher level of theory (DLPNO-CCSD(T)//B97D3:xTB-GFN2)

The first two tasks are encoded in the ACO-FROST program<sup>24</sup> (see also *Icy Grain Model and  $\text{NH}_3$  Binding Site Sampling* subsection). An extensive benchmark applied to the ammonia case is reported in *Methodology*, where we demonstrate the performance of the chosen methodology, highlighting its excellent compromise between accuracy and computational cost. Moreover, we also demonstrated in a dedicated section that the same distribution calculated at the full xTB-GFN2 level is similar to that at the ONIOM(DLPNO-CCSD(T)//B97D3:xTB-GFN2) level, which confirms the robustness of GFN2 despite the fact that its cost is orders of magnitude smaller than those of DFT and DLPNO-CCSD(T).

We highlight a particular aspect that needs to be treated with particular care: the reference of the bare water grain. This attention is due to the cooperativity and mobility of the H-bond network that, when the bare grain is optimized after removing the adsorbate, can lead to strong rearrangements which may result in a negative deformation energy (which is almost always a positive quantity). For this reason, we propose and compare two different references for the bare icy surface, which somehow mimic the two experimental techniques used to study such a phenomenon: TPD (each reference is obtained after adsorption, i.e. the  $\text{NH}_3$  and the bare grain structure reoptimized) and calorimetry (the reference is the starting optimized bare grain, before site sampling).

The final ZPE- and BSSE-corrected BE distribution (BH(0)) for ammonia shows, as expected from our 77 unbiased samples, all the possible interactions of  $\text{NH}_3$  with a water surface, acting as a H-bond donor and/or acceptor. This variety of BE is made possible by the large number of chemically different binding sites that the built icy grain model presents (not only in terms of dangling species but also from a morphological point of view of the global structure). Using an unsupervised machine-learning clustering technique, we correlate the structures and their BH(0). The two clusters found with the ML algorithm can be approximated by two Maxwell–Boltzmann distribution functions with the first peak at around 34 kJ/mol (or  $\sim 4000$  K) and the second peak at  $\sim 15$  kJ/mol (or  $\sim 1800$  K). As expected, the asymmetrical shape at low BH(0) is due to ammonia acting as a H-bond donor, while at high BH(0) we found ammonia acting as both donor and acceptor from a variety of ice dangling hydrogen atoms whose propensity to make H-bonds is modulated by the cooperativity of the H-bond network within the grain. The first peak of the  $\text{NH}_3$  BH(0) distribution matches very well with the data in the literature, from both experimental and theoretical works. In contrast, we show for the first time the presence of a second peak at lower BH(0). We discuss how this second peak may explain the longstanding puzzle of the presence of ammonia in cold and dense ISM.

In summary, the major novelty of our work is the development of a framework with a general applicability to simulate all statistically meaningful binding sites of a species adsorbed on an icy surface, with high accuracy at reasonable computational cost.

It allows producing realistic BE distributions of interstellar molecules, which is a breakthrough with important implications in astrochemistry. Our results point toward a more complex scenario about BEs than it has been thought in the past, as BE in astrochemical models are very often assumed to have a single or very few values, which is an oversimplification of the reality.

Finally, the presence of a low BE definitively has an important effect on our understanding of the chemical evolution of the molecular ISM.

**Online Database.** To easily handle the large data set of BE samples (atomic coordinates and BH(0) values), we developed and made publicly available a Web site<sup>85</sup> based on the molecule hyperactive JSmol plugin (Jmol: an open-source Java viewer for chemical structures in 3D). The extended electronic version of the calculated results, the 77 optimized structures at the ONIOM(B97D3/aug-cc-pVTZ:xTB-GFN2) level, are available at [https://tinaccil.github.io/Jmol\\_BE\\_NH3\\_visualization/](https://tinaccil.github.io/Jmol_BE_NH3_visualization/).



## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsearthspacechem.2c00040>.

Dataframe.csv presenting all the binding energy (BE) information for each sample, where all the values are referenced to the TPD approach (see the discussion in the paper), all the of computed structures (.xyz files) at the ONIOM(B97D3/aug-cc-pVTZ:xtb-GFN2) level, in the structures folder, with the complex structures (grain\_molX.xyz) and the Model zone (CP\_grain\_molX.xyz), the number inside (X), and the file name referenced to the sample ("n\_sample", in the Dataframe), bare grain and ammonia (bare\_grain.xyz and nh3.xyz), optimized respectively at xTB-GFN2 and B97D3/aug-cc-pVTZ, input examples for all of the programs in the example\_input folder, and the script to interface (i.e., xtb-Gaussian.sh) xtb to Gaussian16 (ZIP)

## ■ AUTHOR INFORMATION

### Corresponding Author

Piero Ugliengo – Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; [orcid.org/0000-0001-8886-9832](https://orcid.org/0000-0001-8886-9832); Email: [piero.ugliengo@unito.it](mailto:piero.ugliengo@unito.it)

### Authors

Lorenzo Tinacci – Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), 38000 Grenoble, France; [orcid.org/0000-0001-9909-9570](https://orcid.org/0000-0001-9909-9570)

Aurèle Germain – Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; [orcid.org/0000-0001-7856-0516](https://orcid.org/0000-0001-7856-0516)

Stefano Pantaleone – Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi di Perugia, 06123 Perugia, Italy; [orcid.org/0000-0002-2457-1065](https://orcid.org/0000-0002-2457-1065)

Stefano Ferrero – Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain

Cecilia Ceccarelli – Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), 38000 Grenoble, France; [orcid.org/0000-0001-9664-6292](https://orcid.org/0000-0001-9664-6292)

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsearthspacechem.2c00040>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This project has received funding within the European Union's Horizon 2020 research and innovation programme from the European Research Council (ERC) for the project "The Dawn of Organic Chemistry" (DOC), grant agreement no. 741002, and from the Marie Skłodowska-Curie for the project "Astro-Chemical Origins" (ACO), grant agreement no. 811312. S.P. and P.U. acknowledge the Italian Space Agency for cofunding the Life in Space Project (ASI N. 2019-3-U.O). Support from the Italian MUR (PRIN 2020, Astrochemistry beyond the second period elements, Prot. 2020AFB3FX) is gratefully acknowledged. CINES-OCCIGEN HPC is kindly acknowl-

edged for the generous allowance of supercomputing time through the A0060810797 project. L.T. is grateful to Giovanni Bindi for insightful discussions on ML classification and to J.L., L.M., Pt.F., and finally the LATEX community for insights on TikZ and PGFPlots packages. Finally, we wish to acknowledge extremely useful discussions with Prof. Grotobape.

## ■ REFERENCES

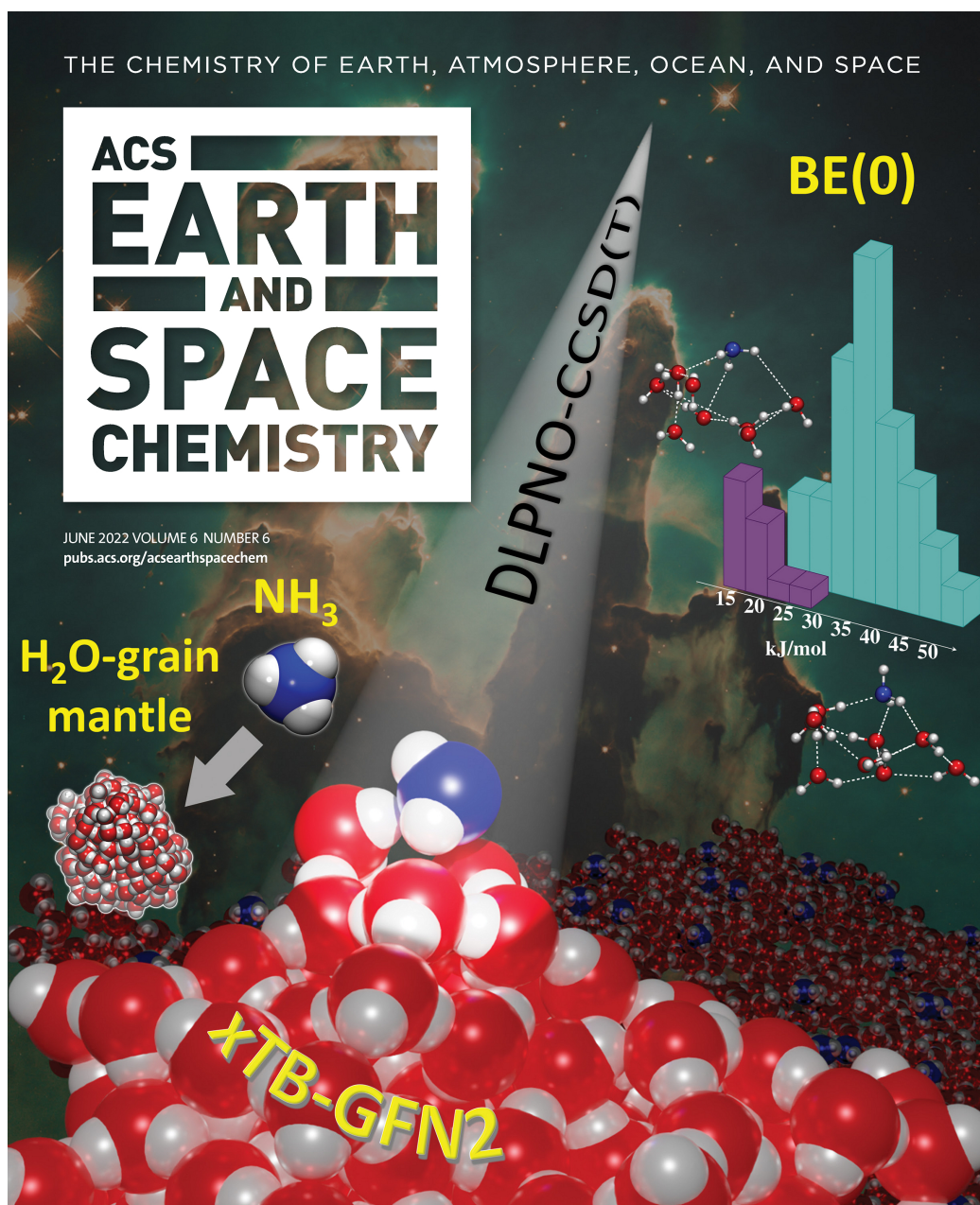
- (1) Tielens, A.; Hagen, W. Model calculations of the molecular composition of interstellar grain mantles. *Astron. Astrophys.* **1982**, *114*, 245–260.
- (2) Molpeceres, G.; Rimola, A.; Ceccarelli, C.; Kästner, J.; Ugliengo, P.; Maté, B. Silicatemediated interstellar water formation: A theoretical study. *Monthly notices of the Royal Astronomical Society* **2019**, *482*, 5389–5400.
- (3) Dulieu, F.; Amiaud, L.; Congiu, E.; Fillion, J.-H.; Matar, E.; Momeni, A.; Pirronello, V.; Lemaire, J. Experimental evidence for water formation on interstellar dust grains by hydrogen and oxygen atoms. *Astronomy & Astrophysics* **2010**, *512*, A30.
- (4) Jing, D.; He, J.; Brucato, J.; De Sio, A.; Tozzetti, L.; Vidali, G. On water formation in the interstellar medium: Laboratory study of the O + D reaction on surfaces. *Astrophysical Journal Letters* **2011**, *741*, L9.
- (5) Oba, Y.; Watanabe, N.; Hama, T.; Kuwahata, K.; Hidaka, H.; Kouchi, A. Water formation through a quantum tunneling surface reaction, OH<sup>+</sup> H<sub>2</sub>, at 10 K. *Astrophysical Journal* **2012**, *749*, 67.
- (6) Taquet, V.; Ceccarelli, C.; Kahane, C. Multilayer modeling of porous grain surface chemistry-I. The GRAINOBLE model. *Astronomy & Astrophysics* **2012**, *538*, A42.
- (7) Blake, G. A.; Sutton, E.; Masson, C.; Phillips, T. Molecular abundances in OMC-1: The chemical composition of interstellar molecular clouds and the influence of massive star formation. *Astrophys. J.* **1987**, *315*, 621–645.
- (8) Charnley, S.; Tielens, A.; Millar, T. On the molecular complexity of the hot cores in Orion A-Grain surface chemistry as 'The last refuge of the scoundrel'. *Astrophysical Journal* **1992**, *399*, L71–L74.
- (9) Ceccarelli, C.; Loinard, L.; Castets, A.; Tielens, A.; Caux, E. The hot core of the solartype protostar IRAS 16293–2422: H<sub>2</sub>CO emission. *Astron. Astrophys.* **2000**, *357*, L9–L12.
- (10) Bachiller, R.; Martin-Pintado, J.; Fuente, A. High-velocity hot ammonia in bipolar outflows. *Astrophysical Journal* **1993**, *417*, L45.
- (11) Flower, D. R.; Forêts, G. P. d. Grain-mantle erosion in magnetohydrodynamic shocks. *Mon. Not. R. Astron. Soc.* **1994**, *268*, 724–732.
- (12) Lefloch, B.; Ceccarelli, C.; Codella, C.; Favre, C.; Podio, L.; Vastel, C.; Viti, S.; Bachiller, R. L1157-B1, a factory of complex organic molecules in a solar-type starforming region. *Monthly Notices of the Royal Astronomical Society: Letters* **2017**, *469*, L73–L77.
- (13) Penteado, E.; Walsh, C.; Cuppen, H. Sensitivity analysis of grain surface chemistry to binding energies of ice species. *Astrophysical Journal* **2017**, *844*, 71.
- (14) Wakelam, V.; Loison, J.-C.; Mereau, R.; Ruaud, M. Binding energies: New values and impact on the efficiency of chemical desorption. *Molecular Astrophysics* **2017**, *6*, 22–35.
- (15) Das, A.; Sil, M.; Gorai, P.; Chakrabarti, S. K.; Loison, J.-C. An Approach to Estimate the Binding Energy of Interstellar Species. *Astrophysical Journal Supplement Series* **2018**, *237*, 9.
- (16) Ferrero, S.; Zamirri, L.; Ceccarelli, C.; Witzel, A.; Rimola, A.; Ugliengo, P. Binding Energies of Interstellar Molecules on Crystalline and Amorphous Models of Water Ice by Ab Initio Calculations. *Astrophysical Journal* **2020**, *904*, 11.
- (17) 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.
- (18) 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 and Space Chemistry* **2018**, *2*, 720–734.

- (19) Pantaleone, S.; Enrique-Romero, J.; Ceccarelli, C.; Ferrero, S.; Balucani, N.; Rimola, A.; Ugliengo, P. H<sub>2</sub> Formation on Interstellar Grains and the Fate of Reaction Energy. *Astrophysical Journal* **2021**, *917*, 49.
- (20) Amiaud, L.; Fillion, J.; Baouche, S.; Dulieu, F.; Momeni, A.; Lemaire, J. Interaction of D<sub>2</sub> with H<sub>2</sub>O amorphous ice studied by temperature-programmed desorption experiments. *J. Chem. Phys.* **2006**, *124*, 094702.
- (21) Bovolenta, G.; Bovino, S.; Vöhringer-Martinez, E.; Saez, D. A.; Grassi, T.; Vogt-Geisse, S. High level ab initio binding energy distribution of molecules on interstellar ices: Hydrogen uoride. *Molecular Astrophysics* **2020**, *21*, 100095.
- (22) Molpeceres, G.; Zaverkin, V.; Kästner, J. Neural-network assisted study of nitrogen atom dynamics on amorphous solid water - I. adsorption and desorption. *Monthly notices of the Royal Astronomical Society* **2020**, *499*, 1373–1384.
- (23) He, J.; Acharyya, K.; Vidali, G. Binding energy of molecules on water ice: Laboratory measurements and modeling. *Astrophysical Journal* **2016**, *825*, 89.
- (24) Germain, A.; Tinacci, L.; Pantaleone, S.; Ceccarelli, C.; Ugliengo, P. Computer Generated Realistic Interstellar Icy Grain Models: Physico-chemical Properties and Interaction with NH<sub>3</sub>. *ACS Earth & Space Chemistry* **2020**, *6*, 1286–1298.
- (25) Watts, J. D.; Gauss, J.; Bartlett, R. J. Coupled-cluster methods with noniterative triple excitations for restricted open-shell Hartree-Fock and other general single determinant reference functions. Energies and analytical gradients. *J. Chem. Phys.* **1993**, *98*, 8718–8733.
- (26) Sauer, J. Ab Initio Calculations for Molecule-Surface Interactions with Chemical Accuracy. *Acc. Chem. Res.* **2019**, *52*, 3502–3510.
- (27) Song, L.; Kästner, J. Formation of the prebiotic molecule NH<sub>2</sub>CHO on astronomical amorphous solid water surfaces: accurate tunneling rate calculations. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29278–29285.
- (28) Molpeceres, G.; Kästner, J. Adsorption of H<sub>2</sub> on amorphous solid water studied with molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7552–7563.
- (29) Sameera, W.; Senevirathne, B.; Andersson, S.; Maseras, F.; Nyman, G. ONIOM (QM: AMOEBA09) Study on binding energies and binding preference of OH, HCO, and CH<sub>3</sub> radicals on hexagonal water ice (Ih). *J. Phys. Chem. C* **2017**, *121*, 15223–15232.
- (30) Sameera, W.; Senevirathne, B.; Andersson, S.; Al-Lbadi, M.; Hidaka, H.; Kouchi, A.; Nyman, G.; Watanabe, N. CH<sub>3</sub>O Radical Binding on Hexagonal Water Ice and Amorphous Solid Water. *J. Phys. Chem. A* **2021**, *125*, 387–393.
- (31) 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.
- (32) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, 024109.
- (33) Rezac, J.; Fanfrik, J.; Salahub, D.; Hobza, P. Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of non-covalent complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749–1760.
- (34) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB: An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (35) Redhead, P. A. Thermal desorption of gases. *Vacuum* **1962**, *12*, 203–211.
- (36) King, D. A. Thermal desorption from metal surfaces: A review. *Surf. Sci.* **1975**, *47*, 384–402.
- (37) Collings, M. P.; Anderson, M. A.; Chen, R.; Dever, J. W.; Viti, S.; Williams, D. A.; McCoustra, M. R. A laboratory survey of the thermal desorption of astrophysically relevant molecules. *Mon. Not. R. Astron. Soc.* **2004**, *354*, 1133–1140.
- (38) Cheung, A.; Rank, D. M.; Townes, C.; Thornton, D. D.; Welch, W. Detection of NH<sub>3</sub> Molecules in the Interstellar Medium by Their Microwave Emission. *Phys. Rev. Lett.* **1968**, *21*, 1701.
- (39) Winnewisser, G.; Churchwell, E.; Walmsley, C. Ammonia in absorption in the direction of Sagittarius B2. *Astron. Astrophys.* **1979**, *72*, 215–223.
- (40) Liszt, H.; Lucas, R.; Pety, J. Comparative chemistry of diffuse clouds-V. Ammonia and formaldehyde. *Astronomy & Astrophysics* **2006**, *448*, 253–259.
- (41) Morris, M.; Zuckerman, B.; Palmer, P.; Turner, B. Interstellar ammonia. *Astrophysical Journal* **1973**, *186*, 501–528.
- (42) Umamoto, T.; Mikami, H.; Yamamoto, S.; Hirano, N. The ortho-to-para ratio of ammonia in the L1157 outflow. *Astrophysical Journal Letters* **1999**, *525*, L105.
- (43) Mundy, L. G.; Wootten, H.; Wilking, B. A. The circumstellar structure of IRAS 16293-2422-(C-18) O, NH<sub>3</sub>, and CO observations. *Astrophysical Journal* **1990**, *352*, 159–166.
- (44) Irvine, W.; Goldsmith, P.; Hjalmarsen, Å. *Interstellar Processes*; Springer: 1987; pp 560–609.
- (45) Crapsi, A.; Caselli, P.; Walmsley, M. C.; Tafalla, M. Observing the gas temperature drop in the high-density nucleus of L 1544. *Astronomy & Astrophysics* **2007**, *470*, 221–230.
- (46) Salinas, V. N.; Hogerheijde, M. R.; Bergin, E. A.; Cleeves, L. I.; Brinch, C.; Blake, G. A.; Lis, D. C.; Melnick, G. J.; Panić, O.; Pearson, J. C.; et al. First detection of gasphase ammonia in a planet-forming disk-NH<sub>3</sub>, N<sub>2</sub>H<sup>+</sup>, and H<sub>2</sub>O in the disk around TW Hydrae. *Astronomy & Astrophysics* **2016**, *591*, A122.
- (47) Knacke, R.; McCorkle, S.; Puetter, R.; Erickson, E.; Krätschmer, W. Observation of interstellar ammonia ice. *Astrophysical Journal* **1982**, *260*, 141–146.
- (48) Grimme, S.; Bannwarth, C.; Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1–86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
- (49) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angewandte Chemie - International Edition* **2020**, *59*, 15665–15673.
- (50) Meng, L.; Tong, X.; Fan, S.; Cheng, C.; Chen, B.; Yang, W.; Hou, K. A universal generating algorithm of the polyhedral discrete grid based on unit duplication. *ISPRS International Journal of Geo-Information* **2019**, *8*, 146.
- (51) Frisch, M. J., et al. *Gaussian16 Rev. B.01*; Gaussian Inc.: 2016.
- (52) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1327.
- (53) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.* **1999**, *461*, 1–21.
- (54) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD (T)]. *J. Chem. Phys.* **2018**, *148*, 011101.
- (55) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (56) Chung, L. W.; Sameera, W.; 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. *Chem. Rev.* **2015**, *115*, 5678–5796.
- (57) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry* **2006**, *27*, 1787–1799.
- (58) 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.
- (59) Kendall, Jr; Dunning, T., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.



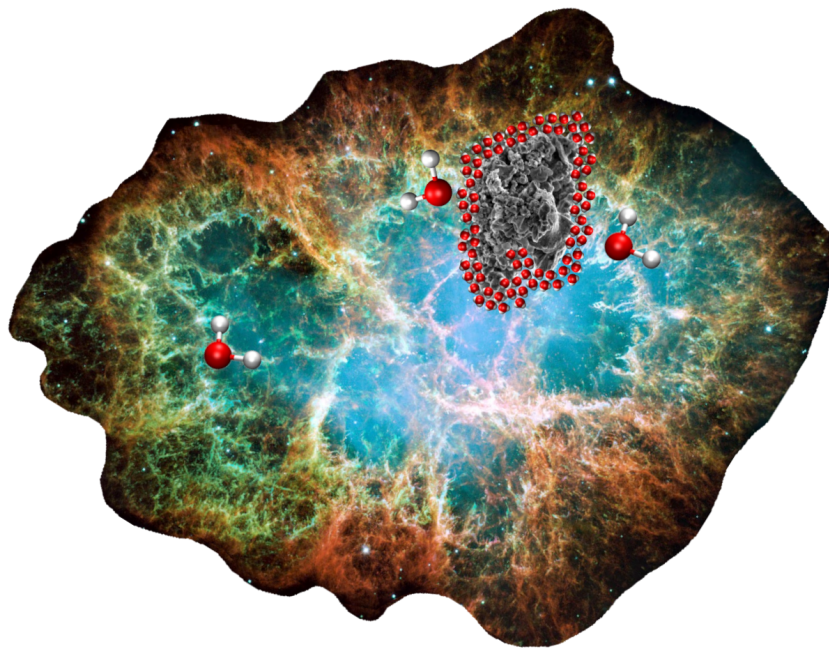
- (60) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (61) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (62) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (63) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.
- (64) 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-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (65) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
- (66) Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Spöner, J. Quantum chemical benchmark study on 46 RNA backbone families using a dinucleotide unit. *J. Chem. Theory Comput.* **2015**, *11*, 4972–4991.
- (67) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
- (68) Hewat, A. W.; Riekel, C. The crystal structure of deuteroammonia between 2 and 180 K by neutron powder profile refinement. *Acta Crystallogr., Sect. A* **1979**, *35*, 569–571.
- (69) Hasegawa, T. I.; Herbst, E. Three-Phase chemical models of dense interstellar clouds- Gas dust particle mantles and dust particle surfaces. *Mon. Not. R. Astron. Soc.* **1993**, *263*, 589.
- (70) Minissale, M.; Aikawa, Y.; Bergin, E.; Bertin, M.; Brown, W. A.; Cazaux, S.; Charnley, S. B.; Coutens, A.; Cuppen, H. M.; Guzman, V. Thermal Desorption of Interstellar Ices: A Review on the Controlling Parameters and Their Implications from Snowlines to Chemical Complexity. *ACS Earth and Space Chemistry* **2022**, *6*, 597.
- (71) Tait, S. L.; Dohnálek, Z.; Campbell, C. T.; Kay, B. D. n-alkanes on MgO (100). II. Chain length dependence of kinetic desorption parameters for small n-alkanes. *J. Chem. Phys.* **2005**, *122*, 164708.
- (72) Freedman, D.; Diaconis, P. On the histogram as a density estimator: L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **1981**, *57*, 453–476.
- (73) McElroy, D.; Walsh, C.; Markwick, A.; Cordiner, M.; Smith, K.; Millar, T. The UMIST database for astrochemistry. *Astronomy & Astrophysics* **2013**, *550*, A36.
- (74) Kruczkiewicz, F.; Vitorino, J.; Congiu, E.; Theulé, P.; Dulieu, F. Ammonia snow lines and ammonium salts desorption. *Astronomy & Astrophysics* **2021**, *652*, A29.
- (75) Martín-Doménech, R.; Caro, G. M.; Bueno, J.; Goesmann, F. Thermal desorption of circumstellar and cometary ice analogs. *Astronomy & Astrophysics* **2014**, *564*, A8.
- (76) Germain, A.; Corno, M.; Ugliengo, P. Computing Binding Energies of Interstellar Molecules by Semiempirical Quantum Methods: Comparison Between DFT and GFN2 on Crystalline Ice. *Computational Science and Its Applications - ICCSA 2021*. 2021; pp 632–645.
- (77) Germain, A.; Ugliengo, P. Modeling interstellar amorphous solid water grains by tightbinding based methods: comparison between GFN-XTB and CCSD (T) results for water clusters. *International Conference on Computational Science and Its Applications*. 2020; pp 745–753.
- (78) Le Gal, R.; Hily-Blant, P.; Faure, A.; Pineau des Forêts, G.; Rist, C.; Maret, S. Interstellar chemistry of nitrogen hydrides in dark clouds. *Astronomy & Astrophysics* **2014**, *562*, A83.
- (79) Jonusas, M.; Leroux, K.; Krim, L. N + H surface reaction under interstellar conditions: Does the NH/NH<sub>2</sub>/NH<sub>3</sub> distribution depend on N/H ratio? *J. Mol. Struct.* **2020**, *1220*, 128736.
- (80) Grassi, T.; Bovino, S.; Caselli, P.; Bovolenta, G.; Vogt-Geisse, S.; Ercolano, B. A novel framework for studying the impact of binding energy distributions on the chemistry of dust grains. *Astronomy & Astrophysics* **2020**, *643*, A155.
- (81) Caselli, P.; Walmsley, C.; Tafalla, M.; Dore, L.; Myers, P. CO depletion in the starless cloud core L1544. *Astrophysical Journal Letters* **1999**, *523*, L165.
- (82) Keto, E.; Caselli, P. Dynamics and depletion in thermally supercritical starless cores. *Mon. Not. R. Astron. Soc.* **2010**, *402*, 1625–1634.
- (83) Aikawa, Y.; Wakelam, V.; Hersant, F.; Garrod, R. T.; Herbst, E. From prestellar to protostellar cores. II. Time dependence and deuterium fractionation. *Astrophysical Journal* **2012**, *760*, 40.
- (84) Sipilä, O.; Caselli, P.; Redaelli, E.; Juvela, M.; Bizzocchi, L. Why does ammonia not freeze out in the centre of pre-stellar cores? *Mon. Not. R. Astron. Soc.* **2019**, *487*, 1269–1282.
- (85) <http://www.jmol.org/>.

## 8.1 ACS Earth &amp; Space Chemistry - Issue Cover



# Theoretical water binding energy distribution and the water snow-line

The hereafter article it was submitted to The Astrophysical Journal in mid December.



## Theoretical water binding energy distribution and snowline in protoplanetary disks

LORENZO TINACCI,<sup>1,2</sup> AURÈLE GERMAIN,<sup>2</sup> STEFANO PANTALEONE,<sup>2</sup> CECILIA CECCARELLI,<sup>1</sup> NADIA BALUCANI,<sup>3</sup> AND PIERO UGLIENGO<sup>2</sup>

<sup>1</sup>Université Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France

<sup>2</sup>Dipartimento di Chimica and Nanostructured Interfaces and Surfaces (NIS) Centre, Università degli Studi di Torino, via P. Giuria 7, 10125, Torino, Italy

<sup>3</sup>Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, 06123 Perugia, Italy

(Received –; Revised –; Accepted –)

Submitted to The Astrophysical Journal

### ABSTRACT

Water is one of the most important and abundant molecule in star forming regions. In protoplanetary disks, where planets and comets form, H<sub>2</sub>O is in the gas or solid form, depending on the dust temperature, *i.e.* the distance from the center, and its binding energy (BE). Not surprisingly, several experimental and theoretical studies of the H<sub>2</sub>O BE have been published. We report new ab initio calculations carried out on a large model of interstellar ice, where we identified 144 different adsorption sites. The BE associated with those sites ranges between 14.2 kJ/mol (1705 K) and 61.6 kJ/mol (7390 K). The distribution of the computed BEs as a function of BE follows a Gaussian peaked at 35.4 kJ/mol (4230 K) with a standard deviation of 9.7 kJ/mol (1160 K). The computed pre-exponential factor ( $\nu$ ) ranges between  $9 \times 10^{12}$  and  $6 \times 10^{14} \text{ s}^{-1}$ . We evaluated the impact of the newly calculated BE and  $\nu$  distributions on the snowline of a generic protoplanetary disk. We found that the region where water is frozen onto the ice is much smaller (a factor 10 smaller radius) than that computed with the single BE (5600 K) and  $\nu$  ( $2 \times 10^{12} \text{ s}^{-1}$ ) values commonly adopted by astrochemical models. Besides,  $\sim 10\%$  of water remains frozen in relatively warm ( $\sim 150$  K) regions, where the single BE and  $\nu$  model would predict a full release of the ice in the gas-phase. This last aspect may have an impact on the quantity trapped in the planetesimals eventually forming rocky planets.

*Keywords:* astrochemistry — ISM: astrochemistry; snowline; protoplanetary disk; cold molecular cloud; molecules

### 1. INTRODUCTION

Water is the second most abundant molecule in the Universe after H<sub>2</sub>. It is “*easily*” formed both in the gas-phase (Elitzur & de Jong 1978; Langer & Graedel 1989; Kaufman & Neufeld 1996; Ceccarelli et al. 1996; Hollenbach et al. 2009) and, with a higher efficiency, on the surface of the sub-micron dust grains that permeate the Inter-Stellar Medium (ISM) (Tielens & Hagen 1982; Dulieu et al. 2010; Oba et al. 2009; Lamberts et al. 2016; Molpeceres et al. 2019) (see also the review by van Dishoeck et al. 2013).

Water vapor is observed in large quantities, with abundances up to about  $2 \times 10^{-4}$  with respect to H<sub>2</sub>, in a wide variety of warm ( $T \geq 50$  K) objects: galactic hot cores and hot corinos, innermost regions of protoplanetary disks, protostellar shocks, nearby galaxies (*e.g.* Cernicharo et al. 1994; Combes & Wiklind 1997; Fischer et al. 1999; Ceccarelli et al. 1999; González-Alfonso et al. 2004; Herpin et al. 2012; Kristensen et al. 2012; Imanishi et al. 2022) and the farthest

galaxies at  $z \geq 3$  (*e.g.* Omont et al. 2013; Yang et al. 2016, 2020; Pensabene et al. 2022)(see also van Dishoeck et al. 2021). Water vapor is also observed, but with a much lower abundance ( $\leq 10^{-8}$ ), in cold ( $T \leq 50$  K) regions such as molecular clouds, prestellar cores and the outer protoplanetary disks (Melnick & Bergin 2005; Caselli et al. 2010; Hogerheijde et al. 2011; Podio et al. 2013). Finally, large quantities of frozen water, up to about  $2 \times 10^{-4}$  in abundance (as in warm objects), are observed in cold objects. The frozen water is believed to constitute icy mantles that envelop the interstellar dust grains, prevalently in Amorphous Water Surface (AWS) (*e.g.* Leger et al. 1979; Gibb et al. 2004; Boogert et al. 2015). In summary, water is copiously present in the molecular ISM, and particularly in the regions forming new solar-like planetary systems, with a rather constant abundance,  $\sim 2 \times 10^{-4}$ , when considering its two different forms, gaseous and solid.

Probably it is not by chance that water is a crucial molecule in terrestrial life, for its multiple roles in the emergence and sustainability of life (*e.g.* Westall & Brack 2018) and proba-



bly its relatively large abundance in the progenitor of the Solar System, whose analogues are the presently forming planetary systems. However, whether water was incorporated, in what quantity and when in the newly formed Earth are still debated issues (*e.g.* Morbidelli et al. 2019). One way to answer this question is to understand the fate of the water synthesised during the formation of solar-like planetary systems (*e.g.* Morbidelli et al. 2016; Hartmann et al. 2017). Since rocky planets as well as asteroids and comets, which may later in the evolution bring water to the newly formed planet, are formed by the aggregation of the interstellar dust grains, the above question translates into how much of the interstellar water remains frozen on the dust grain mantles during the aggregation process. In this respect, the crucial parameter that determines whether water is in the gaseous or solid state is, along with the temperature, its binding energy (hereinafter BE).

In protoplanetary disks, where planets, asteroids and comets are formed, the distance from the star where water freezes-out and becomes a solid is called “snowline” (*e.g.* Hartmann et al. 2017). Therefore, the snowline is a crucial parameter in all models aiming to explain the origin of water on Earth and clearly depends on the assumed water BE distribution. The larger the BE, the larger is the region of the protoplanetary disk where water remains frozen around the dust grains and, therefore, passed to the nascent planets, asteroids and comets. On the contrary, low water BEs would allow a smaller fraction of water being inherited by the planets, asteroids and comets from the protoplanetary disk matter.

Finally, it is worth mentioning that a crucial parameter in the models focused on the terrestrial water origin is represented by the its deuteration, namely the fraction of singly- (or doubly-) deuterated water with respect to the fully hydrogen bearing one, HDO/H<sub>2</sub>O (or D<sub>2</sub>O/H<sub>2</sub>O) (*e.g.* Alexander 2017). The Earth’s oceans have a HDO/H<sub>2</sub>O ratio about ten times that of the original elemental D/H ratio (Lécuyer et al. 1998). There is ample consensus that this enhanced water deuteration has been inherited from the first phases of the Solar System formation (*e.g.* Ceccarelli et al. 2014). It is, therefore, of interest to understand the difference in the BE distribution of HDO with respect to the H<sub>2</sub>O.

While several experimental studies of the water BE have been carried out for more than 30 years (see Sec. 2.1), theoretical studies on (very) small water surfaces have been published only in the last five years (see Sec. 2.2). In this work, we report new ab initio calculations of the water BE distribution, obtained on a simulated icy grain composed of 200 water molecules, which is, in turn, based on a specifically developed open access computer code, called ACO-FROST<sup>1</sup>

(Germain et al. 2022a). To our best knowledge, this is the first theoretical study of the water BEs on such large water amorphous ice, which allows to derive a BE distribution and not only a single value. We have already applied the new grain model and method used here to the adsorption of ammonia and found that its BE distribution presents two peaks. While the first one around 34 kJ/mol (~4000 K) was already known from experimental and theoretical studies, the second smaller peak at about 15 kJ/mol (~1800 K) had never been detected before (Tinacci et al. 2022). It is, therefore, of major interest to understand whether the water BE distribution has a similar behavior.

The article is organised as follows. After reviewing the literature of the previous water BE estimates (Sec. 2), we describe in detail the adopted methodology (Sec. 3) and the results (Sec. 4) of our simulations. We then discuss the comparison of our new BEs with respect to the previous studies (Sec. 5) and the astrophysical implications of the newly computed water BE distribution on the snowline of protoplanetary disks (Sec. 6), and conclude the article (Sec. 7). The data of our study are publicly available on-line (Sec. 8).

## 2. REVIEW OF THE PUBLISHED WATER BE

### 2.1. Experimental studies

Generally, in experimental studies, water molecules are condensed on an inert cold finger and then the rate of desorption upon heating the sample is measured. This technique is called Temperature-Programmed Desorption (TPD). The thermal desorption of a species bound to a substrate can be approximately described by the Polanyi-Wigner equation (Kolasinski 2002), where the rate of desorption  $k(T)$  is given by:

$$k(T) = -\frac{dN}{dT} = N^n \underbrace{\nu \exp\left(-\frac{E_{des}}{T}\right)}_{k_{thdes}}, \quad (1)$$

where  $N$  is the number density of molecules adsorbed on the surface,  $n$  is the order of the desorption,  $k_{thdes}$  is the thermal desorption rate constant,  $\nu$  is the pre-exponential factor (also called prefactor),  $T$  is the temperature of the surface, and  $E_{des}$  is the activation energy for desorption, which is approximately BE.

Experiments differ in the substrate (the cold finger), how the iced sample is condensed and treated (namely the form of the ice), how the desorption carried out in time (temperature ramp) is monitored and how the  $E_{des}$  is extracted from the desorption curve. It is important to emphasize that, regardless the method, BE is evaluated always together with the prefactor.

The substrates used are numerous and have a relatively low impact on the derived BE if more than one monolayer (ML)

<sup>1</sup> [https://github.com/aurelegermain/ACO-FROST\\_grain\\_generator](https://github.com/aurelegermain/ACO-FROST_grain_generator)

**Table 1.** Values of the water BE and prefactor as measured in different experiments using different substrates, ice samples and methods, described in the main text. Units are in kJ/mol and K, for the values in parenthesis.

BE	Prefactor	Substrate	Ice sample	Method	Ref.
40.0±0.1 (4815±15)	$2 \times 10^{12}$ <sup>(b)</sup>	CsI	Amorphous Unannealed Multilayer	IR spectra	1
42.2±0.4 (5070±50)	$2 \times 10^{12}$ <sup>(b)</sup>	CsI	Crystalline Annealed Multilayer	IR spectra	1
49.8±0.8 (5988±101)	$(2.8 \pm 1.0) \times 10^{30}$ <sup>(a)</sup>	Sapphire	Crystalline Multilayer	OIM	2
43.4±2.9 (5222±348)	$1.82 \times 10^{21}$ <sup>(a)</sup>	Gold	Crystalline Multilayer	QCM	3
43.4±2.9 (5222±348)	Not indicated	Graphite (0001)	Crystalline Multilayer	HREELS	4
48.3±1.0 (5809±120)	$(3.99 \pm 0.8) \times 10^{15}$ <sup>(b)</sup>	Ru(001)	Crystalline Multilayer	QMS	5
46.9±0.9 (5640±108)	$(3.26 \pm 0.7) \times 10^{15}$ <sup>(b)</sup>	Ru(001)	Amorphous Multilayer	QMS	5
48.0±0.5 (5773±60)	$1 \times 10^{30 \pm 2}$ <sup>(a)</sup>	Gold	Crystalline Multilayer	QMS	6
46.6 (5600)	$1 \times 10^{30 \pm 2}$ <sup>(a)</sup>	Gold	Amorphous Multilayer	QMS	6
58.2±0.8 (6995±101)	$1 \times 10^{32.6 \pm 0.3}$ <sup>(a)</sup>	Ru(001)	Crystalline Multilayer	OIM	7
39.9±2.9 (4799±349)	$4 \times 10^{26}$ <sup>(a)</sup>	HOPG	Crystalline Multilayer	QMS	8
46.0±3.0 (5533±361)	$9 \times 10^{14 \pm 1}$ <sup>(b)</sup>	HOPG	? Monolayer	QMS	9
46.0±3.0 (5533±361)	$9 \times 10^{14 \pm 1}$ <sup>(b)</sup>	HOPG	? Multilayer	QMS	9
39.9±0.8 (4799±96)	$1 \times 10^{27 \pm 1}$ <sup>(a)</sup>	HOPG	Crystalline Multilayer	QMS	10
39.9 (4800)	$1 \times 10^{12}$ <sup>(b)</sup>	Amorphous silicate	Sub-monolayer	QMS	11
42.9±0.5 (5165±55)	$32.1_{-8.5}^{+12.8} \times 10^{27}$ <sup>(a)</sup>	KBr	Amorphous Multilayer	QMS	12
34-36 (4090-4330)	$1 \times 10^{12}$ <sup>(b)</sup>	Amorphous silica	Sub-monolayer (0.4)	QMS	13
49.3±2 (5930±240)	$1 \times 10^{28}$ <sup>(a)</sup>	Amorphous silica	? Multilayer	QMS	13
48.7±0.3 (5857±36)	$1 \times 10^{12}$ <sup>(a)</sup>	KBr	Multilayer	QCM	14
45.3±1.9 (5454±232)	Not indicated	Gold	? Multilayer	mm/smm	15
64.2±0.3 (7721±36)	$1 \times 10^{17}$ <sup>(a)</sup>	Copper	Amorphous Multilayer	QCM	16

NOTE—BE are reported in kJ/mol and K (in parenthesis). (a) refers to zero order desorption, in units of mol/cm<sup>2</sup>s<sup>-1</sup>; (b) refers to first order desorption, in units of s<sup>-1</sup>; “?” replace the ice phase, amorphous or crystalline, if it was not explicitly indicated in the cited articles. *References:* [1] Sandford & Allamandola (1988); [2] Haynes et al. (1992); [3] Sack & Baragiola (1993); [4] Chakarov et al. (1995); [5] Speedy et al. (1998); [6] Fraser et al. (2001); [7] Smith et al. (2003); [8] Bolina et al. (2005); [9] Ulbricht et al. (2006); [10] Brown & Bolina (2007); [11] Dulieu et al. (2013); [12] Martín-Doménech et al. (2014); [13] Collings et al. (2015); [14] Potapov et al. (2018); [15] Yocum et al. (2019); [16] Rosu-Finsen et al. (2022).

of water molecules (1 ML meaning that the substrate is completely covered by water molecules) is deposited on the substrate, as the water molecules will be desorbed from other water molecules. For a coverage inferior to 1 ML, the BEs could be impacted depending on the wetting of the substrate, *i.e.* if the water molecules are evenly spread over the substrate or if they coagulate and form multilayered clusters as sometime reported (Collings et al. 2015). On the other hand, when the coverage is larger than 1 ML, the BE correlates with the coverage regime as pointed out by Rosu-Finsen et al. (2022). In their article, beside the suggested value report in our Tab. 1, these authors report a BE of 50.3 kJ/mol and a prefactor of  $1 \times 10^{30}$  cm<sup>-2</sup>s<sup>-1</sup> for a water film thickness of 53 nm, and a BE of 69.2 kJ/mol and a prefactor of  $1 \times 10^{35}$  cm<sup>-2</sup>s<sup>-1</sup> for a 101 μm film.

Table 1 lists 21 different water-BE measurements, obtained with 9 different substrates: CsI for Caesium Iodide (Sandford & Allamandola 1988), Sapphire (Haynes et al. 1992), Cop-

per (Rosu-Finsen et al. 2022), Gold (Sack & Baragiola 1993; Fraser et al. 2001; Yocum et al. 2019), Graphite cut in the (0001) plane (Chakarov et al. 1995), Ru(001) for Ruthenium substrate cut in the (001) plane (Speedy et al. 1998; Smith et al. 2003), HOPG for highly oriented pyrolytic graphite (Bolina et al. 2005; Ulbricht et al. 2006; Brown & Bolina 2007), Amorphous silicate (Dulieu et al. 2013; Collings et al. 2015), and KBr for Potassium bromide (Martín-Doménech et al. 2014).

Depending on how water molecules are condensed onto the sample, the ice can be crystalline or amorphous, multilayered, monolayered, or sub-monolayered, and annealed or unannealed. In Table 1, annealed and unannealed indicate when the sample was heated and subsequently allowed to cool down (annealed), resulting in a phase transition from amorphous to crystalline if the ice sample was originally amorphous and if the annealing was done at high enough temperature.



Also, in Table 1, we indicate the different properties of the ice samples only if they were explicitly reported in the relevant articles. For samples where “*crystalline*” or “*amorphous*” is not reported, we could have assumed one or the other depending on the way water was deposited on the sample, but amorphous water ice undergoes a phase transition to crystalline ice at around 140–150 K (Jenniskens et al. 1995; Löfgren et al. 2003), *i.e.* around the peak desorption temperature of water, which implies that the ice sample could be amorphous, crystalline, or a mixture of the two. As H<sub>2</sub>O is the most important component of the ices, a multilayer approach is the most popular one.

The desorption is monitored via Infrared Spectroscopy (especially the earliest works used this technique), optical interference methods (OIM), Quartz Crystal Microbalance and their resonance frequency (QCM), high-resolution electron energy loss spectroscopy (HREELS), and, nowadays mass spectroscopy with help of quadrupole mass spectrometers (QMS). Results from a new millimeter/submillimeter (mm/submm) technique (Yocum et al. 2019) are also reported in Table 1.

Finally, different authors use either the zero or first order of the Polanyi-Wigner equation.

Table 1 summarises the different estimates of the water BE values along with the different characteristics of the experiments, as briefly described above. Globally, the measured water BE ranges between 34 to 64 kJ/mol, where this range includes a variety of ice samples, multilayers regimes, and substrates, not to mention a different way to extract the BE value (zero or first order) and prefactors. When considering only the experiments on amorphous multilayers samples, which are the closest to our simulations, where BE and the prefactor are estimated at the first order, the measured BE range is smaller, between 40 to 47 kJ/mol (or  $\sim 4800$  to  $5640$  K), with a prefactor between  $\sim 10^{12}$  and  $\sim 10^{15} \text{ s}^{-1}$ , with the outsider value of 64.2 kJ/mol (7721 K) obtained by Rosu-Finsen et al. (2022).

## 2.2. Theoretical studies

In theoretical studies, the way the icy grain surface is modeled and the method used to compute the binding energy is of high importance. In Table 2, we report the different values of the water BE obtained from theoretical studies. The computed BE ranges from 22 to 70 kJ/mol. Contrary to experimental studies, computational studies of water are scarce. In Wakelam et al. (2017) the icy grain surface is modeled using a cluster composed of one water molecule and the BE value is computed using M06-2X coupled to an aug-cc-pVTZ basis set. They derive a water BE equal to 38 kJ/mol. Evidently, this model cannot represent accurately the interstellar

icy grains, whose sizes are sub-micrometric and whose state is thought to be mainly amorphous (Boogert et al. 2015). To correct for this approximation, Wakelam et al. (2017) propose to apply a coefficient of proportionality derived by comparing TPD experiments of different species against the computed BE for a single water molecule cluster. By applying this coefficient to their computed BE, Wakelam et al. (2017) obtained the aforementioned water BE of 38 kJ/mol (4600 K), relatively close to the values obtained from experiments on amorphous ice (see Subsec. 2.1).

Das et al. (2018) adopted a similar approach (MP2/aug-cc-pVDZ), but decided to study the impact on the BE of the cluster size by using water models ranging from one to six molecules. They showed that increasing the ice-model size can give results closer to the experimental ones, without the help of a correction coefficient. That said, while providing interesting methods for simple BE computations, the two above works do not take into account the importance of the hydrogen bonds cooperativity present in large water clusters adsorption, bringing sites exhibiting strong BE (as was pointed out before by Ferrero et al. 2020).

Ferrero et al. (2020) used two different periodic (crystalline and amorphous) slabs to model the icy grains. The BE value for the crystalline water model is almost 10 kJ/mol higher than the maximum value found using the amorphous ice model. This difference can be explained by the different length of hydrogen bond chains in the crystalline ice (in principle infinite) with respect to the amorphous one in which they break due to the loss of long range order. The BEs with the amorphous model, was between 30 and 51 kJ/mol (3600 and 6100 K), closer to both the experimental ones (40–47 kJ/mol) and the previous computational results.

Duflot et al. (2021) studied the adsorption of water using an ONIOM hybrid method similar to the one presented in this work. As Ferrero et al. (2020), they adopted both crystalline and amorphous ice models. Both types of models are composed of a fixed geometry low QM part and an unfixed high QM part. The two crystalline models have approximately 160 water molecules in the low QM zone ( $\omega\text{B97X-D/6-31+G}^{**}:\text{PM6}$ ) and 20 water molecules in the high QM zone (CBS/DLPNO-CCSD(T):PM6). The amorphous slab is generated by classical molecular dynamics simulations. Several adsorption are performed on it and, for each adsorption, approximately twenty water molecules around the adsorbate are included in the high QM zone (CBS/DLPNO-CCSD(T):PM6) and 120 to 150 molecules around it are treated as a low QM zone ( $\omega\text{B97X-D/6-31+G}^{**}:\text{PM6}$ ). For both models the obtained results are in good agreement with the experimental BEs listed in Table 1.

Bovolenta et al. (2022) adopted 20 amorphous clusters, of 22 water molecules each, coming from high temperature molecular dynamics followed by annealing at 10K. For each

**Table 2.** Values of the water BE as computed in different theoretical studies using various methods and ice grain models. Units are in kJ/mol and K, for the values in parenthesis.

BE	Ice model	Method	Ref.
38.2 (4600)	Water monomer	M06-2X/A-VTZ <sup>a,c</sup>	1
22.2 (2670)	Water tetramer	MP2/A-VDZ <sup>b,c</sup>	2
34.6 (4166)	Water hexamer	MP2/A-VDZ <sup>b,c</sup>	2
59.9 (7200)	Crystalline periodic	B3LYP-D3/A-VTZ <sup>a</sup>	3
<b>Min:</b> 30.0 (3605)			
<b>Max:</b> 50.8 (6111)	Amorphous periodic	B3LYP-D3/A-VTZ <sup>a</sup>	3
$\mu$ : 41.1 (4941)			
<b>Min:</b> 33.1 (3980)			
<b>Max:</b> 51.7 (6220)	Crystalline cluster	ONIOM(DLPNO-CCSD(T)/CBS // $\omega$ B97X-D/6-31+G**: <b>PM6</b> )	4
$\mu$ : 40.7 (4897)			
<b>Min:</b> 34.6 (4166)			
<b>Max:</b> 54.4 (6545)	Amorphous ice slab	ONIOM(DLPNO-CCSD(T)/CBS // $\omega$ B97X-D/6-31+G**: <b>PM6</b> )	4
$\mu$ : 45.1 (5419)			
<b>Binding mode 1:</b>			
$\mu$ : 22.7 (2725)			
$\sigma$ : 3.7 (449)			
<b>Binding mode 2:</b>			
$\mu$ : 34.0 (4087)	Amorphous clusters	$\omega$ – PBE/def2-TZVP	5
$\sigma$ : 2.8 (338)			

NOTE—The Total BE average excluding the first three values (not accurate) is:  $50.8 \pm 10.7$  kJ/mol ( $6110 \pm 1287$  K). <sup>a</sup> A-VTZ refers to aug-cc-pVTZ. <sup>b</sup> A-VDZ refers to aug-cc-pVDZ. <sup>c</sup> refers to cases in which BSSE and ZPE corrections were not applied.  $\mu$  is the distribution mean and  $\sigma$  the standard deviation, **Max** the maximum value and **Min** the minimum value. *References:* [1] Wakelam et al. (2017); [2] Das et al. (2018); [3] Ferrero et al. (2020); [4] Dufflot et al. (2021); [5] Bovolenta et al. (2022).

cluster, the adsorbate randomly samples the surface until at least 225 binding energies structure are computed through various selection processes. Geometries are first subjected to a cost-effective optimization (HF-3c/MINIX), then a single-point computation ( $\omega$  – PBE/def2-TZVP with ZPE and BSSE correction) is performed to obtain more accurate binding energies. Then, the BEs are separated into two modes depending on whether the water molecules make 1 or 2 hydrogen bonds with the surface (respectively Binding mode 1 and Binding mode 2 in Table 2). Finally a BE distribution is obtained and fitted using a Gaussian function, whose mean and standard deviation values are reported in Table 2.

### 3. METHODOLOGY

The methodology adopted in this article in order to compute the BE distribution is based on previous works by our group (Tinacci et al. 2022; Germain et al. 2022a).

The BE (positive for a bounded system) is defined as the opposite of the interaction energy ( $\Delta E$ ), which is the difference between the energies of complex (between the grain and

the adsorbate, *i.e.*  $E_c$ ) and the sum of energies of the isolated adsorbate ( $E_{ads}^{iso}$ ) and the isolated grain ( $E_{grn}^{iso}$ ), as follows:

$$BE = -\Delta E = E_{ads}^{iso} + E_{grn}^{iso} - E_c . \quad (2)$$

The BE can be decomposed in two terms: (i) the pure electronic interaction ( $BE_e$ ) corrected, if needed, for the Basis Set Superposition Error (BSSE), and (ii) the geometry deformation energy ( $\delta E_{def}$ ) contribution. When each electronic energy is corrected by the zero point energy (ZPE) term derived from the harmonic frequency calculation, the BE can be corrected for these contributions giving the Binding Enthalpy at 0 K (BH(0)), as:

$$BH(0) = \underbrace{BE_e - \delta E_{def}}_{BE} - \Delta ZPE . \quad (3)$$

where  $\Delta ZPE$  is the difference between each ZPE terms. The mathematical details are reported in the appendix A.

#### 3.1. Water-ice grain model and strategy to compute BEs

The initial BE computation is performed in the same way as in our previous work (Germain et al. 2022a). A cluster of 200 water molecules is used as model for the interstellar icy grains (see 200-water grain model at [https://aurelegermain.github.io/JSmol\\_grain/](https://aurelegermain.github.io/JSmol_grain/)) (Germain et al. 2022b). Using the ACO-FROST code (Germain et al. 2022c), we generate a spherical grid of 162 points evenly spaced around the cluster. Each of these points is replaced by three water molecules with a different random orientation. After projecting every water molecule between 2.5 Å and 3.0 Å from the grain we obtain a total of 486 starting adsorption positions.

The strategy to compute the water BE distribution consists of five consequently steps.

- (i) We fix the geometry of the grain model and run a GFN2 geometry optimisation of the adsorbed water molecule only for each starting position.
- (ii) Starting from point (i), all molecules within a 5 Å region around the relaxed adsorbed molecule were allowed to relax at GFN2 level.
- (iii) On the resulting optimized (ii) structure, a further check is carried out to ensure the 5 Å region has not changed its number of water molecules due to the geometry relaxation. In case that happened, the (ii)-(iii) cycle is repeated until no compositional changes in the 5 Å region is detected.
- (iv) On the final (iii) structures, the two-layer ONIOM(QM:SQM) model chemistry is carried out. The Model zone (corresponding to the 5 Å region) is treated with a high level quantum mechanical method, while both the Model zone and the whole cluster (the Real system) are treated at GFN2 level. the resulting energies are then combined in a subtractive way as defined by the ONIOM method.
- (v) During the ONIOM geometry optimization all atoms outside the Model zone are kept fixed; mechanical embedding and polarization of the Model zone are taken into account by activating the corresponding keywords in the computer program. In the frequency calculations (within the harmonic approximation), only the normal modes related to the nuclei inside the Model zone are taken into account, keeping fixed all the other nuclei. As for steps (ii) and (iii) we ensure that the Model zone remains with the same number of water molecules during the geometry optimization. The isolated grain surface ( $E_{gm}^{iso}$ ) is computed following our previous approach (called “TPD”) which resembles closely the physics of the experimental TPD process (Tinacci et al. 2022), relaxing the QM system after withdrawing the adsorbed water molecule.

In order to deal with the large number of samples, we simplify the previously adopted procedure (Tinacci et al. 2022) to save computer resources while keeping the accuracy of the computed BE as high as possible.

### 3.2. Computational methods

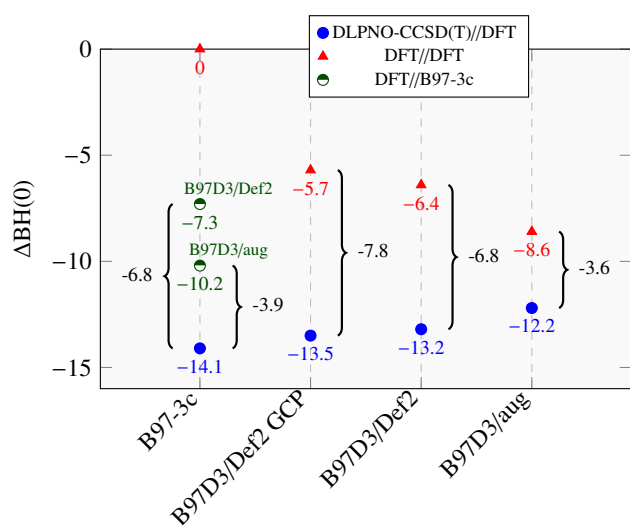
After the preliminary geometric optimization via GFN2-xTB (Grimme et al. 2017; Bannwarth et al. 2019), discussed above, we refine each BE samples via multilevel ONIOM (Mayhall et al. 2010) QM:QM2 approach implemented in ORCA (v.5.0.2) (Neese 2018). The B97-3c (Brandenburg et al. 2018) method is used for the high level ONIOM Model zone for the geometries and the related vibrational harmonic frequencies. Default settings for the SCF and the geometry optimizer are adopted. In the few cases in which imaginary frequencies are found, the optimization is restarted with a tighter DFT integral grid (Grid3 in ORCA) ensuring to achieve a PES minimum. Subsequently, the ONIOM energies are refined using the DLPNO-CCSD(T) (Guo et al. 2018) as a method to treat the Model zone. This method is coupled with the aug-cc-pVTZ (Kendall Jr 1992) as primary basis set, while the aug-cc-pVTZ/C (Weigend et al. 2002) is used as auxiliary basis set for the resolution of the identity (RI) approximation in electron repulsion integrals. The DLPNO calculations are carried out with a tight-PNO set up and the default settings for the SCF, in order to achieve higher accuracy. Due to the possible non-negligible error that stems from the Basis Set Superposition Error (BSSE) in post Hartree-Fock method, the dimer energy at DLPNO-CCSD(T) level is corrected with the Counterpoise method (Boys & Bernardi 1970).

Rendering of molecule images are obtained via the VMD software (Humphrey et al. 1996), while the graphics elaboration and plots via TikZ and PGFPlots  $\LaTeX$  packages.

#### 3.2.1. B97-3c benchmark

To validate the accuracy of B97-3c functional we adopted a twofold strategy: we compare the BE computed at B97-3c level with respect to the less approximate B97D3 method (Grimme 2006; Grimme et al. 2011) coupled with the aug-cc-pVTZ basis, whose performance was already extensively studied in our previous article (Tinacci et al. 2022).

We limited the accuracy check to one  $\text{H}_2\text{O}$ /grain system as a test case, for which we compare structures and  $\text{BH}(0)$ s at B97D3 level with different basis sets: (i) B97D3/aug-cc-pVTZ, (ii) B97D3/Def2-TZVP, (iii) B97D3/Def2-TZVP with geometrical CounterPoise (gCP) (Kruse & Grimme 2012) correction. All the DFTs structures and their related ZPE corrections are obtained starting from the ONIOM(B97-3c:xTB-GFN2) optimized structure. The usual single point energy refinement at DLPNO-CCSD(T)/aug-cc-pVTZ level is performed on the minimum obtained by each different setup. All BEs are corrected for their BSSE. The results of



**Figure 1.** Difference  $\Delta BH(0)$  between the  $BH(0)$  B97-3c reference value (67.8 kJ/mol) and the  $BH(0)$ s computed with the reported QM methods in the ONIOM(QM:xTB-GFN2) framework. All the BEs are BSSE (when needed) and ZPE corrected. With Def2 and aug we mean the Def2-tZVP and aug-cc-pVTZ basis set, respectively. All quantities are in kJ/mol.

Fig. 1 show that: (i) the DLPNO-CCSD(T) BE is almost insensitive to the level of the optimized geometry, varying less than 2 kJ/mol across the considered cases; (ii) plain DFT results converge within 4 kJ/mol to the DLPNO-CCSD(T) results; (iii) B97-3c BE are off by a significant amount, while providing a robust and quick method for the geometry optimization.

### 3.2.2. Accuracy of the DLPNO-CCSD(T) approximation

As pointed out recently, the DLPNO approximation to the CCSD(T) method could have significant additive errors with increasing the size of the system (Sandler et al. 2021; Pavošević et al. 2017; Al-Hamdani & Tkatchenko 2019). The main error, excluding the Basis Set incompleteness error and the algorithm to solve the triplets (T), resides primarily in the two parameters that control the size of the correlation space.

Altun et al. (2020) provided a solution by using an easy extrapolation procedure, similar but conceptually different to the two-point CBS strategy, to recover the Complete PNO Space (CPS) and correct the additive error on the electron pair truncation correlation energy. Recently, they also provided an extensive study of the impact of such correction (Altun et al. 2021) and the dependence of DLPNO-CCSD(T) error related to RI approximation and the BSSE. Unfortunately, the impact of this correction and error on our target quantity, BE, is missing in their articles. Indeed, due to its intrinsic subtractive nature, this quantity can have a compensating error.

We carried out a benchmark, reported in Appendix D with an extended explanation of the adopted method. As expected, the BE seems to have a compensation error. Furthermore the BSSE, that has a non-negligible impact in post Hartree-Fock method, is not affected by the correlation error, the treatment of the triplet and the RI approximation. The RI error is also negligible.

In conclusion, the CPS strategy is probably useful in larger systems than ours. A scheme like CPS(7/6) and CPS(6/5) and different combination of primary and secondary basis sets could be used to save computational time while keeping the accuracy still high enough. For these reasons we have decided to keep our DLPNO-CCSD(T) setup, and do not apply the CPS strategy.

### 3.3. Procedure to obtain unique BE sites

To obtain unique BE sites, we developed a pruning procedure on the obtained ONIOM complex structures. This procedure is requested by considering the possibility to arrive in the same PES minima, starting from slightly different geometrical initial conditions. The redundancy cases are identified in the following way:

- (i) we calculate all the combinations between BEs Model zone structure that have the same number of atoms,
- (ii) for each combination, we align the two structures and compute the RMSD and  $|\Delta BE|$ ,
- (iii) if the two structures have an RMSD below 0.5 Å and  $|\Delta BE_e| < 0.75$  kJ/mol are considered identical.

The 0.5 Å and 0.75 kJ/mol threshold parameters were selected after an exhaustive inspection of the possible redundant BE sites.

### 3.4. The desorption rate prefactor

As pointed out by previous articles (Minissale et al. 2022; Ferrero et al. 2022; Tinacci et al. 2022), attention should be given to the way in which the prefactor in the desorption rate ( $k_{thdes}$  in Eq. 1) is computed. As in the aforementioned articles, we here adopted the Tait et al. (2005) formula:

$$\nu(T) = \frac{k_B T}{h} \left( \frac{2\pi m k_B T}{h} \right) A \frac{\sqrt{\pi}}{\sigma h^3} (8\pi^2 k_B T)^{\frac{3}{2}} \sqrt{I_x I_y I_z}, \quad (4)$$

where  $k_B$  is the Boltzmann constant,  $m$  the mass of the molecule,  $h$  the Planck constant,  $A$  is the surface area per adsorbed molecules (usually assumed to be  $10^{-19} \text{m}^2$ ),  $I_i$  is the  $i$ -esimal adsorbate principal moment of inertia, and  $\sigma$  is the symmetry adsorbate rotation factor. For  $\text{H}_2\text{O}$ , the principal moments of inertia are 1.83, 1.21, 0.62 a.m.u. Å<sup>2</sup>,  $\sigma$  is 2 and  $m$  is 18 a.m.u.

In general, the Tait et al. (2005) prefactor does not depend on the adsorption site due to the immobile particle approximation that sets the vibrational partition functions contribution ( $q_{\text{vib}}^{\text{TST}}$ ) to unity. To take into account the latter, Eq. 4 can



be multiply by  $q_{\text{vib}}^{\text{TST}}(\text{T})$ , that includes the contribution of the harmonic vibrational partition function of the isolated adsorbate,  $q_{\text{ads}}^{\text{iso}}$ , and that of the isolated surface,  $q_{\text{grn}}^{\text{iso}}$ , divided by the harmonic vibrational partition function of complex,  $q_c$ , as follows:

$$q_{\text{vib}}^{\text{TST}}(\text{T}) = \frac{q_{\text{ads}}^{\text{iso}} q_{\text{grn}}^{\text{iso}}}{q_c} \approx \exp\left(-\frac{k_B \text{T}}{h\nu_f}\right). \quad (5)$$

The last term has been derived by fitting through an exponential function the whole  $q_{\text{vib}}^{\text{TST}}(\text{T})$  using the  $\nu_f$  as a free parameter.

## 4. RESULTS

### 4.1. Binding Energy distribution

As described in the previous section (§ 3.3), the structural pruning and cleaning gives 144 unique BE values. Note that all the obtained structures are PES minima, *i.e.* imaginary frequency are not present. Due to the very low ( $\leq 100$  K) T in the dense clouds, we only considered the ZPE correct BE (*vide supra*) and no further correction was included as it is less than about 3 kJ/mol (see Appendix C). In general, the differences between the two sets of computations are below (1-3 kJ/mol) the commonly accepted chemical accuracy (5 kJ/mol) and consequently reside in the margin of error of our computations.

The distribution of the final BH(0) values (Eq. 3), organized in a bin width following the Freedman Diaconis estimator (Freedman & Diaconis 1981), is shown in Fig. 2. With respect to the ammonia distribution, which presents an asymmetric BE distribution (Tinacci et al. 2022), H<sub>2</sub>O has a symmetric distribution with a data dispersion ranging from 11.6 to 64.2 kJ/mol. The distribution is well reproduced by a unnormalized Gaussian function:

$$\mathcal{G}(x, \alpha, \mu, \sigma) = \frac{\alpha}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (6)$$

with the following fitted parameters: mean ( $\mu$ ) of 35.4 kJ/mol, standard deviation ( $\sigma$ ) of 9.73 kJ/mol and the prefactor ( $\alpha$ ) of 763.76 kJ/mol.

### 4.2. Energetic & Structural information

As discussed in the Section 3, the BH(0) can be decomposed in different contributions, which help to understand the form of the BE distribution. Fig. 3 shows the results of the BE decomposition in  $\Delta\text{ZPE}$ ,  $\text{BE}_e$  and  $\delta\text{E}_{\text{def}}$ , which can be summarized as follow:

- (i) the  $\Delta\text{ZPE}$  contribution is rather sparse, as only a rough correlation exists with the  $\Delta\text{ZPE}$  ranging from 5-15 kJ/mol;
- (ii) the  $\text{BE}_e$  correlates significantly with the BH(0). As expected the values are much higher than the BH(0) as  $\text{BE}_e$  does not include the deformation energy;

- (iii) the  $\delta\text{E}_{\text{def}}$  does not correlate with the BH(0) showing a rather large spread of values from 0-50 kJ/mol.

More insight of the nature of the chemical interactions, specifically the network of H-bonds involving the adsorbed water molecule with the surface molecules, can be extracted from the structural feature of the adsorbed complex is shown in Fig. 4.

We categorized each case in terms of the number of H-bonds in which the adsorbed water is involved. Fig. 4 (left) shows that in the vast majority of cases (90 adsorption sites) the adsorbed water enjoyed 2 H-bonds, spanning a  $\text{BE}_e$  average range of 58 kJ/mol. When water is involved in 1 H-bond, only 12 adsorption sites are involved and the  $\text{BE}_e$  values are on the lower site of the  $\text{BE}_e$  distribution ( $\text{BE}_e < 60$  kJ/mol and average range of 37 kJ/mol). The case of 3 H-bonds shows a  $\text{BE}_e$  range peaked at about 69 kJ/mol, with the lowest  $\text{BE}_e$  value at 42 kJ/mol. These data are in agreement with the obvious expectation that the higher the number of H-bonds the higher the average  $\text{BE}_e$  values of the corresponding distribution. Another common concept about the H-bond geometrical features is the anti-correlation between the H $\cdots$ O bond length and the O-H $\cdots$ O angle. Fig. 4 (right) shows, indeed, the expected trend, particularly for the cases involving 2 H-bonds. In all cases, the O-H $\cdots$ O angle ranges 120-180 degrees.

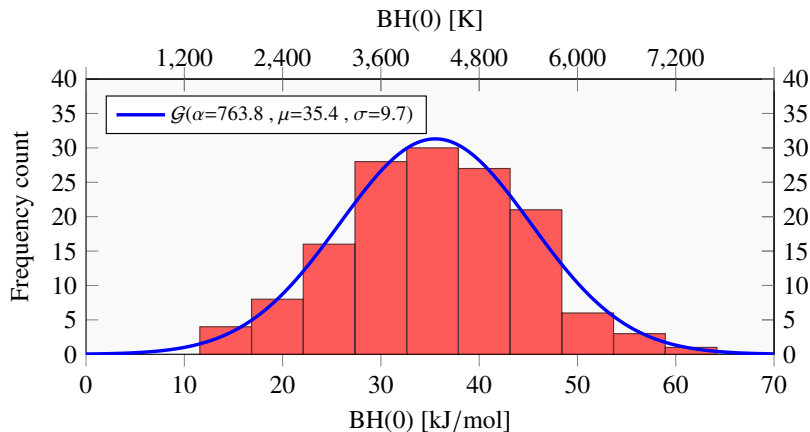
### 4.3. BE distribution of singly and doubly deuterated water

The degree of deuteration of water, namely the HDO/H<sub>2</sub>O abundance ratio, is very important in understanding the origin of the terrestrial water and, more generally, to elucidate where and when water molecules are formed in astrophysical objects (see Introduction). For this reason, we studied the BE distribution of singly- (HDO) and doubly- (D<sub>2</sub>O) deuterated water which are only affected by the difference in the  $\Delta\text{ZPE}$  with respect to the un-deuterated case.

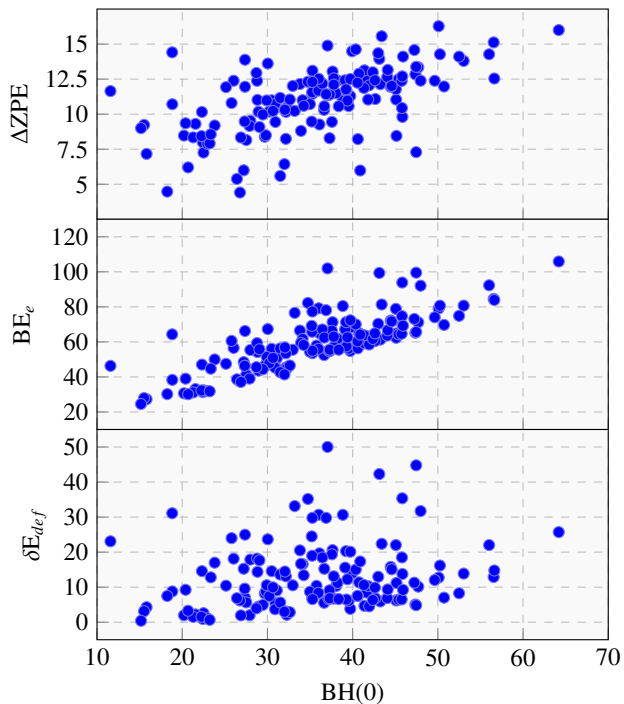
Figure 5 shows the correlation between the  $\Delta\text{ZPE}$  for the non-deuterated water and the singly-/doubly-deuterated ones. For the case of HDO the arithmetic average of the two symmetric deuterated cases is reported. Figure 5 shows a lowering of the  $\Delta\text{ZPE}$  by a factor of 0.908 or 0.826 for one or two D substitutions, shifting BH(0) by 1 or 2 kJ/mol, respectively, a very modest decrease compared to the H<sub>2</sub>O case.

### 4.4. The desorption rate prefactor

Figure 6 shows the desorption rate prefactor, DSP, as a function of the temperature, computed with the Tait et al. (2005) formula (Eq. 4) and the one corrected for the harmonic partition functions, named CDSP (Eq. 5: see below), respectively. CDSP varies between  $\sim 7 \times 10^{13}$  and  $\sim 3 \times 10^{15}$  s<sup>-1</sup> within the T range 50-250 K, *i.e.* about a factor 40. By assuming a desorption temperature of 120 K, as it is often the case in astrochemical models, the CDSP is  $\sim 8 \times 10^{14}$  s<sup>-1</sup>,



**Figure 2.** BSSE corrected BH(0) distribution at DLPNO-CCSD(T) level. Structures and ZPEs are calculated at the ONIOM(B97-3c:xTB-GFN2) level. The solid blue curve is the  $\mathcal{G}(\text{hist}(\text{BH}(0)), \alpha, \mu, \sigma)$  unnormalized Gaussian best fit function for the histogram (see Eq. 6).



**Figure 3.** Correlation plots between BH(0) and  $\Delta\text{ZPE}$ ,  $\text{BE}_e$ ,  $\delta\text{E}_{def}$ . All the quantity are units of kJ/mol. All BH(0) and  $\text{BE}_e$  are BSSE corrected. BH(0) decomposition is illustrated shortly in Eq. 3 and extensively in appendix A.

while the DSP is about  $3 \times 10^{15} \text{ s}^{-1}$ , about a factor 4 smaller than the CDSP. When considering the Hasegawa et al. (1992) rough estimate our CDSP is about two order of magnitude larger at 120 K.

Figure 7 shows the average (over the 144 adsorption cases) and its 95% confidence level of the correction term from the harmonic partition function as a function of temperature. As already pointed out (*vide supra*) at 120 K the correction fac-

tor is around 4, but can become up to 10 at higher temperature.

The correction factor (Eq. 5) as a function of the temperature is shown in Fig. 7. The figure shows the average of the correction factor obtained for all the 144 computed BE sites and the 95% confidence level band. In general, the correction due to the inclusion of the harmonic vibrational partition functions can be larger than one order of magnitude, and it is larger at larger temperatures, as expected. Finally, by fitting against the average values using the exponential expression of Eq. 5 we derive an effective frequency of about  $82 \text{ cm}^{-1}$ .

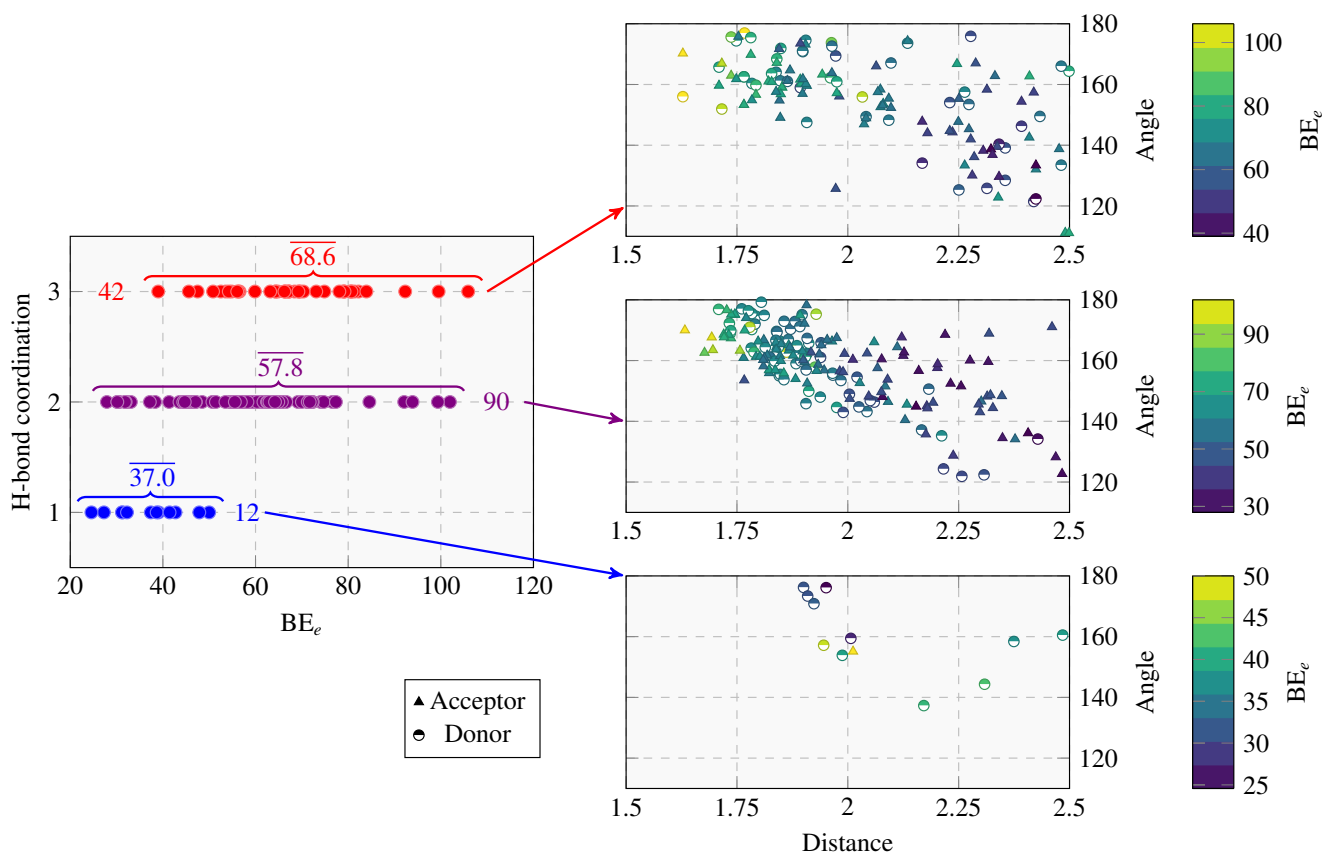
## 5. COMPARISON WITH THE LITERATURE DATA

### 5.1. Experimental data

As discussed in Sec. 2.1, the experimentally measured BE on amorphous water ice lie between 40 to 47 kJ/mol (or  $\sim 4800$  to  $5640 \text{ K}$ ), with a prefactor between  $\sim 10^{12}$  and  $\sim 10^{15} \text{ s}^{-1}$ , if one excludes the outsider value of 64 kJ/mol (7720 K) obtained by Rosu-Finsen et al. (2022). The BE values are clearly larger than the BE average value of Fig. 2, *i.e.* 35.4 kJ/mol.

The comparison between computed and experimentally derived BE is not at all straightforward. What our approach (and also all the other theoretical estimates of Table 2) simulates, is the heat released,  $\Delta\text{H}(0)$ , when a single water molecule is adsorbed on an icy model grain. The resulting BE distribution maps the different amount of heat released by adsorption at different grain sites. These  $\Delta\text{H}(0)$  data would therefore be directly comparable to the heat measured in a microcalorimetric experiment, as for instance, when water is adsorbed at an amorphous silica surface or in an acidic zeolite (Bolis et al. 2006). In the astrochemical context, TPD is the method of choice, for many technical and experimental conditions not easily amenable in a microcalorimeter. In TPD, however, as previously described (*vide supra* Subsec. 2.1),





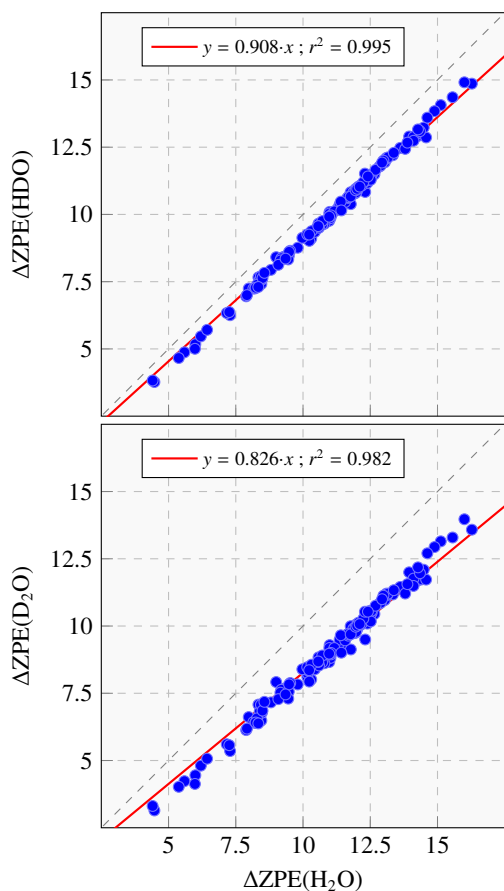
**Figure 4.** Correlation plots related to the H-bond structural features with respect to the pure electronic BE interaction ( $BE_e$ ). All quantities are in Å, kJ/mol and Degree. Left:  $BE_e$  versus number of H-bonds; over-bar numbers as average of the  $BE_e$  distribution; plain numbers refer to the number of corresponding cases with the specific number of H-bonds. Right: correlation plots of the OH...O distance versus the O-H...O angle. Different symbols refer to water acting as H-bond donor/acceptor, respectively.

the BE is worked out numerically from the Polanyi-Wigner equation, together with the prefactor. A one by one comparison with the computed BE is, therefore, only meaningful when also the theoretically computed prefactor is of the same order of magnitude of that derived from the fitted experimental data. Furthermore, in the simulation, the icy grain structure is almost unaltered during the adsorption, but for few water molecules surrounding the adsorbate. This is not the case during the temperature ramp adopted in TPD, where the heat transferred to the bulk grain can significantly restructure the surface site distribution altering the BE values along the experiment.

For the specific case of water BE, another subtle difference bias the comparison with the computed data. Indeed, computed BE are relative to a single specific water molecules used to probe different adsorption sites. In the TPD, it is almost impossible, either to select specific water molecule to be desorbed or to ensure that the most exterior water molecules are desorbed with respect to more hidden ones. In the latter case, due to the many H-bonds involved, the BE values would be closer to the latent water of liquefaction (44

kJ/mol). Insight that can be the case comes from the work by Collings et al. (2015) in which they measured a BE of 35 kJ/mol and a prefactor  $1 \times 10^{12}$  for water adsorbed on an amorphous silicate, at 0.4 ML of water coverage. And, last but not the least, the heat transferred during the TPD run can induce water surface diffusion to new sites coming from the restructured ice, biasing the BE distribution toward higher values (Minissale et al. 2022). In summary, the soft nature of the icy grains introduces criticalities in the TPD experiments and does not allow for a straightforward comparison between computed and experimentally derived BE values. A better description and the link between the two approaches could be achieved via a micro-kinetic model that takes into account the BE distribution and the diffusion across the sites (He & Vidali 2014).

Finally, the observed difference between measured and computed BEs raises the question of which of the two techniques is more reliable in providing the BE in astrochemical models. Also under this aspect, the use of TPD derived BEs is not straightforward. Indeed, the timescale of desorption is dramatically different in TPD experiments (hours or days)



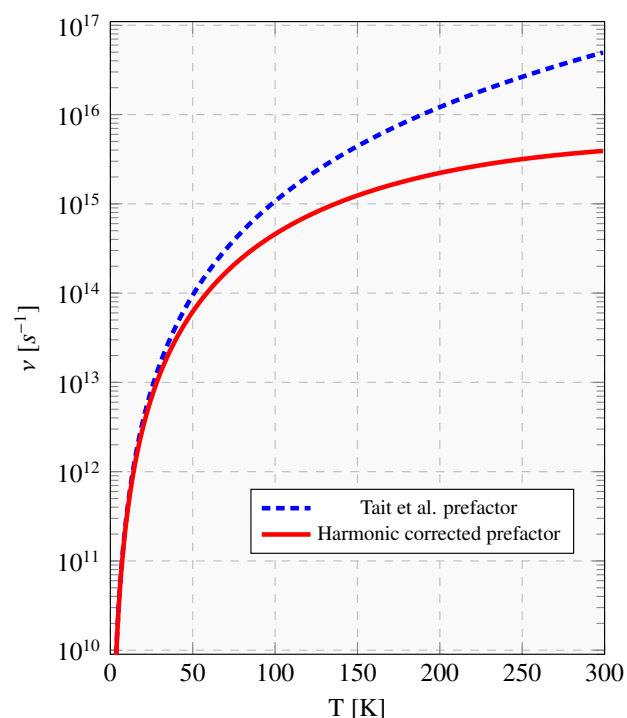
**Figure 5.** Correlation plots between  $\Delta ZPE$ s calculated for the water ( $\Delta ZPE(H_2O)$ : x-axis), and mono-deuterated ( $\Delta ZPE(HDO)$ : upper panel) and bi-deuterated ( $\Delta ZPE(D_2O)$ : bottom panel) water. For the mono deuterated case, the  $\Delta ZPE$  is the average between the two equivalent deuterated positions. All the quantities are in units of kJ/mol.

and in the ISM (thousands or millions years). In the latter the desorption is hugely slower than in the former, likely allowing just a minor adjustment of the ice surface. In this respect, therefore, the theoretically computed BEs have a better appeal to be chosen as data input in the astrochemical models.

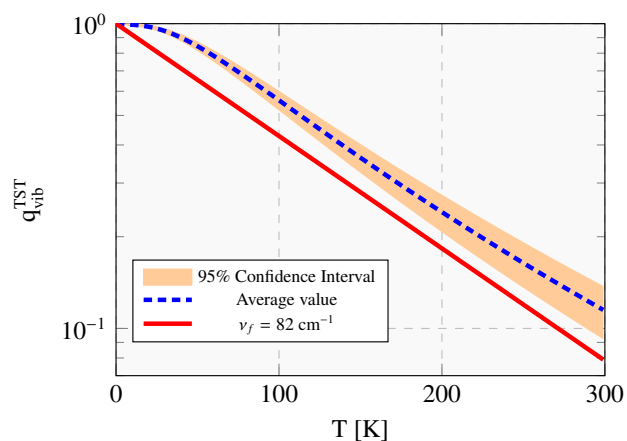
### 5.2. Theoretical data

The comparison of the results obtained in this work and the other computational studies is more straightforward than with experiments. In theoretical studies, the differences may arise by the different adopted ice models, the level of theory and (if present) the ZPE correction.

Wakelam et al. (2017) used a single water molecule for the ice model. Their computed BE are systematically smaller than those measured in experiments and the larger the BE the larger the difference (see their Fig. 1). These authors decided to apply a proportionality factor to bring the two values closer, which, as explained above, may not be justified and

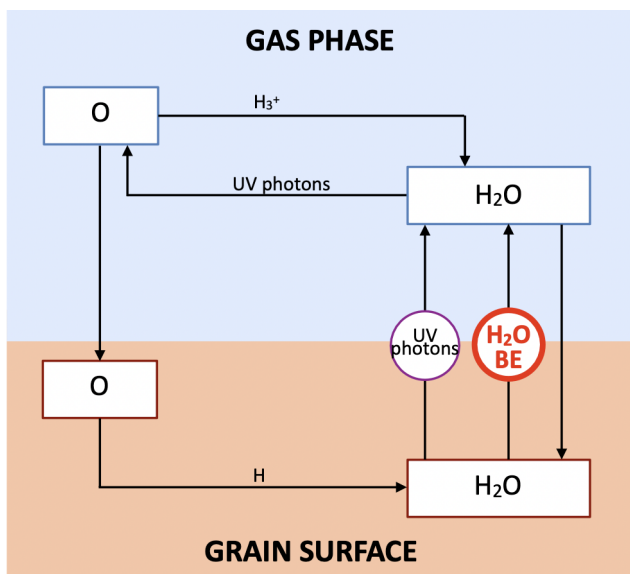


**Figure 6.** Temperature dependence of the desorption prefactor computed by using the Tait et al. (2005) formula (Eq. 4: dashed blue line) and with the inclusion of the correction of the harmonic vibrational partition functions (CSDP, Eq. 5: solid red line).



**Figure 7.** Correction of the desorption prefactor (Eq. 5) as a function of temperature. The dashed blue line shows the  $q_{\text{vib}}^{\text{TST}}$  average value of the water adsorption samples and its confidence interval pictured in orange. In solid red line is reported the fitting function of  $q_{\text{vib}}^{\text{TST}}$  average value obtained with Eq. 5.

wise. Their water BE, based only on the theoretical computations, is 38 kJ/mol, indicating that the use of a single molecule for the ice model leads to sample a binding site whose BE is larger than the mean value of our Gaussian BE distribution.



**Figure 8.** Simplified scheme of the gaseous water chemistry. Gaseous water can be formed by a chain of reactions starting from  $O + H_3^+$ , and by the photo- or thermal- desorption of the water frozen on the grain surface. It is destroyed by the photo-dissociation or by the freezing onto the grain surface. When landing from the gas onto the grain surface, atomic oxygen is assumed to be instantaneously hydrogenated into frozen water.

Das et al. (2018) increased the number of water molecules (from one to six) simulating the ice but only reported the  $H_2O$  BE without any ZPE and BSSE correction. The latter is of fundamental importance for the accuracy of Post-Hartree–Fock theories, and even more by the relatively small adopted basis set. For this reason, and against the arguments provided by Das et al. (2018), we think that a comparison with their computations would not be useful.

Ferrero et al. (2020) found water BEs ranging from 30 and 50 kJ/mol on a periodic amorphous ice model, namely shifted towards the larger BEs with respect to our Fig. 2 Gaussian distribution. The difference with the present results can be tracked down to the different icy surface model, the level of theory (they used a low accuracy HF-3c level for geometry optimization) and, likely, the limited adsorption site sampling procedure.

In the article by DufLOT et al. (2021), a procedure similar to the present one (ONIOM(CBS/DLPNO-CCSD(T):PM6)//ONIOM( $\omega$ B97X-D/6-31+G\*\*):PM6)) was adopted to compute a ZPE-corrected BE. BE values of  $40.1 \pm \sim 10$  kJ/mol and  $40.7 \pm \sim 10$  kJ/mol have been computed on crystalline cluster and amorphous ice slab, respectively. These data are in agreement with the present values with a difference of about 5 kJ/mol, despite the fact that a very different methodology was adopted to build up the underneath ice and the sampling (eight site for each case) of all the BE sites.

Regarding the estimate of the desorption prefactor, it is almost absent in the above mentioned theoretical studies, often assumed to be  $10^{12} - 10^{13}$  (the experimental assumption or the Hasegawa et al. (1992) formula). Recently, Molpeceres et al. (2020, 2022) proposed a new method in which they run molecular dynamics (MD) at GFN2 level for the adsorption of atoms,  $H_2$  and acetaldehyde on icy clusters. The prefactor, named by those authors, attempt frequency of desorption, was computed with an approach never proposed in the related literature on the surface science phenomena. Basically, the position of the adsorbate center of masses radial (COMr) with respect to the grain cluster center is monitored during a short time frame along the MD trajectories. The maximum of the COMr attempts are then worked out from the trajectories and the frequency of desorption is computed as the inverse of the average attempt periods of the simulations. However, if considering the vertical movements of the center of mass can be a smart solution for evaluating the attempt frequency of an adsorbed atom, this could not be the case for complex molecules, which have inner motions and a variety of different bonding interactions with the surface as a function of the change in orientation. Furthermore, a component in the COMr can also be simply due to different rotation of the whole molecule, lying either flat (small COMr) or almost perpendicular to the surface (high COMr), without any attempt to desorb. We notice that the Molpeceres et al. (2020, 2022)’s approach has never been proposed by the surface science community, even if many methods have been proposed in the last years in order to estimate the prefactor (Fichtorn & Miron 2002; Sprowl et al. 2016; Rybicki & Sauer 2022). Furthermore, that procedure implies a rather long MD run which can only be carried out with classical force fields or GFNx level of theory, reducing the final accuracy.

## 6. THE WATER SNOWLINE OF PROTOPLANETARY DISKS

In order to estimate the impact of the new computed  $H_2O$  BE distribution, we used a simple model to predict the snowline position, namely the gaseous and solid water abundances across a generic protoplanetary disk. The goal of the modeling is to show how the snowline of the disk changes using the newly computed BE and  $\nu$  distributions with respect to the single values for the  $H_2O$  (5600 K) and  $\nu$  ( $2 \times 10^{12} \text{ s}^{-1}$ ) adopted by the vast majority of the astrochemical models (e.g. Wakelam et al. 2017; Öberg & Wordsworth 2019). To this end, we adopted a simplified chemistry for the water chemistry, described in § 6.1, and ran it on a generic disk model, previously used in a similar study (Dominik et al. 2005), as described in § 6.2. The results of the modeling are then shown and discussed in § 6.3.

We emphasize that our goal is not to compute in detail the gaseous abundance, but rather the position of the snowline.

### 6.1. Simplified water chemistry

We adopted the simplified scheme of the steady state water chemistry shown in Fig. 8, which accounts for the formation and destruction of gaseous and solid water in the interstellar cold and warm ( $\leq 200$  K) molecular gas. Briefly, the abundance of gaseous water is regulated by the photo-desorption of the frozen water, its photo-dissociation into atomic oxygen, which in part reforms gaseous water<sup>2</sup>, the freezing of gaseous water onto the grain surfaces and its thermal desorption, following the scheme by Dominik et al. (2005). Since the thermal desorption takes the lion's share in this game, the water BE is a crucial parameter in defining the quantity of gaseous and solid water, namely position of the snowline.

### 6.2. Model description

We developed a simplified model for computing the gaseous and solid water abundance across a generic protoplanetary disk, updating and modifying the Dominik et al. (2005) model, *i.e.* adopting the scheme of Fig. 8. Briefly, the model computes the gaseous and frozen abundance of water assuming that: (i) when atomic oxygen freezes onto the grain surfaces it instantaneously forms frozen water; (ii) at each disk position, the  $\text{H}_3^+$  density (from which some gaseous water forms; see above) is equal to the electron density; and (iii) a fraction of the ice is sublimated into the gas-phase, according to the BE distribution computed in this work.

In practice, at steady state, the water gaseous abundance is computed analytically solving the following three equations:

$$\begin{aligned} n_{\text{O}} k_{\text{acc}}^{\text{O}} &= k_{\text{phd}} n_{\text{H}_2\text{O}} - k_{\text{form}}^{\text{H}_2\text{O}} n_{\text{O}} n_{\text{H}_3^+} \\ n_{\text{H}_2\text{O}} (k_{\text{acc}}^{\text{H}_2\text{O}} + k_{\text{phd}}) &= k_{\text{form}}^{\text{H}_2\text{O}} n_{\text{O}} n_{\text{H}_3^+} + k_{\text{phdes}} n_{\text{gr}} k_{\text{thdes}} n_{\text{ice}} \\ n_{\text{H}_2\text{O}} + n_{\text{O}} + n_{\text{ice}} &= A_{\text{ox}} f_{\text{ice}} n \end{aligned}$$

where  $n_{\text{O}}$ ,  $n_{\text{H}_2\text{O}}$ ,  $n_{\text{H}_3^+}$ ,  $n_{\text{gr}}$ ,  $n_{\text{ice}}$  and  $n$  are the number density of gaseous atomic oxygen, water and  $\text{H}_3^+$ , of the dust grains, frozen water (ice) and the total H-nuclei density, respectively;  $k_{\text{acc}}^{\text{O}}$  and  $k_{\text{acc}}^{\text{H}_2\text{O}}$  are the accretion rates onto the grain surfaces of O and  $\text{H}_2\text{O}$ , respectively;  $k_{\text{phd}}$ ,  $k_{\text{phdes}}$  and  $k_{\text{thdes}}$  are the gaseous water photo-dissociation rate and the frozen water photo-desorption and thermal desorption rates;  $k_{\text{form}}^{\text{H}_2\text{O}}$  is the formation of gaseous water from atomic oxygen (see footnote 2);  $A_{\text{ox}}$  is the total volatile oxygen elemental abundance (*i.e.* of the oxygen not trapped in the refractory dust grains);  $f_{\text{ice}}$  is the fraction of the sublimated ice. The latter is computed, at each point, adding up the fraction of ice corresponding to the BE bin where  $k_{\text{thdes}} \geq k_{\text{acc}}^{\text{H}_2\text{O}}$ . Note that  $k_{\text{thdes}}$  is a function of BE as reported in Eq. 2.

<sup>2</sup> At steady state, gas-phase atomic oxygen reacts with  $\text{H}_3^+$  forming  $\text{OH}^+$ ; the formation of  $\text{OH}^+$  is followed by the reactions  $\text{OH}^+ + \text{H}_2 \rightarrow \text{H}_2\text{O}^+$  and  $\text{H}_2\text{O}^+ + \text{H}_2 \rightarrow \text{H}_3\text{O}^+$ , and finally the recombination  $\text{H}_3\text{O}^+ + \text{e}^- \rightarrow \text{H}_2\text{O}$ . In practice, about 1/3 of the gaseous O forms gaseous  $\text{H}_2\text{O}$  (Dominik et al. 2005; Hollenbach et al. 2009).

**Table 3.** Values of the fraction of the ice with a given BE as a function of BE as well as the computed prefactor  $\nu$ . Units for BE are in kJ/mol and K, for the values in parenthesis,  $\text{s}^{-1}$  for  $\nu$ .

BE	$\nu$	Fraction of the ice
14.2 (1705)	$9.64 \times 10^{12}$	0.04
19.5 (2336)	$2.53 \times 10^{13}$	0.08
24.7 (2968)	$5.18 \times 10^{13}$	0.16
30.0 (3599)	$9.10 \times 10^{13}$	0.28
35.2 (4230)	$1.44 \times 10^{14}$	0.30
40.5 (4862)	$2.12 \times 10^{14}$	0.27
45.8 (5493)	$2.95 \times 10^{14}$	0.21
51.0 (6125)	$3.92 \times 10^{14}$	0.06
56.3 (6756)	$5.04 \times 10^{14}$	0.03
61.6 (7387)	$6.29 \times 10^{14}$	0.01

The values of the adopted parameters are from Dominik et al. (2005), with the exception of the  $\text{H}_2\text{O}$  photo-desorption yield, for which we assumed the value measured by Öberg et al. (2009),  $2 \times 10^{-3}$ . The total volatile oxygen elemental abundance  $A_{\text{ox}}$  is assumed equal to  $1 \times 10^{-4}$  with respect to  $\text{H}_2$ . The values of the fraction of the ice with a given BE as a function of BE is reported in Tab. 3 along with the computed prefactor  $\nu$ .

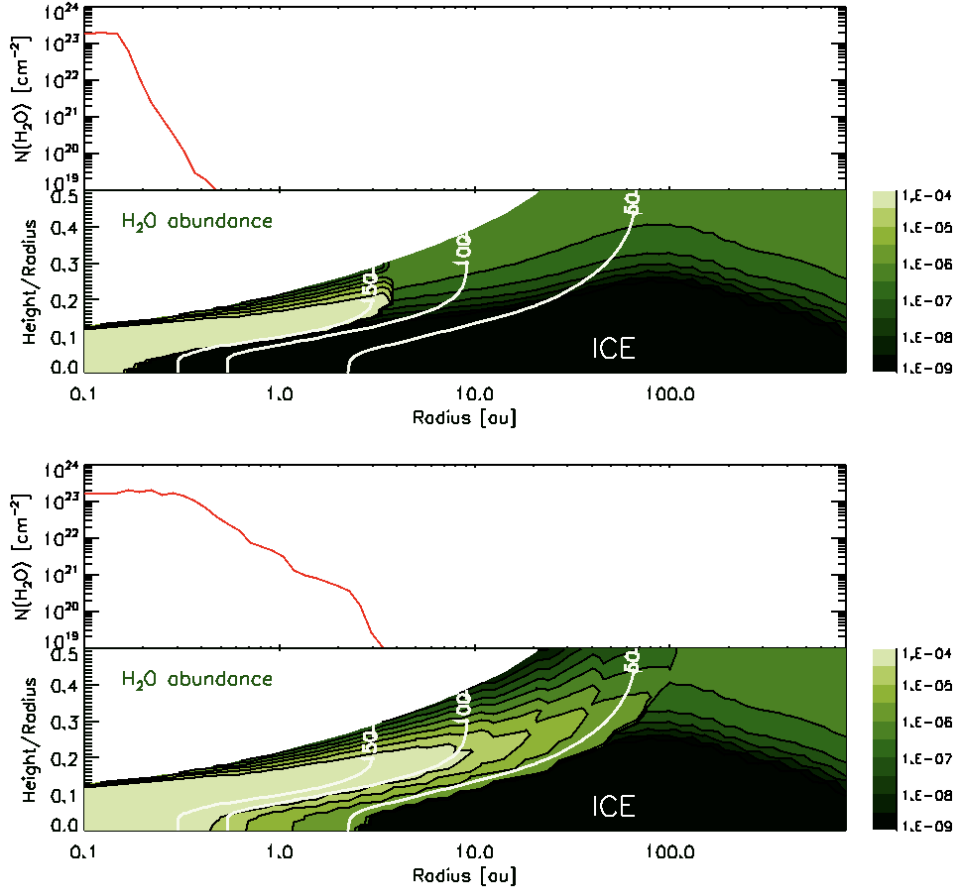
Finally, we used the physical model (temperature and density) of DM Tau, published in Dominik et al. (2005).

### 6.3. Results of the modeling

Figure 9 shows the results of the modeling using the Wakelam et al. (2017) single value for the BE (5600 K) and  $\nu$  ( $2 \times 10^{12} \text{ cm}^{-1}$ ) and the BE and  $\nu$  distributions computed in the present work, respectively.

When the Wakelam et al. (2017) values are used (upper panel in Fig. 9), the water ice is all sublimated at about 150 K (light green region in the figure) and, therefore, the water abundance reaches  $10^{-4}$  in the  $\leq 3$  au zone and, more specifically,  $\leq 0.2$  au in the midplane. In the regions where the water ice does not sublimate, the gaseous water abundance is dominated by the equilibrium set by the photo-desorption of the ices, where it is around  $10^{-7}$ – $10^{-6}$  (dark green region), and freezing onto the grain surfaces, where it drops to  $10^{-8}$  (black region).

However, when the BE and  $\nu$  distributions are used (lower panel in Fig. 9), a larger region has a gaseous water abundance  $\geq 3 \times 10^{-4}$  (light green region), where the dust temperature is  $\geq 100$  K. Most importantly, the region with a gaseous water abundance  $\geq 1 \times 10^{-8}$  is much more extended, both on the disk plan and on the layers above it, whereas the iced



**Figure 9.** Astrochemical modeling results obtained assuming the water BE (5600 K) and  $\nu$  ( $2 \times 10^{12} \text{ cm}^{-1}$ ) from [Wakelam et al. \(2017\)](#) (top figure) and the BE and  $\nu$  distributions computed in this work (bottom figure). *Upper panels:* Gaseous water column density as a function of the radius. *Lower panels:* Water abundance as a function of the radius and the height/radius. The white solid lines indicate where the dust temperature (assumed to be equal to that of the gas) is equal to 150, 100 and 50 K, respectively. In the black region, water is frozen onto the grain ice.

water region is much reduced and starts at a radius of 2 au instead of 0.2 au in this model.

As a result, when the [Wakelam et al. \(2017\)](#) values are used, the gaseous  $\text{H}_2\text{O}$  column density  $N(\text{H}_2\text{O})$  is about  $2 \times 10^{23} \text{ cm}^{-2}$  in the 0.1–0.16 au region and then it drops by more than a factor 100 at  $\geq 0.3$  au. When the BE and  $\nu$  distributions are used, the  $N(\text{H}_2\text{O})$  peak levels up to 0.3 au and then  $N(\text{H}_2\text{O})$  gently decreases and is a factor 100 lower than the peak value at  $\geq 1$  au. This behavior is due to the BE distribution in which about 80% of the BEs are less than the [Wakelam et al. \(2017\)](#) value.

It is worth reminding, at this point, that very few observations have been obtained of the cold water in protoplanetary disks ([Hogerheijde et al. 2011](#); [Podio et al. 2013](#)), all obtained with the FIR satellite Herschel. Unfortunately, these observations, which targeted the ground water line at  $538 \mu\text{m}$ , have a very large beam ( $\sim 30''$ ), hence incapable to constrain the spatial origin of the observed cold water vapour.

Interestingly, on the other side of the BE distribution, about 20% of the BEs are larger than the [Wakelam et al. \(2017\)](#) value. This implies that about 20% of the ice will need a larger dust temperature to sublimate than the one predicted by using the [Wakelam et al. \(2017\)](#) value ( $\sim 150$  K). Indeed, according to our modeling, the abundance of gaseous water is never larger than  $9 \times 10^{-5}$  when using the new BE distribution, namely  $\sim 10\%$  of water stays frozen in the ices. Even if a small fraction, this ice may remain trapped in the planetesimals eventually forming rocky planets and asteroids inside the standard snowline, and, consequently, have an impact on their water content. A fortiori, this consideration applies to the terrestrial water. However, a focused and more sophisticated model, beyond the scope of this work, is needed to fully appreciate the real impact of the newly computed BE distribution on the quantity of water that could have been trapped in the planetesimals that formed our Earth.

## 7. CONCLUSIONS



In this work, we present an extended computational study of the binding energy (BE) of water molecules on the largest so far published icy grain model, formed by 200 water molecules. The grain model was recently built by Germain et al. (2022a) using their ACO-FROST code, which carries out ab initio computations.

We used a multi-step procedure to obtain the BE over the whole set of adsorbing sites on the icy grain. First, we adsorbed the H<sub>2</sub>O molecule and let relax the system over a 5 Å ice sphere surrounding the adsorbed water molecule via a GFN2 multiple optimization. We then refined the computations via the two-layer ONIOM(QM:SQM) methodology to obtain high level accuracy quantum mechanical (QM) computations of the 5 Å ice sphere plus adsorbed H<sub>2</sub>O, while the rest of the ice grain was computed via a semi-empirical QM (QSM) method. From the initial set of 486 adsorbing sites where we computed the H<sub>2</sub>O BE, we extracted 144 unique sites, meaning each site having a different geometry.

The H<sub>2</sub>O BE of the unique sites ranges from 14.2 kJ/mol (1705 K) to 61.6 kJ/mol (7390 K). The distribution of the computed BE follows fairly well a Gaussian distribution, whose peak is at 35.4 kJ/mol (4230 K) and standard deviation is 9.7 kJ/mol (1160 K). We also computed the prefactor, that enters in the desorption rate computation, as a function of BE. It ranges from  $\sim 1 \times 10^{13} \text{ s}^{-1}$  (at lowest BE) to  $\sim 6 \times 10^{14} \text{ s}^{-1}$  (at the largest BE). These values are significantly different from those computed with the Hasegawa et al. (1992) method ( $\sim 2 \times 10^{12} \text{ s}^{-1}$ ) and adopted by several astrochemical models.

The newly computed H<sub>2</sub>O BE distribution compares fairly well with the BE derived from previous experimental TPD studies, when considering the methodological difference between the experimental and computational derivations. For example, the former measure the curve of desorption, from which BE and  $\nu$  are derived, and cannot avoid the effects of the adsorbing structural deformation during the process. On the contrary, computations are unaffected by this problem and, given the large scales involved, very likely they better describe the situation in the interstellar medium.

Finally, we developed a model to describe the water abundance across a generic protoplanetary disk and compute the position of the snowline. We found that the region of gaseous water is more extended (by about a factor 10) when considering the newly computed BE and  $\nu$  distributions than when computed with the single BE and  $\nu$  values reported in Wakelam et al. (2017) and used in many astrochemical models.

Mirroring the gaseous water distribution, the region where water is frozen is much larger (by the same factor). This behavior is due to the fact that about 80% of the BEs in the computed distribution are lower than the Wakelam et al. (2017) value (5600 K).

Furthermore, the 20% BEs with a value larger than the Wakelam et al. (2017) one leads to the incomplete sublimation of the ices at temperatures where the single BE modeling would predict the full release of frozen water into the gas-phase. Specifically, about 10% of water remains frozen at  $\sim 150 \text{ K}$  and this might contribute to the enrichment of water of rocky planets inside the classical snowline. Whether this has an impact on the origin of the terrestrial water would need a dedicated and sophisticated model.

## 8. ONLINE DATABASE

To easily handle the large data set of BE samples (atomic coordinates and BH(0) values), we developed and made publicly available a web site based on the molecule hyperactive JSmol plugin (Jmol: an open-source Java viewer for chemical structures in 3D<sup>3</sup>). This extended electronic version of the calculated results, including all the 144 sample structures at ONIOM(B97-3c:xTB-GFN2) level, are available at [https://tinaccil.github.io/Jmol\\_BE\\_H2O\\_visualization/](https://tinaccil.github.io/Jmol_BE_H2O_visualization/).

This project has received funding within the European Union's Horizon 2020 research and innovation programme from the European Research Council (ERC) for the project "The Dawn of Organic Chemistry" (DOC), grant agreement No 741002, and from the Marie Skłodowska-Curie for the project "Astro-Chemical Origins" (ACO), grant agreement No 811312. PU acknowledge the Italian Space Agency for co-funding the Life in Space Project (ASI N. 2019-3-U.O). CINES-OCCIGEN HPC is kindly acknowledged for the generous allowance of super-computing time through the A0120811498 project. LT is grateful to Jacopo Lupi, Leonardo Miele, Simon Ferrada-Chamorro and Stefano Ferrero for insightful discussions and to the L<sup>A</sup>T<sub>E</sub>X community for the insights on TikZ and PGFPlots packages. Finally, we wish to acknowledge the extremely useful discussions with Prof. F. Dulieu, Prof. Albert Rimola, Dr. P. Theule and Prof. Gretobape and his group.

*Software:* ORCA (Neese 2018), VMD (Humphrey et al. 1996), xTB (Grimme et al. 2017), ACO-FROST (Germain et al. 2022c).

## REFERENCES

- Al-Hamdani, Y. S., & Tkatchenko, A. 2019, The Journal of Chemical Physics, 150, 010901, doi: 10.1063/1.5075487

<sup>3</sup> <http://www.jmol.org/>



- Alexander, C. M. O. 2017, *Philosophical Transactions of the Royal Society of London Series A*, 375, 20150384, doi: [10.1098/rsta.2015.0384](https://doi.org/10.1098/rsta.2015.0384)
- Altun, A., Ghosh, S., Riplinger, C., Neese, F., & Bistoni, G. 2021, *The Journal of Physical Chemistry A*, 125, 9932, doi: [10.1021/acs.jpca.1c09106](https://doi.org/10.1021/acs.jpca.1c09106)
- Altun, A., Neese, F., & Bistoni, G. 2020, *Journal of Chemical Theory and Computation*, 16, 6142, doi: [10.1021/acs.jctc.0c00344](https://doi.org/10.1021/acs.jctc.0c00344)
- Bannwarth, C., Ehlert, S., & Grimme, S. 2019, *Journal of Chemical Theory and Computation*, 15, 1652, doi: [10.1021/acs.jctc.8b01176](https://doi.org/10.1021/acs.jctc.8b01176)
- Bolina, A. S., Wolff, A. J., & Brown, W. A. 2005, *The Journal of Physical Chemistry B*, 109, 16836, doi: [10.1021/jp0528111](https://doi.org/10.1021/jp0528111)
- Bolis, V., Busco, C., & Ugliengo, P. 2006, *The Journal of Physical Chemistry B*, 110, 14849, doi: [10.1021/jp061078q](https://doi.org/10.1021/jp061078q)
- Boogert, A. C., Gerakines, P. A., & Whittet, D. C. 2015, *Annual Review of Astronomy and Astrophysics*, 53, 541, doi: [10.1146/annurev-astro-082214-122348](https://doi.org/10.1146/annurev-astro-082214-122348)
- Bovolenta, G. M., Vogt-Geisse, S., Bovino, S., & Grassi, T. 2022, *The Astrophysical Journal Supplement Series*, 262, 17, doi: [10.3847/1538-4365/ac7f31](https://doi.org/10.3847/1538-4365/ac7f31)
- Boys, S. F., & Bernardi, F. 1970, *Molecular Physics*, 19, 553
- Brandenburg, J. G., Bannwarth, C., Hansen, A., & Grimme, S. 2018, *The Journal of chemical physics*, 148, 064104, doi: [10.1063/1.5012601](https://doi.org/10.1063/1.5012601)
- Brown, W. A., & Bolina, A. S. 2007, *Monthly Notices of the Royal Astronomical Society*, 374, 1006, doi: [10.1111/j.1365-2966.2006.11216.x](https://doi.org/10.1111/j.1365-2966.2006.11216.x)
- Caselli, P., Keto, E., Pagani, L., et al. 2010, *A&A*, 521, L29, doi: [10.1051/0004-6361/201015097](https://doi.org/10.1051/0004-6361/201015097)
- Ceccarelli, C., Caselli, P., Bockelée-Morvan, D., et al. 2014, in *Protostars and Planets VI*, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 859, doi: [10.2458/azu\\_uapress\\_9780816531240-ch037](https://doi.org/10.2458/azu_uapress_9780816531240-ch037)
- Ceccarelli, C., Hollenbach, D. J., & Tielens, A. G. G. M. 1996, *ApJ*, 471, 400, doi: [10.1086/177978](https://doi.org/10.1086/177978)
- Ceccarelli, C., Caux, E., Loinard, L., et al. 1999, *A&A*, 342, L21
- Cernicharo, J., González-Alfonso, E., Alcolea, J., Bachiller, R., & John, D. 1994, *ApJL*, 432, L59, doi: [10.1086/187511](https://doi.org/10.1086/187511)
- Chakarov, D. V., Oesterlund, L., & Kasemo, B. 1995, *Langmuir*, 11, 1201, doi: [10.1021/la00004a028](https://doi.org/10.1021/la00004a028)
- Collings, M. P., Frankland, V., Lasne, J., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 1826, doi: [10.1093/mnras/stv425](https://doi.org/10.1093/mnras/stv425)
- Combes, F., & Wiklind, T. 1997, *ApJL*, 486, L79, doi: [10.1086/310849](https://doi.org/10.1086/310849)
- Das, A., Sil, M., Gorai, P., Chakrabarti, S. K., & Loison, J. C. 2018, *The Astrophysical Journal Supplement Series*, 237, 9, doi: [10.3847/1538-4365/aac886](https://doi.org/10.3847/1538-4365/aac886)
- Dominik, C., Ceccarelli, C., Hollenbach, D., & Kaufman, M. 2005, *ApJL*, 635, L85, doi: [10.1086/498942](https://doi.org/10.1086/498942)
- Duflot, D., Toubin, C., & Monnerville, M. 2021, *Frontiers in Astronomy and Space Sciences*, 8, 24, doi: [10.3389/fspas.2021.645243](https://doi.org/10.3389/fspas.2021.645243)
- Dulieu, F., Amiaud, L., Congiu, E., et al. 2010, *A&A*, 512, A30, doi: [10.1051/0004-6361/200912079](https://doi.org/10.1051/0004-6361/200912079)
- Dulieu, F., Congiu, E., Noble, J., et al. 2013, *Scientific Reports*, 3, 1338, doi: [10.1038/srep01338](https://doi.org/10.1038/srep01338)
- Elitzur, M., & de Jong, T. 1978, *A&A*, 67, 323
- Feller, D. 1993, *The Journal of chemical physics*, 98, 7059, doi: [10.1063/1.464749](https://doi.org/10.1063/1.464749)
- Ferrero, S., Zamirri, L., Ceccarelli, C., et al. 2020, *The Astrophysical Journal*, 904, 11, doi: [10.3847/1538-4357/abb953](https://doi.org/10.3847/1538-4357/abb953)
- Ferrero, S., Grieco, F., Ibrahim Mohamed, A., et al. 2022, *Monthly Notices of the Royal Astronomical Society*, 516, 2586, doi: [10.1093/mnras/stac2358](https://doi.org/10.1093/mnras/stac2358)
- Fichthorn, K. A., & Miron, R. A. 2002, *Physical Review Letters*, 89, 196103, doi: [10.1103/PhysRevLett.89.196103](https://doi.org/10.1103/PhysRevLett.89.196103)
- Fischer, J., Luhman, M. L., Satyapal, S., et al. 1999, *Ap&SS*, 266, 91. <https://arxiv.org/abs/astro-ph/9911310>
- Fraser, H. J., Collings, M. P., McCoustra, M. R. S., & Williams, D. A. 2001, *Monthly Notices of the Royal Astronomical Society*, 327, 1165, doi: [10.1046/j.1365-8711.2001.04835.x](https://doi.org/10.1046/j.1365-8711.2001.04835.x)
- Freedman, D., & Diaconis, P. 1981, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 453, doi: [10.1007/BF01025868](https://doi.org/10.1007/BF01025868)
- Germain, A., Tinacci, L., Pantaleone, S., Ceccarelli, C., & Ugliengo, P. 2022a, *ACS Earth and Space Chemistry*, 6, 1286, doi: [10.1021/acsearthspacechem.2c00004](https://doi.org/10.1021/acsearthspacechem.2c00004)
- . 2022b, *Amorphous grain model of interstellar icy grains*, Zenodo, doi: [10.5281/zenodo.7385781](https://doi.org/10.5281/zenodo.7385781)
- . 2022c, *ACO-FROST: generation of icy grain models and adsorption of molecules*, Zenodo, doi: [10.5281/zenodo.7385933](https://doi.org/10.5281/zenodo.7385933)
- Gibb, E. L., Whittet, D. C. B., Boogert, A. C. A., & Tielens, A. G. G. M. 2004, *The Astrophysical Journal Supplement Series*, 151, 35, doi: [10.1086/381182](https://doi.org/10.1086/381182)
- González-Alfonso, E., Smith, H. A., Fischer, J., & Cernicharo, J. 2004, *ApJ*, 613, 247, doi: [10.1086/422868](https://doi.org/10.1086/422868)
- Grimme, S. 2006, *Journal of computational chemistry*, 27, 1787, doi: [10.1002/jcc.20495](https://doi.org/10.1002/jcc.20495)
- Grimme, S., Bannwarth, C., & Shushkov, P. 2017, *Journal of chemical theory and computation*, 13, 1989, doi: [10.1021/acs.jctc.7b00118](https://doi.org/10.1021/acs.jctc.7b00118)
- Grimme, S., Ehrlich, S., & Goerigk, L. 2011, *Journal of computational chemistry*, 32, 1456, doi: [10.1002/jcc.21759](https://doi.org/10.1002/jcc.21759)
- Guo, Y., Riplinger, C., Becker, U., et al. 2018, *The Journal of chemical physics*, 148, 011101, doi: [10.1063/1.5011798](https://doi.org/10.1063/1.5011798)
- Guo, Y., Riplinger, C., Liakos, D. G., et al. 2020, *The Journal of chemical physics*, 152, 024116, doi: [10.1063/1.5127550](https://doi.org/10.1063/1.5127550)

- Hartmann, L., Ciesla, F., Gressel, O., & Alexander, R. 2017, *Space Science Reviews*, 212, 813, doi: [10.1007/s11214-017-0406-0](https://doi.org/10.1007/s11214-017-0406-0)
- Hasegawa, T. I., Herbst, E., & Leung, C. M. 1992, *The Astrophysical Journal Supplement Series*, 82, 167, doi: [10.1086/191713](https://doi.org/10.1086/191713)
- Haynes, D. R., Tro, N. J., & George, S. M. 1992, *The Journal of Physical Chemistry*, 96, 8502, doi: [10.1021/j100200a055](https://doi.org/10.1021/j100200a055)
- He, J., & Vidali, G. 2014, *Faraday Discussions*, 168, 517, doi: [10.1039/C3FD00113J](https://doi.org/10.1039/C3FD00113J)
- Herpin, F., Chavarría, L., van der Tak, F., et al. 2012, *A&A*, 542, A76, doi: [10.1051/0004-6361/201118413](https://doi.org/10.1051/0004-6361/201118413)
- Hogerheijde, M. R., Bergin, E. A., Brinch, C., et al. 2011, *Science*, 334, 338, doi: [10.1126/science.1208931](https://doi.org/10.1126/science.1208931)
- Hollenbach, D., Kaufman, M. J., Bergin, E. A., & Melnick, G. J. 2009, *ApJ*, 690, 1497, doi: [10.1088/0004-637X/690/2/1497](https://doi.org/10.1088/0004-637X/690/2/1497)
- Humphrey, W., Dalke, A., & Schulten, K. 1996, *Journal of molecular graphics*, 14, 33, doi: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Imanishi, M., Nakanishi, K., Izumi, T., & Baba, S. 2022, *ApJ*, 926, 159, doi: [10.3847/1538-4357/ac3a68](https://doi.org/10.3847/1538-4357/ac3a68)
- Jenniskens, P., Blake, D. F., Wilson, M. A., & Pohorille, A. 1995, *The Astrophysical Journal*, 455, 389, doi: [10.1086/176585](https://doi.org/10.1086/176585)
- Kaufman, M. J., & Neufeld, D. A. 1996, *ApJ*, 456, 611, doi: [10.1086/176683](https://doi.org/10.1086/176683)
- Kendall Jr, T. 1992, *J. Chem. Phys.*, 96, 6796, doi: [10.1063/1.462569](https://doi.org/10.1063/1.462569)
- Kolasinski, K. 2002, *Surface Science* (Chichester: John Wiley & Sons, LTD)
- Kristensen, L. E., van Dishoeck, E. F., Bergin, E. A., et al. 2012, *A&A*, 542, A8, doi: [10.1051/0004-6361/201118146](https://doi.org/10.1051/0004-6361/201118146)
- Kruse, H., & Grimme, S. 2012, *The Journal of chemical physics*, 136, 04B613, doi: [10.1063/1.3700154](https://doi.org/10.1063/1.3700154)
- Lamberts, T., Samanta, P. K., Köhn, A., & Kästner, J. 2016, *Physical Chemistry Chemical Physics* (Incorporating Faraday Transactions), 18, 33021, doi: [10.1039/C6CP06457D](https://doi.org/10.1039/C6CP06457D)
- Langer, W. D., & Graedel, T. E. 1989, *ApJS*, 69, 241, doi: [10.1086/191313](https://doi.org/10.1086/191313)
- Lécuyer, C., Gillet, P., & Robert, F. 1998, *Chemical Geology*, 145, 249, doi: [10.1016/S0009-2541\(97\)00146-0](https://doi.org/10.1016/S0009-2541(97)00146-0)
- Leger, A., Klein, J., de Cheveigne, S., et al. 1979, *Astronomy and Astrophysics*, 79, 256
- Löfgren, P., Ahlström, P., Lausma, J., Kasemo, B., & Chakarov, D. 2003, *Langmuir*, 19, 265, doi: [10.1021/la020218u](https://doi.org/10.1021/la020218u)
- Martín-Doménech, R., Caro, G. M., Bueno, J., & Goesmann, F. 2014, *Astronomy & Astrophysics*, 564, A8, doi: [10.1051/0004-6361/201322824](https://doi.org/10.1051/0004-6361/201322824)
- Mayhall, N. J., Raghavachari, K., & Hratchian, H. P. 2010, *The Journal of chemical physics*, 132, 114107, doi: [10.1063/1.3315417](https://doi.org/10.1063/1.3315417)
- Melnick, G. J., & Bergin, E. A. 2005, *Advances in Space Research*, 36, 1027, doi: [10.1016/j.asr.2005.05.110](https://doi.org/10.1016/j.asr.2005.05.110)
- Minissale, M., Aikawa, Y., Bergin, E., et al. 2022, *ACS Earth and Space Chemistry*, doi: [10.1021/acsearthspacechem.1c00357](https://doi.org/10.1021/acsearthspacechem.1c00357)
- Molpeceres, G., Kästner, J., Herrero, V., Peláez, R., & Maté, B. 2022, *Astronomy & Astrophysics*, 664, A169, doi: [10.1051/0004-6361/202243489](https://doi.org/10.1051/0004-6361/202243489)
- Molpeceres, G., Rimola, A., Ceccarelli, C., et al. 2019, *MNRAS*, 482, 5389, doi: [10.1093/mnras/sty3024](https://doi.org/10.1093/mnras/sty3024)
- Molpeceres, G., Zaverkin, V., & Kästner, J. 2020, *Monthly Notices of the Royal Astronomical Society*, 499, 1373, doi: [10.1093/mnras/staa2891](https://doi.org/10.1093/mnras/staa2891)
- Morbideilli, A., Bitsch, B., Crida, A., et al. 2016, *Icarus*, 267, 368, doi: [10.1016/j.icarus.2015.11.027](https://doi.org/10.1016/j.icarus.2015.11.027)
- Morbideilli, A., Blanc, M., Alibert, Y., et al. 2019, *The Delivery of Water to Protoplanets, Planets and Satellites* (Springer), doi: [10.1007/978-94-024-1628-2](https://doi.org/10.1007/978-94-024-1628-2)
- Neese, F. 2018, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8, e1327, doi: [10.1002/wcms.1327](https://doi.org/10.1002/wcms.1327)
- Oba, Y., Miyauchi, N., Hidaka, H., et al. 2009, *Astrophysical Journal*, 701, 464, doi: [10.1088/0004-637X/701/1/464](https://doi.org/10.1088/0004-637X/701/1/464)
- Öberg, K. I., Linnartz, H., Visser, R., & van Dishoeck, E. F. 2009, *ApJ*, 693, 1209, doi: [10.1088/0004-637X/693/2/1209](https://doi.org/10.1088/0004-637X/693/2/1209)
- Öberg, K. I., & Wordsworth, R. 2019, *AJ*, 158, 194, doi: [10.3847/1538-3881/ab46a8](https://doi.org/10.3847/1538-3881/ab46a8)
- Omont, A., Yang, C., Cox, P., et al. 2013, *A&A*, 551, A115, doi: [10.1051/0004-6361/201220811](https://doi.org/10.1051/0004-6361/201220811)
- Pavošević, F., Peng, C., Pinski, P., et al. 2017, *The Journal of chemical physics*, 146, 174108, doi: [10.1063/1.4979993](https://doi.org/10.1063/1.4979993)
- Pensabene, A., van der Werf, P., Decarli, R., et al. 2022, *Astronomy & Astrophysics*, 667, A9, doi: [10.1051/0004-6361/202243406](https://doi.org/10.1051/0004-6361/202243406)
- Petersson, a., Bennett, A., Tensfeldt, T. G., et al. 1988, *The Journal of chemical physics*, 89, 2193, doi: [10.1063/1.455064](https://doi.org/10.1063/1.455064)
- Podio, L., Kamp, I., Codella, C., et al. 2013, *ApJL*, 766, L5, doi: [10.1088/2041-8205/766/1/L5](https://doi.org/10.1088/2041-8205/766/1/L5)
- Potapov, A., Jäger, C., & Henning, T. 2018, *The Astrophysical Journal*, 865, 58, doi: [10.3847/1538-4357/aad803](https://doi.org/10.3847/1538-4357/aad803)
- Riplinger, C., Sandhoefer, B., Hansen, A., & Neese, F. 2013, *The Journal of chemical physics*, 139, 134101, doi: [10.1063/1.4821834](https://doi.org/10.1063/1.4821834)
- Rosu-Finsen, A., Chikani, B., & Salzmann, C. G. 2022, *Monthly Notices of the Royal Astronomical Society*, 517, 1919, doi: [10.1093/mnras/stac2803](https://doi.org/10.1093/mnras/stac2803)
- Rybicki, M., & Sauer, J. 2022, *Journal of Chemical Theory and Computation*, 18, 5618, doi: [10.1021/acs.jctc.2c00597](https://doi.org/10.1021/acs.jctc.2c00597)
- Sack, N. J., & Baragiola, R. A. 1993, *Physical Review B*, 48, 9973, doi: [10.1103/PhysRevB.48.9973](https://doi.org/10.1103/PhysRevB.48.9973)
- Sandford, S. A., & Allamandola, L. J. 1988, *Icarus*, 76, 201, doi: [10.1016/0019-1035\(88\)90069-3](https://doi.org/10.1016/0019-1035(88)90069-3)
- Sandler, I., Chen, J., Taylor, M., Sharma, S., & Ho, J. 2021, *The Journal of Physical Chemistry A*, 125, 1553, doi: [10.1021/acs.jpca.0c11270](https://doi.org/10.1021/acs.jpca.0c11270)

- Smith, J. A., Livingston, F. E., & George, S. M. 2003, *The Journal of Physical Chemistry B*, 107, 3871, doi: [10.1021/jp022503s](https://doi.org/10.1021/jp022503s)
- Speedy, R. J., Debenedetti, P. G., Smith, R. S., Huang, C., & Kay, B. D. 1998, *The Journal of Chemical Physics*, 105, 240, doi: [10.1063/1.471869](https://doi.org/10.1063/1.471869)
- Sprowl, L. H., Campbell, C. T., & Arnadottir, L. 2016, *The Journal of Physical Chemistry C*, 120, 9719, doi: [10.1021/acs.jpcc.5b11616](https://doi.org/10.1021/acs.jpcc.5b11616)
- Stoychev, G. L., Auer, A. A., & Neese, F. 2017, *Journal of chemical theory and computation*, 13, 554, doi: [10.1021/acs.jctc.6b01041](https://doi.org/10.1021/acs.jctc.6b01041)
- Tait, S. L., Dohnálek, Z., Campbell, C. T., & Kay, B. D. 2005, *The Journal of chemical physics*, 122, 164708, doi: [10.1063/1.1883630](https://doi.org/10.1063/1.1883630)
- Tielens, A. G. G. M., & Hagen, W. 1982, *A&A*, 114, 245
- Tinacci, L., Germain, A., Pantaleone, S., et al. 2022, *ACS Earth and Space Chemistry*, 6, 1514, doi: [10.1021/acsearthspacechem.2c00040](https://doi.org/10.1021/acsearthspacechem.2c00040)
- Ulbricht, H., Zacharia, R., Cindir, N., & Hertel, T. 2006, *Carbon*, 44, 2931, doi: [10.1016/J.CARBON.2006.05.040](https://doi.org/10.1016/J.CARBON.2006.05.040)
- van Dishoeck, E. F., Herbst, E., & Neufeld, D. A. 2013, *Chemical Reviews*, 113, 9043, doi: [10.1021/cr4003177](https://doi.org/10.1021/cr4003177)
- van Dishoeck, E. F., Kristensen, L. E., Mottram, J. C., et al. 2021, *A&A*, 648, A24, doi: [10.1051/0004-6361/202039084](https://doi.org/10.1051/0004-6361/202039084)
- Wakelam, V., Loison, J. C., Mereau, R., & Ruaud, M. 2017, *Molecular Astrophysics*, 6, 22, doi: [10.1016/j.molap.2017.01.002](https://doi.org/10.1016/j.molap.2017.01.002)
- Weigend, F., Köhn, A., & Hättig, C. 2002, *The Journal of chemical physics*, 116, 3175, doi: [10.1063/1.1445115](https://doi.org/10.1063/1.1445115)
- Westall, F., & Brack, A. 2018, *SSRv*, 214, 50, doi: [10.1007/s11214-018-0476-7](https://doi.org/10.1007/s11214-018-0476-7)
- Yang, C., González-Alfonso, E., Omont, A., et al. 2020, *A&A*, 634, L3, doi: [10.1051/0004-6361/201937319](https://doi.org/10.1051/0004-6361/201937319)
- Yang, C., Omont, A., Beelen, A., et al. 2016, *A&A*, 595, A80, doi: [10.1051/0004-6361/201628160](https://doi.org/10.1051/0004-6361/201628160)
- Yocum, K. M., Smith, H. H., Todd, E. W., et al. 2019, *The Journal of Physical Chemistry A*, 123, 8702, doi: [10.1021/acs.jpca.9b04587](https://doi.org/10.1021/acs.jpca.9b04587)

## APPENDIX

## A. BINDING ENERGY FORMALISM ADOPTED

The equation adopted for the calculation of the ONIOM BEs are:

$$BE = -\Delta E = E_{ads}^{iso}(QM) + E_{grn}^{iso}(QM:SQM) - E_c(QM:SQM), \quad (A1)$$

BE can be decomposed in the pure electronic interaction ( $BE_e$ ) corrected for the Basis Set Superposition Error (BSSE) and the deformation energy ( $\delta E_{def}$ ) contributions. The  $BE_e$  is given by:

$$BE_e = E_{ads}^{iso//c}(\mathcal{G}(grn)) + E_{grn}^{iso//c}(\mathcal{G}(ads)) - E_c(QM), \quad (A2)$$

where  $E_{ads}^{iso//c}(\mathcal{G}(grn))$  and  $E_{grn}^{iso//c}(\mathcal{G}(ads))$  are the energies of the isolated adsorbate and the grain in the geometries assumed in the complex ( $iso//c$ ) in presence of the ghost orbitals of the grain  $\mathcal{G}(grn)$  and the adsorbate  $\mathcal{G}(ads)$ , respectively. As the BSSE is already taken into account in the GFN2 method, equation A2 only applies to the QM methods (*vide infra*) on the Model zone. The  $\delta E_{def}$  is defined:

$$\delta E_{def} = \underbrace{(E_{ads}^{iso//c} - E_{ads}^{iso})}_{\delta E_{def}^{ads}} + \underbrace{(E_{grn}^{iso//c} - E_{grn}^{iso})}_{\delta E_{def}^{grn}}, \quad (A3)$$

where  $\delta E_{def}^{ads}$  and  $\delta E_{def}^{grn}$  are the deformation energy of the adsorbate and the surface, respectively. Vibrational frequencies were computed on the Model zone to obtain the zero-point energies (ZPE, Eq. B7), from which the  $\Delta ZPE$  resulted as:

$$\Delta ZPE = ZPE_c - ZPE_{ads}^{iso} - ZPE_{grn}^{iso}. \quad (A4)$$

Including all the above-mentioned contributions, Equation A1 becomes:

$$BH(0) = \underbrace{BE_e}_{BE} - (\delta E_{def}^{grn} + \delta E_{def}^{ads}) - \Delta ZPE. \quad (A5)$$

## B. VIBRATIONAL PARTITION FUNCTION AND ZPE

The Vibrational Partition function  $q_{vib}$  in the harmonic oscillator assumption and for a non-linear molecule is obtained as:

$$q_{vib} = \Pi_i^{3N-6} \frac{1}{1 - \exp\left(\frac{h\nu_i}{k_B T}\right)}. \quad (B6)$$

In this case we choose the first vibrational energy level to be the zero of energy.

The ZPE is obtained as:

$$ZPE = \sum_i^{3N-6} \frac{h\nu_i}{2k_B}. \quad (B7)$$

## C. BINDING ENERGY THERMAL CORRECTION

If thermal correction is taken into account the BE at a given temperature T is defined as follows:

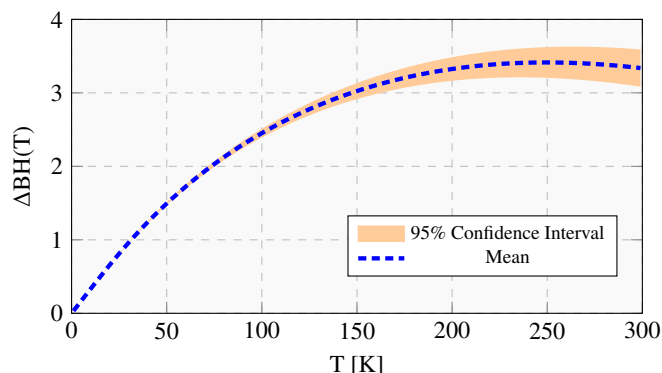
$$BH(T) = BH(0) + 4RT + H_{ads}^{vib}(T) + H_{grn}^{vib}(T) - H_c^{vib}(T), \quad (C8)$$

where the  $4RT$  term ( $R$  is the ideal gas constant and  $T$  the temperature) coming from the classical rotational ( $3/2RT$ ) and translational ( $3/2RT$ ) contributions of the isolated adsorbate molecule (since the water is non-linear) in the rigid rotor approximation and ( $1RT$ ) from the volume work contribution to the enthalpy. The  $H^{vib}(T)$  is the vibrational thermal contribution to the enthalpy (without the ZPE, since already taken into account in Eq. 3), that in the case of rigid rotor harmonic oscillator (RRHO) is:

$$H^{vib}(T) = R \sum_i^{3N-6} \frac{h\nu_i/k_B}{\exp\left(\frac{h\nu_i}{k_B T}\right) - 1}. \quad (C9)$$

In the latter equation the  $h$  is the Plank constant,  $k_B$  the Boltzmann constant,  $\nu$  the normal mode vibrational frequency and  $N$  the number of atoms of the system.

Thermal correction to the Water BE distributions is reported in Figure C.1.



**Figure C.1.** Thermal correction to the Binding energy ( $\Delta BH(T) = BH(T) - BH(0)$ ) average value of the water adsorption samples and its confidence interval as a function of the Temperature (Eq. C9).

#### D. ON DLPNO-CCSD(T) ACCURACY

As pointed out in many articles in the last years, the DLPNO approximation to the CCSD(T) method could have significant additive errors with the size of the system (Sandler et al. 2021; Pavošević et al. 2017; Al-Hamdani & Tkatchenko 2019). The error in such method, excluding the Basis Set incompleteness error and the algorithm to solve the triplets (T), resides primarily in the two parameters that control the size of the correlation space:  $T_{CutPairs}$  and  $T_{CutPNO}$ .  $T_{CutPairs}$  defines the pairs to be treated in the coupled cluster (“Strong pairs”) or in the MP2 (“Weak pairs”) fashion.  $T_{CutPNO}$  controls the dimension of the compact virtual space in function of pair natural orbitals (PNOs) for which the coupled cluster “Strong pairs” equations are solved.

In the article by Altun et al. (2020), they provided a solution on using an easy extrapolation procedure, similar but conceptually different to the two-point CBS strategy (Pettersson et al. 1988; Feller 1993), to recover the Complete PNO Space (CPS) and correct the additive error on the electron pair truncation correlation energy. This extrapolation is applied to the correlation energies obtain with two  $T_{CutPNO}$  values. The equation to recover the correlation energy in the CPS extrapolation limit ( $E_{CPS}$ ), for a given basis set, is:

$$E_{CPS} = E^X + 1.5 \cdot (E^Y - E^X), \quad (D10)$$

where  $E^X$  and  $E^Y$  are the correlation energies with  $T_{CutPNO} = 10^X$  and  $10^Y$  respectively and  $Y = X + 1$ . To calculate the total absolute energy the  $E_{CPS}$  quantity must be added to the DLPNO-CCSD energy, which is indifferent to the PNO cutoff. Recently, the same authors also provided an extensive study, on many chemical different systems, on the impact of such correction (Altun et al. 2021) and the dependence of DLPNO-CCSD(T) error with respect to RI approximation and BSSE. From what we can extract, for our purpose, on their conclusion: (i) DLPNO-CCSD(T) error does not depend on the BSSE error, (ii) with the TightPNO setting or with CPS limit the error linked to the basis set incompleteness is negligibly small, (iii) the RI error for the absolute correlation energy for aug-cc-pVTZ & aug-cc-pVTZ/C basis set is of the order of 0.5 kJ/mol for the water dimer. What is interestingly missing in their articles is the impact of such correction and error on our target quantity: the Binding Energy. In fact, due to its intrinsic subtractive nature, this quantity may have an error compensation. In order to investigate the error, we compute two water absorption cases, within the ONIOM framework, with 15 and 6 water molecules in the model zone each. In this benchmark we study the accuracy and dependence between each of the following cases:

- $T_{CutPNO}$  at  $10^{-6}$  and  $10^{-7}$  with the related CPS(6/7) extrapolation;
- the accurate iterative ( $T_1$ ) (Guo et al. 2018, 2020) algorithm and the approximate ( $T_0$ ) correction (Riplinger et al. 2013);
- the aug-cc-pVTZ/C and “autoaux-max” basis set (which is obtained matching the maximum angular moment of the parent basis set) (Stoychev et al. 2017).

The aug-cc-pVTZ primary basis set are always kept, as well as the TightPNO setting. Many works reported that the approximated algorithm to compute the triplets excitation of the coupled cluster method (the default one in ORCA and here addressed as ( $T_0$ ), meanwhile in the rest of the article is generally called (T)) can affect the accuracy of the calculations with respect to the non-approximated one (i.e. ( $T_1$ )) (Guo et al. 2018, 2020). For this reason we test also this issues. Since Altun et al. (2020) found the auxiliary “autoaux-max” basis set (Stoychev et al. 2017) coupled with the primary aug-cc-pVTZ ensures a RI error free for the water dimer.

**Table 4.** Accuracy on DLPNO-CCSD(T) calculation obtain from structures and related ZPE corrections at ONIOM(B97-3c:xTB-GFN2) level. The CCSD(T)/aug-cc-pVTZ was impossible to be computed for the 15 water system for the choose basis set and our computational resources.

System	15 water				7 water							
	aug & aug/C				aug & aug/C				aug & autoaux-max			
DLPNO-CCSD	(T <sub>0</sub> )	(T <sub>0</sub> )	(T <sub>1</sub> )	(T <sub>1</sub> )	(T <sub>0</sub> )	(T <sub>0</sub> )	(T <sub>1</sub> )	(T <sub>1</sub> )	(T <sub>0</sub> )	(T <sub>0</sub> )	(T <sub>1</sub> )	(T <sub>1</sub> )
$\log(T_{CutPNO})$	-7	-6	-7	-6	-7	-6	-7	-6	-7	-6	-7	-6
BSSE	7.36	7.41	7.38	7.42	4.61	4.59	4.62	4.60	4.64	4.65	4.66	4.66
BH(0)	53.63	52.53	53.92	52.79	32.18	31.49	32.28	31.57	32.13	31.44	32.23	31.53
BH(0) CPS	54.15		54.46		32.54		32.65		32.47		32.58	
BH(0) B97-3c	67.81				36.37							
BH(0) CCSD(T)/aug	-				32.90 (BSSE 4.65)							

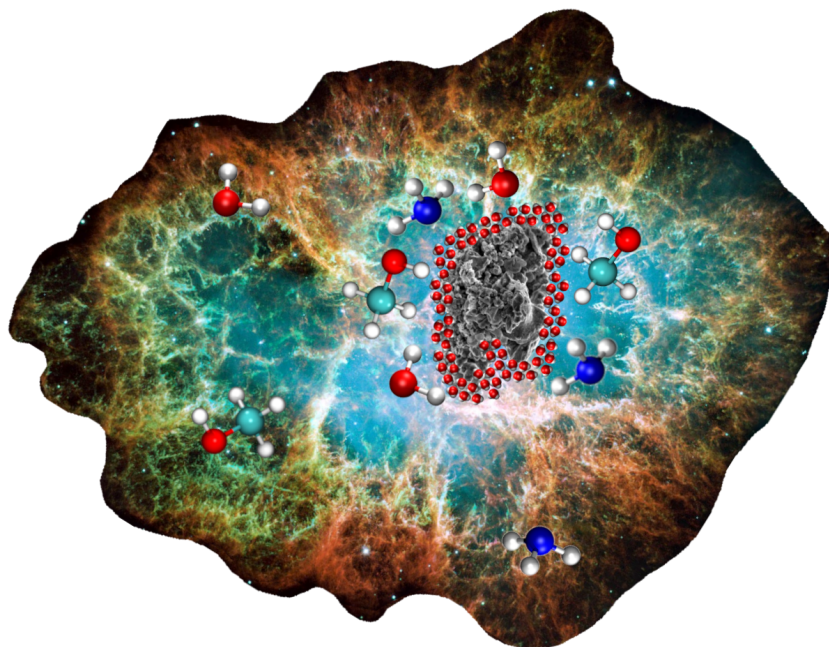
In table 4 is reported the benchmark. The BSSE has no relationship with the DLPNO-CCSD(T) correlation error, the treatment of the Triplets and the RI approximation, and it remain quasi constant below  $\sim 0.05$  kJ/mol in all the configurations. The RI error in the 7 water system is negligible and of the order of  $\sim 0.05$  kJ/mol.

The BE seems to have a compensation error, so we decided to use the (T<sub>0</sub>), aug-cc-pVTZ basis set and not CPS correction to achieve the maximum accuracy/cost ratio for our objectives. The CPS strategy should be useful in bigger systems than our own. In fact, in cases of larger systems, in which drastic decision must be taken to save computational time, one can use BSSE evaluated with (T<sub>0</sub>), CPS(7/6) or CPS(6/5) schemes and different combination of primary and secondary basis set.



# The pre-exponential factor importance on desorption rate for species on water surface: a computational study

The hereafter article will be submitted to The Journal Chemical Physics in the next month.



# The Importance of the Pre-exponential Factor on the Desorption Rate for Species frozen on Water Surfaces: a Computational Study

L. Tinacci,<sup>1,2</sup> S. Pantaleone,<sup>1</sup> C. Ceccarelli,<sup>2</sup> and P. Ugliengo<sup>1</sup>

<sup>1</sup>*Università degli Studi di Torino, Dipartimento di Chimica, via P. Giuria 7, 10125 Torino, Italy*

<sup>2</sup>*Université Grenoble Alpes, Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), 38000 Grenoble, France*

(\*Electronic mail: piero.ugliengo@unito.it)

(Dated: 8 January 2023)

Interstellar grains are essential for the development of chemical complexity in the Interstellar Medium (ISM). In cold molecular clouds, the grains are enveloped by icy mantles where some molecules can be formed and others freeze-out from the gas-phase. Then, in warm enough environments, the iced species are desorbed into the gas-phase, where they can be detected by radio telescopes. In astrochemical models, two quantities govern whether a species is frozen or gaseous: the binding energy of the species and the so-called pre-exponential factor. The former is usually obtained via laboratory experiments and theoretical calculations. Until recently, the latter has been somewhat overlooked by modelers and constant values based on very rough approximations have been often used. In this article, we present new theoretical computations on water, ammonia and methanol adsorbed on water surfaces, both amorphous and crystalline. We show that the pre-exponential factor varies with the desorption temperature. Besides, we discuss the dependence of the simulated TPD spectra of adsorbed water, ammonia and methanol on the approach used to estimate the pre-exponential factor. The results demonstrate that the desorption temperature may change up to 60 K. Finally, we propose a cost-effective strategy to include all the thermal contributions in the partition functions without performing a complete frequency calculation.

## I. INTRODUCTION

The desorption rate is one of the fundamental observable needed to describe the microscopic mechanisms that account for the desorption of an adsorbed species from the surface. In other words, the desorption rate is the main actor that rules whether a species remain stuck on a surface or goes into the gas phase. The desorption rate ( $k_{\text{des}}$ ) is normally mathematically described as an Arrhenius process:

$$k_{\text{des}} = \nu \exp(-E_a/(R T)),$$

where the pre-exponential  $\nu$  is related to the entropy of the desorption mechanism and the activation energy  $E_a$  of the process can be assumed to be the adsorption enthalpy.  $E_a$  is assumed to be approximately equal to what is often called the binding energy (BE).

The determination of the pre-exponential factor (elsewhere also called prefactor) in the chemical desorption mechanism is a long standing issue<sup>1,2</sup> because of the high level of accuracy required in both experimental and computational studies. In fact, from the computational point of view, despite several methods provide a very accurate estimation of the energetic term,<sup>3,4</sup> the prefactor still remains an elusive quantity<sup>1,5</sup>.

The main problem connected to the estimation of the prefactor is the difficulty in describing the vibrational modes, especially those deriving from the free translations and rotations of the adsorbate that become intermolecular vibrational modes once the gas phase molecule is stuck on the surface. In the last decades, different approaches have been proposed to retrieve this fundamental piece of information: the immobile particle approximation<sup>1,6</sup> (where the vibrational frequencies are not considered at all), the harmonic oscillator (*i.e.* the calculation of the Hessian truncating the energy at the second

derivative), and the free and hindered translators<sup>5</sup>. Many approaches were proposed to go beyond the above-mentioned approximations<sup>7-9</sup>, the one with the most general applicability taking into account the full anharmonic treatment of all vibrational modes, solving the 1D nuclei Schrödinger equation for each normal mode or combination of them<sup>10-12</sup>. This treatment is important, especially for high amplitude motions and low frequencies, and for high-frequency stretchings involving light atoms, in particular H. A way to overcome the limitations imposed by the static calculations is to take advantage of Molecular Dynamics (MD) simulations<sup>13</sup> which, through the Fourier transform of the velocity autocorrelation function, provide the anharmonic frequency spectrum of the system<sup>14</sup>.

Usually, all these techniques are applied to the interaction of molecules with metallic, metal oxide surfaces, Zeolite and Metallic-Organic-Frameworks (MOFs). The main difference with respect to a molecular crystal surface is that on the above-mentioned surfaces the assumption of a rigid surface works well, the cohesive energy of the surface itself being orders of magnitude higher than the interactions that govern the adsorption process. This means that during the adsorption process the surface does not undergo any restructuring, at variance of what happens on molecular crystal surfaces. Moreover the above-mentioned structures are usually highly symmetric and, consequently, it is also easier to explore the PES around the adsorption site by means of MD simulations.

In this work, we will focus on the interaction of ammonia, water, and methanol with water surfaces which aim at simulating the icy mantles of interstellar grains. The interactions between the adsorbed species and the surface are mainly driven by hydrogen bonds, the same interactions that compose the surface itself. For this reason, when the desorption process

is studied by means of Temperature-Programmed Desorption (TPD) experiments on this kind of systems, the desorption rate is often overestimated<sup>15</sup>. The reason is that it is very difficult to selectively desorb the adsorbed species only, as well as to prevent the restructuring of the surface, which affects the results.

The desorption rates of interstellar molecules are among the major actors in the predictions of astrochemical microkinetic models. The comparison of those predictions with astronomical observations help to understand several mechanisms occurring in the molecular interstellar medium, including the evolution of the chemical species in the various phases of the formation of a planetary system like the Solar one (dark molecular clouds, prestellar cores, protoplanetary disks, etc.).

The main objective of the present work is to provide quantitative differences in using the prefactor formula commonly employed by the astrochemical community, proposed by Hasegawa & Herbst<sup>16</sup>, the immobile particle approximation proposed by Tait et al.<sup>6</sup> and its extension throughout the harmonic oscillator to obtain the vibrational modes. The challenge is to apply the different approaches to the adsorption of species at “soft” water surfaces, with respect to “hard” metallic surfaces where these theories has been widely applied, and discuss the differences. Furthermore, we will illustrate a way to fit the desorption rates in order to keep the temperature dependence for both prefactor and BE in a way that can be easily implemented in microkinetic simulation programs.

## II. FORMALISM TO DERIVE THE BINDING ENERGY

The equation adopted for the calculation of the BEs is:

$$BE = -\Delta E = E_{\text{ads}}^{\text{iso}} + E_{\text{srf}}^{\text{iso}} - E_{\text{c}} , \quad (1)$$

where  $E_{\text{c}}$  is the energy of the complex and  $E^{\text{iso}}$  the energies of the isolated systems, with “ads” standing for the adsorbate, and “srf” for the surface. The energy of the complex is also corrected for the Basis Set Superposition Error (BSSE)<sup>17</sup>, if not already taken into account by definition in the chosen method.

BE can be decomposed in the pure electronic interaction ( $BE_{\text{e}}$ ) and the deformation energy ( $\delta E_{\text{def}}$ ) contributions. The  $BE_{\text{e}}$  is given by:

$$BE_{\text{e}} = E_{\text{ads}}^{\text{iso}/c} + E_{\text{srf}}^{\text{iso}/c} - E_{\text{c}} , \quad (2)$$

where  $E_{\text{ads}}^{\text{iso}/c}$  and  $E_{\text{srf}}^{\text{iso}/c}$  are the energies of the isolated adsorbate and the grain at the geometries of the complex (iso/c).  $\delta E_{\text{def}}$  is defined as:

$$\delta E_{\text{def}} = \underbrace{(E_{\text{ads}}^{\text{iso}/c} - E_{\text{ads}}^{\text{iso}})}_{\delta E_{\text{def}}^{\text{ads}}} + \underbrace{(E_{\text{srf}}^{\text{iso}/c} - E_{\text{srf}}^{\text{iso}})}_{\delta E_{\text{def}}^{\text{srf}}} ,$$

where  $\delta E_{\text{def}}^{\text{ads}}$  and  $\delta E_{\text{def}}^{\text{srf}}$  are the deformation energies of the adsorbate and the surface, respectively.  $\delta E_{\text{def}}$  is therefore expected to be a positive quantity, even if exceptions sometimes occur (for a detailed discussion, please refer to Ref. 18).

In addition, vibrational frequencies are computed to obtain the zero-point energies (ZPE), from which the  $\Delta ZPE$  results as:

$$\Delta ZPE = ZPE_{\text{ads}}^{\text{iso}} + ZPE_{\text{srf}}^{\text{iso}} - ZPE_{\text{c}} .$$

Including all the above-mentioned contributions, Eq. 1 becomes:

$$BH(0) = \underbrace{BE_{\text{e}} - \delta E_{\text{def}}}_{BE} + \Delta ZPE , \quad (3)$$

If the thermal correction is taken into account the BE at a given temperature T is:

$$BH(T) = BH(0) + 4RT + H_{\text{ads}}^{\text{vib}}(T) + H_{\text{srf}}^{\text{vib}}(T) - H_{\text{c}}^{\text{vib}}(T) , \quad (4)$$

where the  $4RT$  term ( $R$  is the ideal gas constant and  $T$  the temperature) comes from the classical rotational ( $3/2RT$ ) and translational ( $3/2RT$ ) contributions of the isolated adsorbate molecule (non-linear in the selected cases) in the rigid rotor approximation, and  $1RT$  comes from the volume work contribution to the enthalpy. The  $H^{\text{vib}}(T)$  is the thermal contribution to the enthalpy without the ZPE already taken into account in Eq. 3, that in the case of rigid rotor harmonic oscillator (RRHO) is:

$$H_{\text{RRHO}}^{\text{vib}}(T) = R \sum_i^{3N-6} \frac{h \nu_i / k_B}{\exp\left(\frac{h \nu_i}{k_B T}\right) - 1} , \quad (5)$$

where  $h$  is the Planck constant,  $k_B$  the Boltzmann constant,  $\nu$  the normal mode vibrational frequency and  $N$  the number of atoms of the system.

The vibrational enthalpy ( $H^{\text{vib}}(T)$ ) can also be corrected as proposed by the Martin Head-Gordon group<sup>19</sup>, where the rigid rotor-harmonic oscillator approximation can be improved by treating low-lying vibrational modes as free translational and rotational modes via a quasi-RRHO (q-RRHO) model:

$$H_{\text{q-RRHO}}^{\text{vib}}(T) = R \sum_i^{3N-6} \omega(\nu_i) \frac{h \nu_i / k_B}{\exp\left(\frac{h \nu_i}{k_B T}\right) - 1} + T \frac{1 - \omega(\nu_i)}{2} , \quad (6)$$

where  $\omega(\nu_i)$  is defined as follows:

$$\omega(\nu_i) = \frac{1}{1 + (\nu_0 / \nu_i)^4} , \quad (7)$$

and  $\nu_0$  is the threshold frequency for considering the vibrational modes as free translations and rotations ( $< \nu_0$ ) or harmonic vibrations ( $> \nu_0$ ). In the present work,  $\nu_0$  has been set to  $100 \text{ cm}^{-1}$ .

## III. DESORPTION RATE CONSTANT

### A. Formalism

The desorption process can be formulated with the Polanyi-Wigner equation<sup>20</sup>:

$$-\frac{dN}{dt} = N^i \underbrace{\nu \exp(-E_a / (R T))}_{k_{\text{des}}(T)} . \quad (8)$$

where  $N$  is the number of adsorbed species on the surface,  $i$  is the order of the process,  $\nu$  is the pre-exponential factor (also called prefactor),  $E_a$  the activation energy,  $R$  the ideal gas constant,  $T$  the temperature and  $t$  the time. Typically, the physisorption process of small molecules is a non-activated process, so that the activation energy is assumed to be equal to the BE ( $E_a \approx BE$ ) of the species.

Usually, in the theoretical and experimental astrochemical community, the prefactor is estimated via the following equation proposed by Hasegawa & Herbst<sup>16</sup>:

$$\nu_{\text{HH}} = \sqrt{\frac{2 E_a}{\pi^2 m_M A}}, \quad (9)$$

where  $A$  is the surface area per adsorbed molecule (normally assumed to be  $10^{-19} \text{ m}^2$ , and this value is kept constant along this work) and  $m_M$  is the mass of the adsorbed species. This equation usually leads to values between  $10^{12} - 10^{13} \text{ s}^{-1}$ , as reported in Ref. 16. As an alternative, in some cases the prefactor can be inferred by experiments (*e.g.* in Ref. 21), running different temperature ramps in the same TPD experiment. However, the majority of BEs and  $\nu$  present in literature are obtained at a given temperature peak and, consequently, they are constant, *i.e.* without any dependency on the temperature<sup>15</sup>.

Using the Transition State Theory (TST), we can model the desorption rate as an unimolecular process  $C \rightarrow M+S$  ( $C$  is the adsorbed complex,  $M$  and  $S$  are respectively the isolated adsorbed molecule and surface), were a gas-like transition state with a barrier and structure assumed to be equal to the BE (corrected for the ZPE and thermal contributions, *i.e.*  $BH(T)$ ) and not interacting (isolated) adsorbed molecule and surface, respectively. The associated rate can be written as:

$$k_{\text{des}}(T) = \nu_{\text{TST}}(T) \exp\left(-\frac{BH(T)}{R T}\right), \quad (10)$$

where the prefactor  $\nu_{\text{TST}}(T)$  is expressed as:

$$\nu_{\text{TST}}(T) = \frac{k_B T}{h} \frac{\ddagger q}{q} \kappa,$$

and where  $\ddagger q$  and  $q$  are the partition functions of the transition state (TS) and the complex, and  $\kappa$  the transmission coefficient that, for an irreversible process such as the desorption, can be assumed to be equal to 1<sup>22</sup>. The above equation can be expanded by factorizing the different terms of the internal partition function for each involved species, leading to:

$$\nu_{\text{TST}}(T) = \frac{k_B T}{h} \frac{\ddagger q_{\text{trans}}^{2D}(M) \ddagger q_{\text{rot}}(M)}{\underbrace{q_{\text{vib}}(C)}_{q_{\text{vib}}^{\text{TST}}}} \frac{q_{\text{vib}}^{\ddagger}(M) q_{\text{vib}}^{\ddagger}(S)}{q_{\text{vib}}(C)}. \quad (11)$$

Here we do not report the  $q_{\text{elec}}$  since, at room temperature, it is equal to the degeneracy of the electronic state, *i.e.*  $q_{\text{elec}} = 1$ , except for radicals or species in excited state. The  $k_B T/h$  term represents the translational degree of freedom that leads the system to the TS (gas phase), *i.e.* the translational motion of

the desorbing molecule normal to the surface. In other words, it represents the third dimension of the 2D partition function ( $\ddagger q_{\text{trans}}^{2D}(M)$ ).  $k_B T/h$  is the key actor in the estimation of the prefactor and, at a temperature around 100K, it has a value of  $\approx 2 \times 10^{12} \text{ s}^{-1}$ . The  $\ddagger q_{\text{trans}}^{2D}(M)$  is approximated to the translational partition function of a noble gas adsorbed on a surface:

$$\ddagger q_{\text{trans}}^{2D}(M) = \left(\frac{2 \pi m_M k_B T}{h^2}\right) A.$$

On the other hand, the rotational partition function can be expressed within the rigid rotor approximation:

$$\ddagger q_{\text{rot}}(M) = \frac{\sqrt{\pi}}{\sigma h^3} (8 \pi^2 k_B T)^{\frac{3}{2}} \sqrt{I_x I_y I_z},$$

where  $I_x$ ,  $I_y$ ,  $I_z$  are the principal moment of inertia in a.m.u.<sup>2</sup> and  $\sigma$  is the symmetry factor that classically identifies the indistinguishable rotational configurations of a specific molecule. The latter equation can be expressed in terms of rotational constants ( $A$ ,  $B$  and  $C$  in  $\text{cm}^{-1}$ ), formalism preferred by the astronomical community, leading to  $\frac{\sqrt{\pi}}{\sigma} \left(\frac{\pi k_B T}{h c}\right)^{\frac{3}{2}} \sqrt{\frac{1}{A B C}}$ . Finally, the  $q_{\text{vib}}$  for the harmonic assumption is obtained as:

$$q_{\text{vib}} = \prod_i^{3N-6} \frac{1}{1 - \exp\left(-\frac{h \nu_i}{k_B T}\right)}. \quad (12)$$

Eq. 11 is usually approximated in the limit of a completely immobile particle in the adsorbed state<sup>1,6,15</sup> assuming  $q_{\text{vib}}^{\text{TST}}$  equal to one, thus leading to:

$$\nu_{\text{TST}}^{\text{Tait}}(T) = \frac{k_B T}{h} \left(\frac{2 \pi m_M k_B T}{h^2}\right) A \frac{\sqrt{\pi}}{\sigma h^3} (8 \pi^2 k_B T)^{\frac{3}{2}} \sqrt{I_x I_y I_z}. \quad (13)$$

This approximation relies on the assumption that, below the typical desorption temperature of H-bonded species ( $< 200\text{K}$ ), no normal modes are excited and, consequently, the vibrational partition function leads to 1. This approximation was applied to rigid surfaces, where the cohesive energies of the surface are orders of magnitude greater than the adsorption of the species over it. However, for “soft” surfaces (*i.e.* where the molecules are connected by intermolecular weak forces, like the H-bond pattern in the ice), it is possible that low-frequency phonons play a not negligible role on the  $q_{\text{vib}}^{\text{TST}}$  term. In the selected samples, the forces keeping the adsorbate bound to the surface are of the same nature of those connecting the molecule of the surface itself and, in this specific scenario, the above-mentioned approximation is not necessarily valid. The major contributor that leads to different results with respect to the approximated form are the 6 hindered rotations and translations of the adsorbed molecule (we define the “6HRT” as an acronym to identify these modes) converted to internal vibrations upon adsorption. Equation 13 does not depend on the final structure and, therefore, it is not associated to a specific site and BE, giving a constant number for each species. Indeed, as reported by Ref. 1, the prefactor calculated by Eq. 13 must be considered as an upper limit. The

vibrational partition functions ratio,  $q_{\text{vib}}^{\text{TST}}$ , can be expanded as:

$$\frac{q_{\text{vib}}^{\ddagger}(\text{M})q_{\text{vib}}^{\ddagger}(\text{S})}{q_{\text{vib}}(\text{C})} \approx \frac{q_{\text{vib}}^{\ddagger}(\text{M})}{\underbrace{q_{\text{vib}}^{\text{intra}}(\text{M//C})}_{q_{\text{vib}}^{\text{TST}}(\text{M})}} \frac{q_{\text{vib}}^{\ddagger}(\text{S})}{\underbrace{q_{\text{vib}}(\text{S//C})}_{q_{\text{vib}}^{\text{TST}}(\text{S})}} \frac{1}{q_{\text{vib}}^{\text{6HRT}}}, \quad (14)$$

where  $q_{\text{vib}}^{\text{6HRT}}$  is the vibrational partition function only considering the six hindered rotations and translations of the adsorbed species on the surface,  $q_{\text{vib}}^{\text{intra}}(\text{M//C})$  is the partition function for the 6N-6 intramolecular vibrations of the adsorbed species on the surface and  $q_{\text{vib}}(\text{S//C})$  the partition function of the surface normal modes excluding the contributes from the adsorbed species.

Using the correction for low-frequency vibrations of Eq. 6 and the correction for the entropy proposed by Grimme et al.<sup>23</sup>, the partition function can be written as follow:

$$q_{\text{vib}} = \prod_i^{3N-6} \omega(v_i) q_{\text{vib}}(v_i) + (1 - \omega(v_i)) q_{\text{rot}}(v_i). \quad (15)$$

where  $\omega(v_i)$  is the dumping function defined in Eq. 7 and  $q_{\text{rot}}(v_i)$  is the  $v_i$  normal mode weighted rotational partition function, *i.e.*:

$$q_{\text{rot}}(v_i) = \frac{\sqrt{8 \pi^2 k_B T}}{h} \sqrt{\frac{\mu(v_i) I_{\text{av}}}{\mu(v_i) + I_{\text{av}}}},$$

$I_{\text{av}}$  is the average inertia tensor and  $\mu(v_i)$  is equal to  $\frac{h}{8\pi v_i}$ . As pointed out in Grimme's paper, this type of approach can easily correct the entropy without a full exploration of the PES like in the hindered rotator and translator approach<sup>5</sup>.

## B. TPD Spectrum and Temperature Peak

The TPD spectrum is computed by using Eq. 8. For this reason, the time is expressed as a function of the temperature through the experimental surface heating rate ( $\beta = dT/dt$ ) obtaining the first-order at the monolayer coverage regime (*i.e.*  $i=1$ ):

$$-\frac{dN}{dT} = \frac{N}{\beta} v(T) \exp\left(-\frac{BH(T)}{RT}\right).$$

The maximum temperature peak is found by differentiating the above equation as a function of T and setting  $d^2T/dT^2 = 0$ . Since both the prefactor  $v$  and BH are functions of T, the derivative has to be solved numerically by integrating the differential equation with, for example, the Euler method. In the approximation that  $v$  and BH are not temperature dependent, the temperature peak is obtained by solving the root of the following equation numerically:

$$\frac{BH(0)}{RT^2} - \frac{v}{\beta} \exp\left(-\frac{BH(0)}{RT}\right) = 0. \quad (16)$$

TABLE I. Binding energies of molecular cases and their decomposition. All values are in kJ/mol.

Species	Sample	BH(0)	BE <sub>e</sub>	ΔZPE	δE <sub>def</sub>
H <sub>2</sub> O	Mol 1	30.9	45.3	-10.7	3.7
	Mol 2	44.6	71.7	-12.0	15.1
	Mol 3	56.6	84.6	-15.1	12.9
NH <sub>3</sub>	Mol 1	24.3	57.8	-5.8	27.8
	Mol 2	34.1	54.1	-9.1	10.9
	Mol 3	44.5	69.1	-10.8	13.9
CH <sub>3</sub> OH	Mol 1	20.3	31.8	-10.2	1.3
	Mol 2	35.9	54.2	-8.4	9.9
	Mol 3	55.1	75.3	-9.6	10.6

*Note:* The samples are obtain at ONIOM(DLPNO-CCSD(T)/aug-cc-pVTZ:xTB-GFN2//B97-3c:xTB-GFN2) level of theory. We remind that the normal BE (Eq. 3) is the sum of BE<sub>e</sub> and δE<sub>def</sub>.

## IV. COMPUTATIONAL METHODS & MODELS

### A. Molecular calculations

The cluster model made by 200 water molecules is the same as in Refs. 18, 24, and 25, as well as large part of the methodology (such as the ONIOM approach and the constrained optimization, only high-level molecules are free to relax, see Figure 1),<sup>18</sup>. Small modifications to the above references are the computational code and the DFT method used to carry out geometry optimizations and frequency calculations. Specifically, we relied on the ORCA program (v.5.0.2)<sup>26</sup>, which implements a full integration with the xTB code<sup>27,28</sup>, used for the treatment of the low-level zone. As regards the DFT high-level method, the B97-3c<sup>29</sup> functional was used, which already incorporates a very well-balanced basis set, without the need for any BSSE correction. The ONIOM energies were refined using the DLPNO-CCSD(T)<sup>30</sup> coupled with the aug-cc-pVTZ<sup>31</sup> as the primarily basis set and the aug-cc-pVTZ/C<sup>32</sup> as the auxiliary basis set for the resolution of the identity (RI) approximation in electron repulsion integrals. "Tight PNO" and "Tight SCF" settings were used in DLPNO-CCSD(T) calculations. All the calculations were corrected for the (BSSE)<sup>17</sup>. As described in our previous cited paper<sup>18</sup>, every sample has its own reference isolated surface obtained by deleting the adsorbed molecule in the complex and then re-optimized it. Frequency calculations were calculated numerically, by differentiation of the first analytical derivatives using the center-difference formula (*i.e.* two displacements for each atom in each direction).

*Computed cases:* The desorption processes were simulated by removing the adsorbate from the previously mentioned 200-water cluster. Three adsorption sites were studied for three different molecules: water, ammonia and methanol. In Table I and in Fig. 1 the energetic and structural features of these samples are presented.

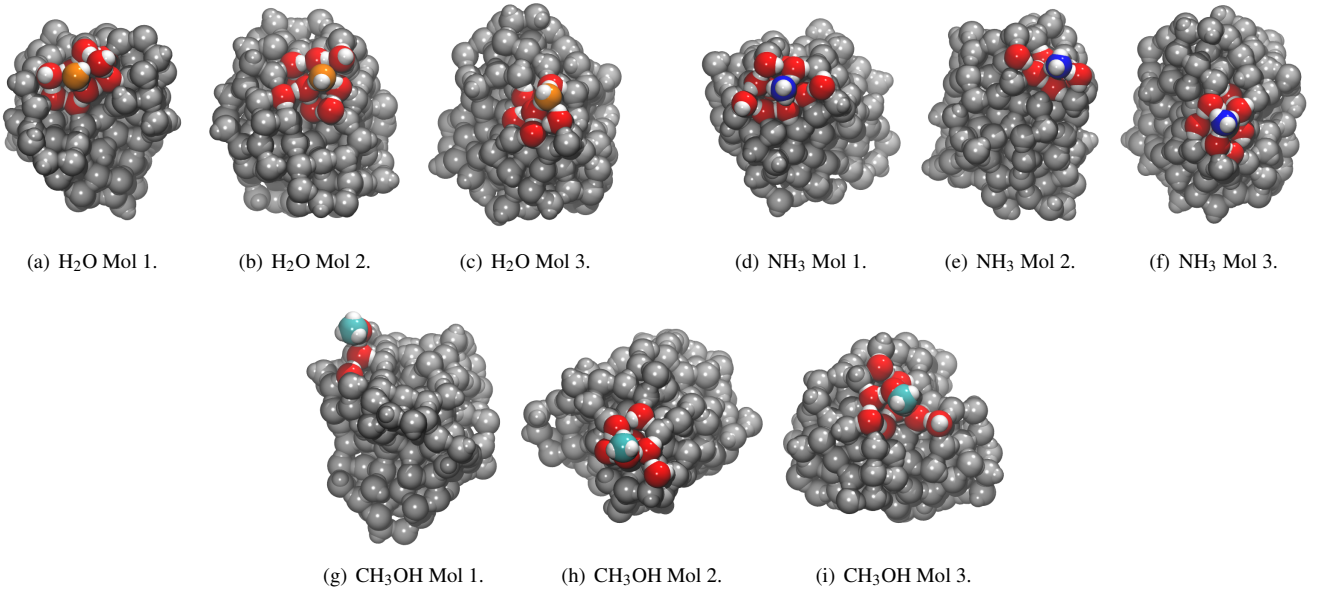


FIG. 1. Optimized geometries of the molecular samples: water (a-c), ammonia (d-f) and methanol (g-i). The atoms in the ONIOM high-level zone are shown in colors while the ONIOM low-level zone is in gray. Atom color legend: oxygen in red, hydrogen in white, nitrogen in blue and carbon in cyan. In subfigures (a), (b) and (c) the oxygen of the desorbing water molecule is highlighted in orange. Rendering of all the molecule images has been obtained via the VMD software<sup>33</sup>. The extend electronic version is available at the following link: [https://tinaccil.github.io/Jmol\\_BE\\_prefactor\\_visualization](https://tinaccil.github.io/Jmol_BE_prefactor_visualization).

## B. Periodic calculations

Periodic calculations on the experimental unit cell of P-Ice (Ice Ih) and on its surface models were carried out with the CRYSTAL17 code.<sup>34,35</sup> To be consistent with molecular calculations the B97-3c method was used.<sup>29</sup> The  $\Gamma$ -centered k-point grid was generated via the Monkhorst-Pack scheme, and the shrinking factor was set to 6, for a total number of 21 independent k-points in the irreducible First Brillouin Zone (FBZ). Tolerances of the integral calculation were set to  $10^{-7}$  ( $10^{-6}$  default) Hartree for Coulomb overlap and Coulomb penetration in the direct space. The DFT integration grid was set to 99 radial points and 1454 angular points (XXLGRID) in a Lebedev scheme. The SCF tolerance on energy convergence was increased to  $10^{-11}$  Hartree. In geometry optimizations, both atomic positions and cell vectors were relaxed, setting the tolerances for the convergence of the maximum allowed gradient and the maximum atomic displacement to  $4 \cdot 10^{-5}$  Hartree/Bohr and  $12 \cdot 10^{-5}$  Bohr, respectively. Frequency calculations at the  $\Gamma$ -point have been performed with the central-difference formula and the IR intensities computed through the Barry phase approach.<sup>36,37</sup> Moreover, in order to obtain a more reliable vibrational correction of the cohesive energy, phonon dispersion of the bulk structure was evaluated by means of the supercell ( $2 \times 2 \times 2$ ) approach (the shrinking factor was accordingly reduced to 3), thus breaking the translational symmetry and allowing for a larger sampling in the FBZ than in  $\Gamma$ -point only. The results on the cell parameters, cohesive energy and IR spectrum are available in the

Appendix.

To simulate the ice surface, the bulk structure was cut along the (100) direction, ranging the thickness from 1 up to 4 layers, in order to identify the minimum thickness that leads to converged surface properties (see Appendix).

The adsorption process was simulated by adding the adsorbate molecule to the 3-layer ice surface, while the desorption process was treated more carefully, due to possible high restructuring during the geometry optimization. More details are reported in Appendix A.

Frequency calculations at the  $\Gamma$ -point have been performed with the central-difference formula. All the atoms were set free to displace and, for comparison, we also let vibrate only the first coordination sphere of the water molecules close to the adsorbate. The aim is to compare the results obtained from a full frequency calculation versus a smaller fragment, in terms of the ZPE correction and the resulting partition functions, in order to save large part of the computational cost required by a full frequency calculation.

*Computed cases:* As for the molecular cases, the adsorbate was removed from the  $2 \times 1$  3-layers ice surface. In the case of water, two different desorption processes were simulated: the first one by removing an adsorbed water molecule on the (100) surface (hereafter called “Cry Ads”, see Fig. 2a), the second one by removing a molecule from the bare (100) bare ice surface (hereafter called “Cry Des”, see Fig. 2b, the removed water molecule is highlighted in green). In Table II and in Fig. 2 the energetic and structural features of these samples are presented.



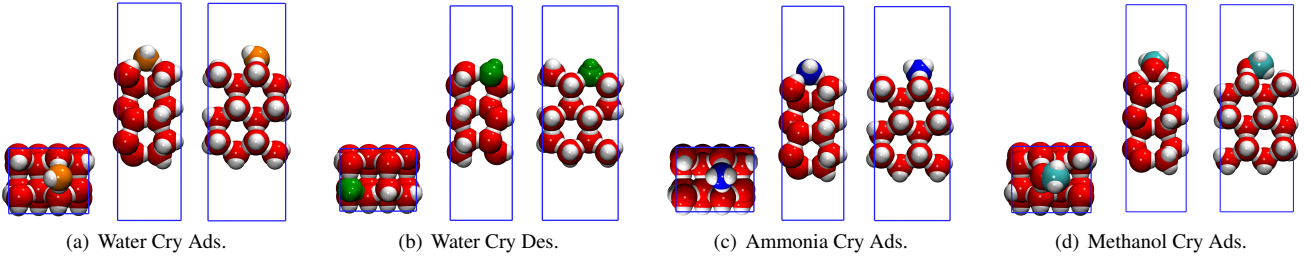


FIG. 2. Optimized geometries of the periodic samples with three different orthogonal perspectives of the periodic cell. Atom color legend: oxygen in red, hydrogen in white, nitrogen in blue and carbon in cyan. In subfigures a) and b) the two desorbing water cases “Cry Ads” (orange) and “Cry Des” (green) are reported, respectively. Rendering of all the molecule images have been obtained via the VMD software<sup>33</sup>. The extend electronic version is available at the following link: [https://tinaccil.github.io/Jmol\\_BE\\_prefactor\\_visualization](https://tinaccil.github.io/Jmol_BE_prefactor_visualization).

TABLE II. Binding energies of periodic cases and their decomposition. All values are in kJ/mol.

Species	Sample	BH(0)	BE <sub>c</sub>	$\Delta ZPE$	$\delta E_{\text{def}}$
H <sub>2</sub> O	Cry Ads	50.4	63.7	-13.4	5.2
	Cry Des	71.1	86.5	-15.3	15.7
NH <sub>3</sub>	Cry Ads	55.2	67.5	-12.3	7.5
CH <sub>3</sub> OH	Cry Ads	60.1	70.2	-10.1	6.0

Note: The samples are obtain at full B97-3c level of theory. We remind that the normal BE (Eq. 3) is the sum of BE<sub>c</sub> and  $\delta E_{\text{def}}$ .

## V. RESULTS

The test cases described in the previous section were used to study the dependency of the different prefactors on simulated TPD spectra, the thermal correction to the BE and the vibrational partition function on the desorption rate, respectively. In Tables I and II, we present an overview of the BE, for molecular and periodic systems, respectively, and its decomposition according to Eq. 3. In Table III, we show the computed properties regarding the TPD spectra.

### A. Thermal correction of the binding energy

In computational astrochemistry, it is common to report the BE corrected only for the ZPE (Eq. 3) and without any thermal correction (Eq. 4). This approximation makes sense when temperatures are very low, such as in cold molecular clouds ( $\approx 10$  K), but at higher temperatures thermal correction may become not negligible. Using the samples described in Tables I and II, we computed and reported in Fig. 3 the aforementioned contributions, identified as  $\Delta BH(T)$ , as a function of the temperature.

From the  $\Delta BH(T_{\text{peak}})$  column of Table III and the left panels of Fig. 3 (where the  $\Delta BH(T)$  is presented in function of the temperature), it can be seen the relatively limited contribution of the thermal correction to the BH(0). For the majority of the cases, the thermal correction is in the range of 1.5–4 kJ/mol, representing  $\approx 5\%$  of the total BE, to compare with the ZPE

correction which accounts for  $\approx 20\%$  of the BE.

The periodic “Cry Des” sample represents a peculiar case. This is a limit case in which water, composing a perfect crystalline surface, is removed from the latter creating a relatively unstable surface. This sample shows a larger  $\Delta BH(T)$  due to the vibrational frequencies of the reference surface ( $H_{\text{srf}}^{\text{vib}}(T)$ ; in this case the surface with the hole) and in fact the  $H_{\text{srf}}^{\text{vib}}(T)/H_{\text{c}}^{\text{vib}}(T)$  ratio goes from 1 to 0.98 in the considered temperature range and, consequently, the correction is only due to the  $4RT$  contribution since the  $H_{\text{mol}}^{\text{vib}}(T)$  is almost zero for all the studied samples.

Another outlier is the ammonia “Mol 1” case, for which the thermal correction remains low at all studied temperatures, the main contribution of which is due to the important rearrangement of the surface ( $\delta BE_{\text{def}}$  in Table I). This important rearrangement leads to a  $H_{\text{srf}}^{\text{vib}}(T)/H_{\text{c}}^{\text{vib}}(T)$  ratio of 0.27–0.87 in the considered temperature range, thus decreasing the  $4RT$  contribution.

In the right panels of Fig. 3, the differences between a full RRHO (Eq. 5) and q-RRHO (Eq. 6) correction to the BH(0) are reported. The correction is always positive, except for the water Mol 3 and Cry Des cases, in which it is almost constant to 0, meaning that  $\Delta BH(T)(\text{RRHO})$  is always larger than  $\Delta BH(T)(\text{q-RRHO})$ . In almost all other cases, the differences remain restricted to a range of  $\approx 0$ –1 kJ/mol, below the typical desorption temperature (200 K). Due to the low correction using the q-RRHO with respect to the RRHO at our operating temperatures and the fact that q-RRHO theory was developed to correct enthalpies at higher temperatures, we suggest to use the normal RRHO to correct the BH(0).

### B. Vibrational partition function and 6HRT importance

The difference between the immobile particle approximation (Eq. 13, Tait et al. Ref. 6) and the full TST approach (Eq. 14) in computing the desorption prefactor is the  $q_{\text{vib}}^{\text{TST}}$  term decomposed in Eq. 14.

In the present work, we have relied on the vibrational harmonic frequency calculations paradigm. To verify if the surface phonons play a not negligible contribution to the  $q_{\text{vib}}^{\text{TST}}$

TABLE III. Molecular and periodic desorption cases.  $T_{\text{peak}}$  is extracted from the simulated TPD spectra of Fig. 5 and 6, using  $\beta = 0.04$  K/s and  $A = 10^{-19}$   $\text{N}_a/\text{m}^2$ . The prefactors,  $v_{\text{TST}}$  and  $v_{\text{TST}}^{\text{Tait}}$ , are computed at the  $T_{\text{peak}}$  obtained with the simulations carried out with  $v_{\text{TST}}$  prefactor.  $v_{\text{HH}}$  is computed using BH(0) and  $A 10^{-19}$   $\text{N}_a/\text{m}^2$ . The energy values are in kJ/mol, peak temperatures in Kelvin and prefactors in  $\text{s}^{-1}$ .

Species	Sample	BH(0)	$\Delta\text{BH}(T_{\text{peak}})$	$v_{\text{TST}}(T_{\text{peak}})$	$v_{\text{TST}}^{\text{Tait}}(T_{\text{peak}})$	$v_{\text{HH}}(\text{BH}(0))$	$T_{\text{peak}}$	$T_{\text{peak}}^{\text{Tait}}$	$T_{\text{peak}}^{\text{HH}}$
$\text{H}_2\text{O}$	Mol 1	30.9	2.7	$5.9 \cdot 10^{14}$	$1.3 \cdot 10^{15}$	$1.9 \cdot 10^{12}$	105	103	123
	Mol 2	44.6	3.0	$1.4 \cdot 10^{15}$	$4.0 \cdot 10^{15}$	$2.2 \cdot 10^{12}$	146	142	173
	Mol 3	56.6	3.8	$7.0 \cdot 10^{15}$	$8.4 \cdot 10^{15}$	$2.5 \cdot 10^{12}$	180	179	221
	Cry Ads <sup>a</sup>	50.4	3.3	$1.4 \cdot 10^{15}$	$5.9 \cdot 10^{15}$	$2.4 \cdot 10^{12}$	163	158	193
	Cry Des <sup>b</sup>	71.1	6.1	$1.3 \cdot 10^{16}$	$1.5 \cdot 10^{16}$	$2.8 \cdot 10^{12}$	214	213	267
$\text{NH}_3$	Mol 1	24.3	1.3	$1.2 \cdot 10^{14}$	$9.3 \cdot 10^{14}$	$1.7 \cdot 10^{12}$	85	81	95
	Mol 2	34.1	2.1	$2.7 \cdot 10^{14}$	$2.8 \cdot 10^{15}$	$2.0 \cdot 10^{12}$	116	110	132
	Mol 3	44.5	2.8	$1.1 \cdot 10^{15}$	$6.0 \cdot 10^{15}$	$2.3 \cdot 10^{12}$	145	140	171
	Cry Ads <sup>a</sup>	55.2	2.9	$1.3 \cdot 10^{15}$	$1.2 \cdot 10^{16}$	$2.6 \cdot 10^{12}$	177	169	208
$\text{CH}_3\text{OH}$	Mol 1	20.3	1.7	$2.3 \cdot 10^{16}$	$2.8 \cdot 10^{16}$	$1.1 \cdot 10^{12}$	64	64	84
	Mol 2	35.9	1.8	$4.7 \cdot 10^{16}$	$1.7 \cdot 10^{17}$	$1.5 \cdot 10^{12}$	107	104	139
	Mol 3	55.1	1.9	$8.5 \cdot 10^{16}$	$6.7 \cdot 10^{17}$	$1.9 \cdot 10^{12}$	158	151	206
	Cry Ads <sup>a</sup>	60.1	1.4	$2.8 \cdot 10^{16}$	$9.3 \cdot 10^{17}$	$2.0 \cdot 10^{12}$	174	162	221

Note: The molecular samples (Mol) are obtained with ONIOM(DLPNO-CCSD(T)/aug-cc-pVTZ:xTB-GFN2//B97-3c:xTB-GFN2), while the crystalline (Cry) samples are computed with a full B97-3c level of theory.

<sup>a</sup> Crystalline sample “adsorption” on surface.

<sup>b</sup> Crystalline sample “desorption” from surface.

term and the importance of the six hindered rotations and translations (6HRT), we calculated the vibrational frequencies on the periodic systems at three levels: i) on the whole system (“Full”), ii) on the first coordination shell of the desorbed species (“Fragment”, described in Appendix A 2 a), iii) on the adsorbed molecule only, using for the bare surface only its electronic energy (“Mol. Fragment”). In Fig. 4, the  $q_{\text{vib}}^{\text{TST}}$  quantity is only presented for the crystalline (Cry) samples (Table II), due to the 2-layers ONIOM approach and to the constraints applied to molecular cases which could bias this fine analysis. Looking at all the crystalline samples of Fig. 4, the surface phonons do not have an important contribution to the  $q_{\text{vib}}^{\text{TST}}$  term. In other words, surface vibrations do not change in the studied range of temperature before and after the adsorption ( $q_{\text{vib}}^{\text{TST}}(\text{S})$  term in Eq. 14). The term  $q_{\text{vib}}^{\text{TST}}(\text{M})$  was also studied, even if not presented in Fig. 4, and for all the cases it is always  $\approx 1$ , meaning that the intramolecular vibrations for the adsorbed and isolated molecule are never excited even at high temperatures.

To identify the importance of the 6HRT modes for the adsorbed molecules, we performed frequency calculations for the atoms of the desorbed molecule only (6HRT Mol. Fragment). From these calculations, the 6HRT were identified by visual inspection. Due to the coupling with all the other atoms in the “Full” and “Fragment” systems, it was not possible to isolate the 6HRT modes of the desorbing species. For this reason, we do not report these modes for those systems and prefer to use the full  $q_{\text{vib}}^{\text{TST}}$  expression instead of  $1/q_{\text{vib}}^{\text{6HRT}}$ . This behavior is due to the fact that the intermolecular interactions of the molecule composing the surface have the same nature of those between the desorbing molecule and the surface (H-bonds).

From Fig. 4 one can see that the  $1/q_{\text{vib}}^{\text{6HRT}}$  (6HRT Mol. Fragment) gives the same results of the  $q_{\text{vib}}^{\text{TST}}$  calculated for the

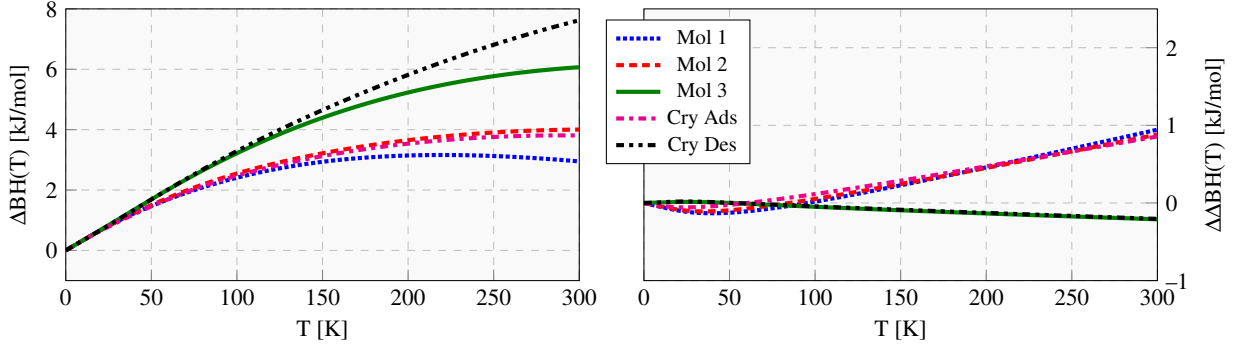
“Full” and “Fragment” systems. The results do not change even if considering only the 6HRT or also the internal normal modes (all vibrations from “Mol. Fragment”). This is particularly important to save a lot of computational resources, due to the large cost of performing a full vibrational analysis of the system. In this case, we have clearly shown that it is sufficient to calculate only the vibrational contribution of the desorbing molecule to obtain a good approximation of the  $q_{\text{vib}}^{\text{TST}}$  and, consequently, of the prefactor.

In addition, we studied the impact of considering the vibrational partition function with Eq. 15. We do not present the results here since we found that no major difference is present with respect to using the normal vibrational partition function formulation (Eq. 12). The details are reported in the Appendix A.

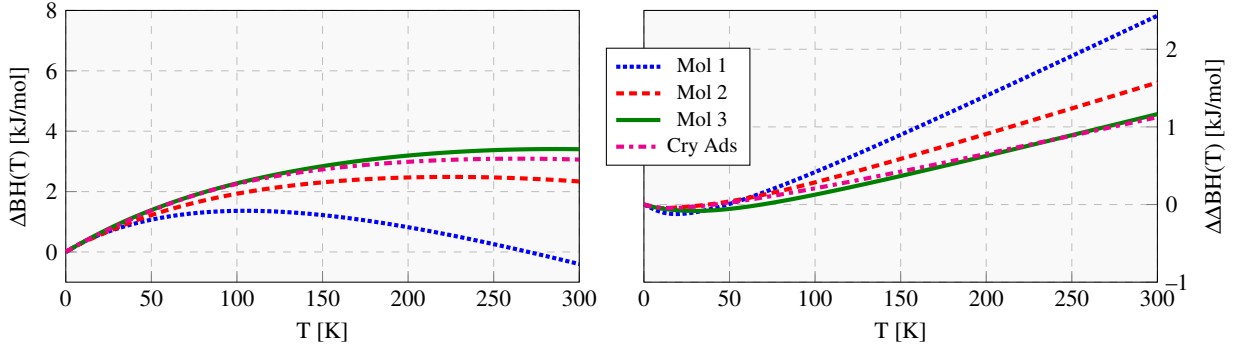
### C. Pre-exponential factor impact on TPD spectra

For each sample, the desorption rate was calculated with the three different prefactor models discussed before:  $v_{\text{TST}}$  (Eq. 11),  $v_{\text{TST}}^{\text{Tait}}$  (Eq. 13, proposed by Tait et al.<sup>6</sup>), and  $v_{\text{HH}}$  (Eq. 9, proposed by Hasegawa & Herbst<sup>16</sup>). The temperature peaks of simulated TPD experiments were calculated according to Eq. 16, and the simulated TPD spectra are reported in Figs. 5 and 6 for the molecular and periodic samples, respectively. Table III summarises the major parameters of Eq. 16.

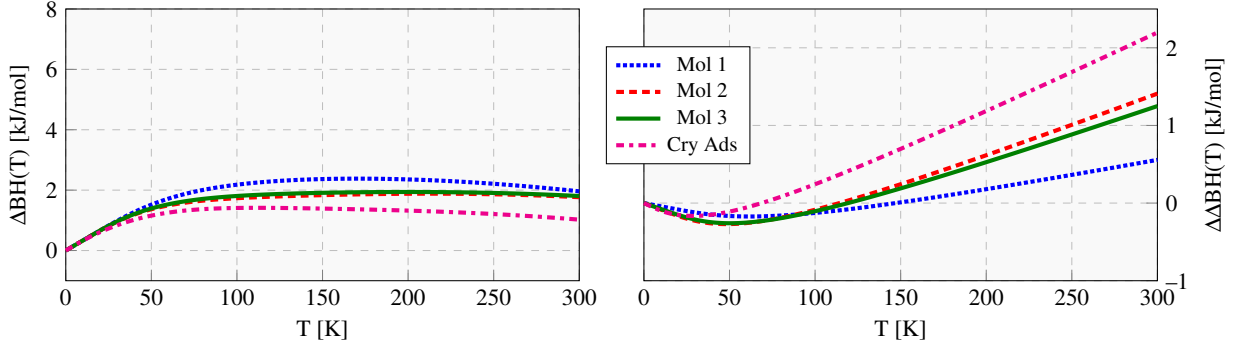
The main message that can be extracted from the simulated TPD spectra is that, for the water and ammonia samples,  $v_{\text{HH}}$  is between 3 and 4 orders of magnitude lower than  $v_{\text{TST}}^{\text{Tait}}$  (calculated at the  $T_{\text{peak}}$ ). The difference is even larger for the methanol samples, where it is 4-5 orders of magnitude lower. This result can be attributed to the oversimplification of  $v_{\text{HH}}$  that, contrarily to the TST approach, does not take into ac-



(a) Thermal correction of water BH(0) (left panel) and comparison with the quasi-RRHO approach (right panel).



(b) Thermal correction of ammonia BH(0) (left panel) and comparison with the quasi-RRHO approach (right panel).



(c) Thermal correction of methanol BH(0) (left panel) and comparison with the quasi-RRHO approach (right panel).

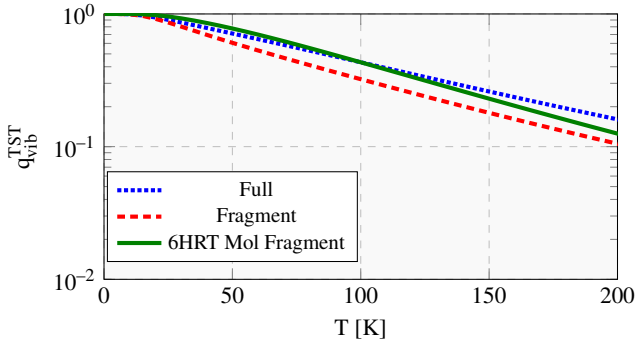
FIG. 3. *Left panels:* Thermal correction of BE ( $\Delta BH(T) = BH(T) - BH(0)$ ) as a function of the temperature (Eq. 4) for all the cases of Table III. *Right panels:* Vibrational enthalpy difference between the RRHO (Eq. 5) and quasi-RRHO (Eq. 6) approaches, *i.e.*  $\Delta\Delta BH(T) = \Delta BH(T)(RRHO) - \Delta BH(T)(q-RRHO)$ .

count the rotational motions. Indeed, as discussed in Tait et al.<sup>6</sup>, the prefactor moment of inertia increases with the molecular size.

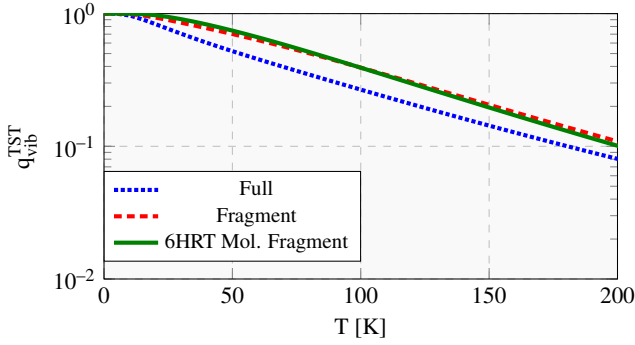
About the difference between  $v_{TST}^{Tait}$  and  $v_{TST}$ , as pointed out in Section III, the immobile particle approximation described by Tait et al.<sup>6</sup> represents an upper limit to the prefactor while the  $v_{TST}$  is, as shown in Fig. 4, between 0 and 1 order of magnitude lower. This difference results in a different  $T_{peak}$  of  $\approx 3K$  for all the molecular samples and  $\approx 12K$  for the periodic ‘‘Cry Ads’’ samples. The discrepancy between the two above-mentioned samples can be attributed to the 6HRT

modes of periodic adsorbed species that are not constrained as in the molecular cases. In the ‘‘Cry Des’’ case, every difference in  $T_{peak}$  is found between  $v_{TST}^{Tait}$  and  $v_{TST}$ , since the  $q_{vib}^{TST}$  is almost equal to 1 in all the studied T range (see Fig. A 2 b in Appendix).

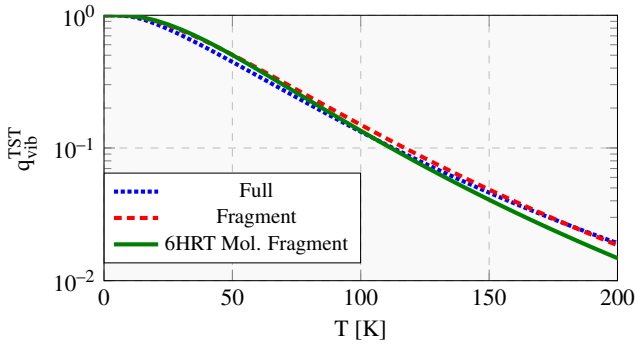
Finally, we want to highlight the effect of the  $\beta$  factor in the simulated TPD experiments. Lower  $\beta$ s correspond to narrower peaks and lower  $T_{peak}$ , and *viceversa*. Therefore, when  $\beta$  decreases the difference in simulated TPDs using different prefactor models will be minimized. This is expected at least



(a) Water vibrational partition function ratio.



(b) Ammonia vibrational partition function ratio.



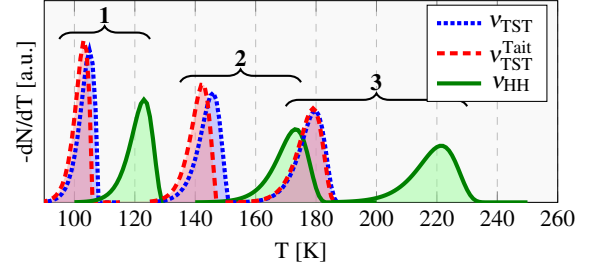
(c) Methanol vibrational partition function ratio.

FIG. 4. Vibrational partition function ratio (Eq. 14) for water, ammonia and methanol in the crystalline (Cry Ads) cases. “Full”, “Fragment”, and “6HRT Mol. Fragment” stand for complete, and partial (first coordination sphere of the desorbing molecule and desorbing molecule only) vibrational frequencies calculations.

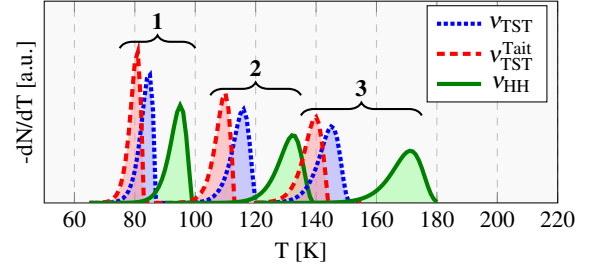
for the TPD computed with  $v_{\text{TST}}^{\text{Tait}}$  and  $v_{\text{TST}}$ .

#### D. Desorption rate Temperature dependencies

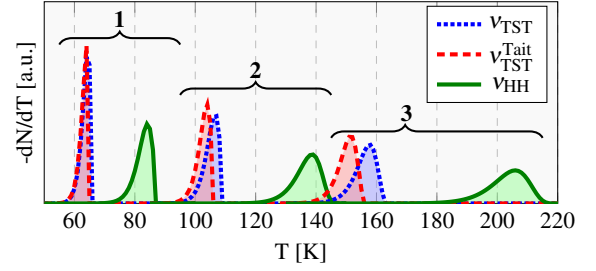
In the majority of the simulated cases, the assumption of modelling  $k_{\text{des}}$  with Eq. 10, with  $v$  and the BH calculated at



(a) Water



(b) Ammonia



(c) Methanol

FIG. 5. Simulated TPD spectra for molecular (Mol) samples. The spectra are computed as first-order desorption and heating rate  $\beta = 0.04\text{K/s}$ . Each sample is identified by the number above the curly bracket.

the  $T_{\text{peak}}$ , *i.e.*:

$$k_{\text{des}}(T) = v_{\text{TST}}(T_{\text{peak}}) \exp\left(-\frac{\text{BH}(T_{\text{peak}})}{R T}\right),$$

describes with very good accuracy the desorption process. In fact, as shown in Fig. 3 and in the  $\Delta\text{BH}(T_{\text{peak}})$  column of Table III, the thermal corrections to the  $\text{BH}(0)$  remain constant in the 50–300 K range for almost all the samples. On the other hand, the prefactor has the highest change (one order of magnitude) in the 50–200 K range for all the considered samples (see  $v_{\text{TST}}(T_{\text{peak}})$  column in Table III for different samples). This behavior is mainly due to the importance of  $T$  in the exponential factor of  $k_{\text{des}}$ , overcoming the thermal correction to the  $\text{BH}(0)$  and to the prefactor. Supporting the above discussion, in an experiment involving a multilayer regime of ethane, Luna et al. 21, experimentally found that the prefactor is dependent on the temperature in a desorption mechanism described by the Polanyi-Wigner equation.

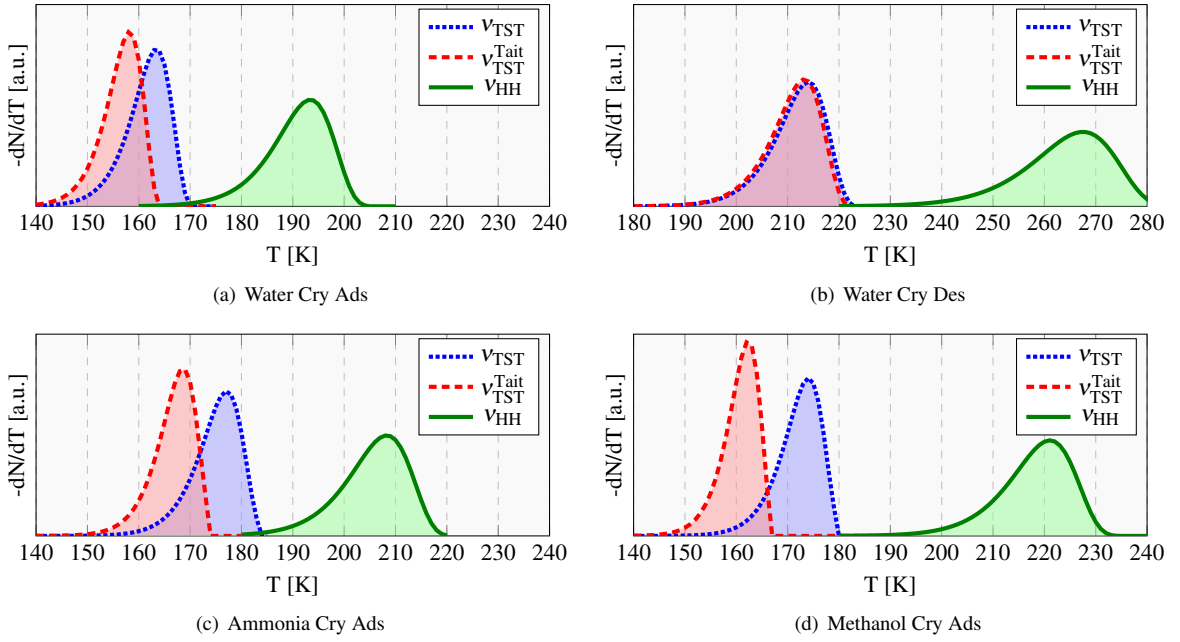


FIG. 6. Simulated TPD spectra for the crystalline (Cry) samples. The spectra are computed as first-order desorption and heating rate  $\beta = 0.04\text{K/s}$ .

To correctly estimate the dependence of  $k_{\text{des}}$  on the temperature, we fitted the calculated rate with the three-parameter modified Arrhenius equation proposed by Kooij<sup>38,39</sup>:

$$k(T) = \alpha \left( \frac{T}{300} \right)^\eta \exp\left( -\frac{\gamma}{RT} \right),$$

where  $\alpha$ ,  $\eta$ , and  $\gamma$  are the fitting parameters. We remind the reader that this expression is used to fit experimental and/or computational results and, consequently, the parameters do not have real physical meaning. For  $\eta = 0$ , we obtain the Arrhenius equation and, accordingly,  $\gamma$  can be approximated to the activation energy of the reaction. Under this assumption, the following equation has been solved to compute the peak temperature:

$$\frac{\gamma}{R} + \eta T - \frac{T^2}{\beta} \alpha \left( \frac{T}{300} \right)^\eta \exp\left( -\frac{\gamma}{RT} \right) = 0, \quad (17)$$

This equation can be easily solved with a numerical algorithm, *e.g.* Bisection or Newton. In Appendix A, we demonstrate the accuracy obtained when the Arrhenius-Kooij equation is used to fit the desorption rate and also the accuracy of the Eq. 17 in the  $T_{\text{peak}}$  calculation.

## VI. CONCLUSIONS

In this work, we study the importance of the pre-exponential factor in the desorption rate from the interstellar icy mantles of some representative species (water, ammonia, methanol), by

means of periodic and molecular DFT simulations. Specifically, we show the effect of using different approaches to calculate the prefactor, from the most used one in the astrochemical community ( $v_{\text{HH}}$ ) to others that explicitly take into account the partition functions of the involved species ( $v_{\text{TST}}$  and  $v_{\text{TST}}^{\text{Tait}}$ ). As expected, for all the samples, either molecular or periodic,  $v_{\text{HH}}$  leads to the typical values discussed in the introduction ( $\sim 10^{12} \text{ s}^{-1}$ ), while  $v_{\text{TST}}$  and  $v_{\text{TST}}^{\text{Tait}}$  may differ from 2 up to 5 or 6 orders of magnitude. These differences largely affect the simulated TPD spectra, where  $v_{\text{HH}}$  and  $v_{\text{TST}}^{\text{Tait}}$  always represent the upper and lower limits, respectively, on the range of desorption temperature. The difference in the  $T_{\text{peak}}$  adopting one approach or another spans from 20 to 60 K, depending on the cases.

In addition, we evaluate the effect of RRHO and q-RRHO approaches to the correction of the vibrational enthalpy, whose difference may become important (up to 8 kJ/mol) at high temperatures (above 100 K).

Finally, we performed frequency calculations either on the full system or only on selected atoms, showing that it is enough to include only the adsorbate molecule in the vibrating atoms to obtain converged values of  $q_{\text{vib}}^{\text{TST}}$ . This is another important result in particular to save computational resources, as performing frequency calculations on all the atoms is very expensive.

## VII. ONLINE SCRIPTS AND DATABASE

The python scripts, with all the examples, can be found in the following GitHub link: <https://github.com/>

**Tinaccil/BE\_prefactor.** To easily handle the data set of BE samples (atomic coordinates and BH(0) values, summarized in Table III), we developed and made publicly available a website based on the molecule hyperactive JSmol plugin (Jmol: an open-source Java viewer for chemical structures in 3D<sup>40</sup>). This extended electronic version of the calculated structures is available at: [https://tinaccil.github.io/Jmol\\_BE\\_prefactor\\_visualization/](https://tinaccil.github.io/Jmol_BE_prefactor_visualization/).

## ACKNOWLEDGMENTS

This project has received funding within the European Union’s Horizon 2020 research and innovation programme from the European Research Council (ERC) for the project “The Dawn of Organic Chemistry” (DOC), grant agreement No 741002, and from the Marie Skłodowska-Curie for the project “Astro-Chemical Origins” (ACO), grant agreement No 811312. PU and SP acknowledge the Italian Space Agency for co-funding the Life in Space Project (ASI N. 2019-3-U.O). CINES-OCCIGEN HPC is kindly acknowledged for the generous allowance of super-computing time through the A0060810797 project. F. Dulieu and P. Theulé are thanks of the insight on the laboratory point of view. LT is grateful to Jacopo Lupi, Leonardo Miele, Franciele Kruczkiewicz and Stefano Ferrero for insightful discussions and to the L<sup>A</sup>T<sub>E</sub>X community for the insights on TikZ and PGFPlots packages. Finally, we wish to acknowledge the extremely useful discussions with Prof. Gretobape.

- <sup>1</sup>J. Fair and R. J. Madix, “Low and high coverage determinations of the rate of carbon monoxide adsorption and desorption from pt (110),” *The Journal of Chemical Physics* **73**, 3480–3485 (1980).
- <sup>2</sup>C. T. Campbell and J. R. Sellers, “The entropies of adsorbed molecules,” *Journal of the American Chemical Society* **134**, 18109–18115 (2012).
- <sup>3</sup>G. Piccini, M. Alessio, J. Sauer, Y. Zhi, Y. Liu, R. Kolvenbach, A. Jentys, and J. A. Lercher, “Accurate adsorption thermodynamics of small alkanes in zeolites. ab initio theory and experiment for h-chabazite,” *The Journal of Physical Chemistry C* **119**, 6128–6137 (2015).
- <sup>4</sup>J. Sauer, “Ab initio calculations for molecule–surface interactions with chemical accuracy,” *Accounts of chemical research* **52**, 3502–3510 (2019).
- <sup>5</sup>L. H. Sprowl, C. T. Campbell, and L. Arnadottir, “Hindered translator and hindered rotor models for adsorbates: Partition functions and entropies,” *The Journal of Physical Chemistry C* **120**, 9719–9731 (2016).
- <sup>6</sup>S. L. Tait, Z. Dohnálek, C. T. Campbell, and B. D. Kay, “n-alkanes on mgo (100). ii. chain length dependence of kinetic desorption parameters for small n-alkanes,” *The Journal of chemical physics* **122**, 164708 (2005).
- <sup>7</sup>M. Jørgensen and H. Gronbeck, “Adsorbate entropies with complete potential energy sampling in microkinetic modeling,” *The Journal of Physical Chemistry C* **121**, 7199–7207 (2017).
- <sup>8</sup>A. Bajpai, P. Mehta, K. Frey, A. M. Lehmer, and W. F. Schneider, “Benchmark first-principles calculations of adsorbate free energies,” *Acs Catalysis* **8**, 1945–1954 (2018).
- <sup>9</sup>C. Waitt, A. R. Miles, and W. F. Schneider, “Adsorbate free energies from dft-derived translational energy landscapes,” *The Journal of Physical Chemistry C* **125**, 20331–20342 (2021).
- <sup>10</sup>G. Piccini and J. Sauer, “Quantum chemical free energies: Structure optimization and vibrational frequencies in normal modes,” *Journal of chemical theory and computation* **9**, 5038–5045 (2013).
- <sup>11</sup>G. Piccini and J. Sauer, “Effect of anharmonicity on adsorption thermodynamics,” *Journal of chemical theory and computation* **10**, 2479–2487 (2014).
- <sup>12</sup>M. Rybicki and J. Sauer, “Rigid body approximation for the anharmonic description of molecule–surface vibrations,” *Journal of Chemical Theory and Computation* (2022), 10.1021/acs.jctc.2c00597.
- <sup>13</sup>K. A. Fichthorn and R. A. Miron, “Thermal desorption of large molecules from solid surfaces,” *Physical Review Letters* **89**, 196103 (2002).
- <sup>14</sup>K. Singwi and M. Tosi, “On the velocity autocorrelation in a classical fluid,” *Physical Review* **157**, 153 (1967).
- <sup>15</sup>M. Minissale, Y. Aikawa, E. Bergin, M. Bertin, W. A. Brown, S. Cazaux, S. B. Charnley, A. Coutens, H. M. Cuppen, V. Guzman, *et al.*, “Thermal desorption of interstellar ices: A review on the controlling parameters and their implications from snowlines to chemical complexity,” *ACS Earth and Space Chemistry* (2022), 10.1021/acsearthspacechem.1c00357.
- <sup>16</sup>T. I. Hasegawa, E. Herbst, and C. M. Leung, “Models of gas-grain chemistry in dense interstellar clouds with complex organic molecules,” *The Astrophysical Journal Supplement Series* **82**, 167 (1992).
- <sup>17</sup>S. F. Boys and F. Bernardi, “The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors,” *Molecular Physics* **19**, 553–566 (1970).
- <sup>18</sup>L. Tinacci, A. Germain, S. Pantaleone, S. Ferrero, C. Ceccarelli, and P. Ugliengo, “Theoretical distribution of the ammonia binding energy at interstellar icy grains: A new computational framework,” *ACS Earth and Space Chemistry* **6**, 1514–1526 (2022).
- <sup>19</sup>Y.-P. Li, J. Gomes, S. Mallikarjun Sharada, A. T. Bell, and M. Head-Gordon, “Improved force-field parameters for qm/mm simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies,” *The Journal of Physical Chemistry C* **119**, 1840–1850 (2015).
- <sup>20</sup>D. A. King, “Thermal desorption from metal surfaces: A review,” *Surface Science* **47**, 384–402 (1975).
- <sup>21</sup>R. Luna, C. Millán, M. Domingo, C. Santonja, and M. Satorre, “Experimental study of the frequency factor in the polanyi–wigner equation: The case of c2h6,” *Vacuum* **122**, 154–160 (2015).
- <sup>22</sup>A. F. Voter and J. D. Doll, “Dynamical corrections to transition state theory for multistate systems: Surface self-diffusion in the rare-event regime,” *The Journal of chemical physics* **82**, 80–92 (1985).
- <sup>23</sup>P. Pracht and S. Grimme, “Calculation of absolute molecular entropies and heat capacities made simple,” *Chemical science* **12**, 6551–6568 (2021).
- <sup>24</sup>A. Germain, L. Tinacci, S. Pantaleone, C. Ceccarelli, and P. Ugliengo, “Computer generated realistic interstellar icy grain models: Physicochemical properties and interaction with nh3,” *ACS Earth and Space Chemistry* **6**, 1286–1298 (2022).
- <sup>25</sup>A. Germain, L. Tinacci, S. Pantaleone, C. Ceccarelli, and P. Ugliengo, “Amorphous grain model of interstellar icy grains,” (2022).
- <sup>26</sup>F. Neese, “Software update: the orca program system, version 5.0,” *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8**, e1327 (2018).
- <sup>27</sup>S. Grimme, C. Bannwarth, and P. Shushkov, “A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (z= 1–86),” *Journal of chemical theory and computation* **13**, 1989–2009 (2017).
- <sup>28</sup>C. Bannwarth, S. Ehlert, and S. Grimme, “Gfn2-xtb - an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions,” *Journal of Chemical Theory and Computation* **15**, 1652–1671 (2019), gFN2.
- <sup>29</sup>J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme, “B97-3c: A revised low-cost variant of the b97-d density functional method,” *The Journal of chemical physics* **148**, 064104 (2018).
- <sup>30</sup>Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese, “Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [dlpno-ccsd (t)],” *The Journal of chemical physics* **148**, 011101 (2018).
- <sup>31</sup>R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, “Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions,” *The Journal of chemical physics* **96**, 6796–6806 (1992).
- <sup>32</sup>F. Weigend, A. Köhn, and C. Hättig, “Efficient use of the correlation consistent basis sets in resolution of the identity mp2 calculations,” *The Journal of chemical physics* **116**, 3175–3183 (2002).



- <sup>33</sup>W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," *Journal of molecular graphics* **14**, 33–38 (1996).
- <sup>34</sup>R. Dovesi, V. Saunders, C. Roetti, R. Orlando, C. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. Harrison, I. Bush, *et al.*, *CRYSTAL17 User's Manual* (2017).
- <sup>35</sup>R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, *et al.*, "Quantum-mechanical condensed matter simulations with crystal," *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8**, e1360 (2018).
- <sup>36</sup>F. Pascale, C. M. Zicovich-Wilson, F. López Gejo, B. Civalleri, R. Orlando, and R. Dovesi, "The calculation of the vibrational frequencies of crystalline compounds and its implementation in the crystal code," *Journal of computational chemistry* **25**, 888–897 (2004).
- <sup>37</sup>C. M. Zicovich-Wilson, F. Pascale, C. Roetti, V. Saunders, R. Orlando, and R. Dovesi, "Calculation of the vibration frequencies of  $\alpha$ -quartz: The effect of hamiltonian and basis set," *Journal of computational chemistry* **25**, 1873–1881 (2004).
- <sup>38</sup>D. Kooij, "Über die zersetzung des gasförmigen phosphorwasserstoffs," *Zeitschrift für physikalische Chemie* **12**, 155–161 (1893).
- <sup>39</sup>K. J. Laidler, "A glossary of terms used in chemical kinetics, including reaction dynamics (iupac recommendations 1996)," *Pure and applied chemistry* **68**, 149–192 (1996).
- <sup>40</sup><http://www.jmol.org/>.
- <sup>41</sup>K. Röttger, A. Endriss, J. Ihringer, S. Doyle, and W. Kuhs, "Lattice constants and thermal expansion of h<sub>2</sub>o and d<sub>2</sub>o ice ih between 10 and 265 k," *Acta Crystallographica Section B: Structural Science* **50**, 644–648 (1994).
- <sup>42</sup>J. S. Chickos and W. E. Acree Jr, "Enthalpies of sublimation of organic and organometallic compounds. 1910–2001," *Journal of Physical and Chemical Reference Data* **31**, 537–698 (2002).

## Appendix A: Appendixes

### 1. B97-3c accuracy

In Table IV a comparison between some experimental and calculated properties of P-ice is presented. The match with experimental results is very good for all the properties, in particular the sublimation enthalpy calculated as:

$$\Delta H^{\text{sub}} = \frac{H_{\text{ice}} - n H_{\text{H}_2\text{O}}}{n}$$

where  $H_{\text{ice}}$  and  $H_{\text{H}_2\text{O}}$  are the enthalpy of P-ice and of a gas phase water molecule, respectively, and  $n$  is the number of water molecules composing the ice bulk structure (see details in Appendix). In Figure 7 the computed IR spectrum is also reported, in comparison with the experimental one. As one can see, the intensities are not well reproduced, and this is a very difficult task in general. However, their position is correctly reproduced, in particular that at high wavenumbers, in the zone of the O–H stretching, which is usually overestimated because of the anharmonicity which usually is not accounted for in simulated IR, due the high computational cost. Please note that no scaling factor has been applied to the spectrum.

### 2. Crystalline surface modelling and specifics

According to the results of Figure 8, the slab model already shows the convergence of the surface formation energy with 2/3 layers. As regards the adsorption modelling, the 3 layer was used, and in Figure 9 the BE vs the supercell size is reported. In this case the convergence means that there are no lateral interactions among the adsorbate molecules of neighboring cells. As one can see, between the  $2 \times 1$  and  $2 \times 2$  supercells there is only 1 kJ/mol of difference, thus indicating that the  $2 \times 1$  supercell already leads to converged BE. Similarly, Figure 10 shows a clear convergence of the BE vs the supercell size. In this case the convergence is less clear with respect to Figure 9, however it seems that BEs for the  $3 \times 2$  supercell are similar to those of the  $2 \times 1$ . In Table V another test was carried out about the adsorption only on the top of the slab (top part of Table V) vs the double top-bottom adsorption (bottom part of Table V).

According to the results obtained from the tests, in the main text only results on the  $2 \times 1$  supercell, the adsorption is simulated only on the top of the slab, and only the atomic positions of the structures are free to relax, while the cell parameters are kept fixed to their optimized values on the bulk structure.

#### a. Surface fragments definition

In order to study the dependency of the phonons on the  $q_{\text{vib}}^{\text{TST}}$  term (Eq. 14) we calculated the harmonic frequency for the whole system, a fragment of it and only the adsorbed

TABLE IV. Experimental (Exp) and optimized (B97-3c) geometrical parameters of P-ice. Cell parameters in Å, cell angle in degrees, cell volume in Å<sup>3</sup>, sublimation enthalpy ( $\Delta H_{\text{Sub}}$ ) in kJ/mol.

	a	c	$\gamma$	c/a	Volume	$\Delta H_{\text{Sub}}$
B97-3c	4.4100	7.2228	119.866	1.638	121.7	-59.2
Exp	4.4800 <sup>a</sup>	7.3100 <sup>a</sup>	120.000 <sup>a</sup>	1.632 <sup>a</sup>	127.1 <sup>a</sup>	-59.2 <sup>b</sup>
$\Delta\%$	1.56	1.19	0.11	-0.38	-0.38	-0.07

<sup>a</sup> K. Röttger et al.<sup>41</sup>

<sup>b</sup> J. S. Chickos et al.<sup>42</sup>

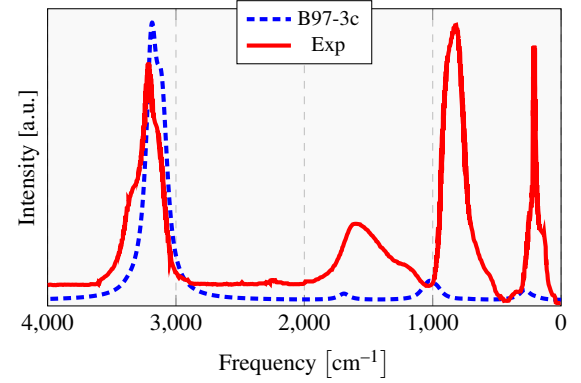


FIG. 7. Normalized IR spectra for P-ice computed at B97-3c level of theory (blue dashed line) and the experimental one (red line). The experimental spectrum is computed at 21 K.

molecules for the periodic samples. The fragment is defined as the first coordination shell of the adsorbed molecules, the case of the Cry ‘‘Cry Ads’’ (for the H<sub>2</sub>O case, the fragment remain the same for all the other samples) is shown in Fig. 11(a) the ‘‘Cry Des’’ in Fig. 11(b).

TABLE V. Different water adsorption cases on crystalline surface.

Sample	BH(0)	BE <sub>e</sub>	$\Delta ZPE$	$\delta E_{\text{def}}$
Cry <sub>TB</sub> <sup>a</sup>	51.1	63.8	12.7	5.2
Cry-A-3l <sub>TB</sub> <sup>a</sup>	70.1	86.0	16.0	16.1
Cry-A-4l <sub>TB</sub> <sup>a</sup>	72.2	86.3	14.1	15.8
Cry-F-3l <sub>TB</sub> <sup>a</sup>	66.9	80.5	13.7	21.6
Cry-F-4l <sub>TB</sub> <sup>a</sup>	68.2	83.2	15.0	18.9
Cry <sub>T</sub> <sup>b</sup>	50.4	63.7	13.4	5.2
Cry-A-3l <sub>T</sub> <sup>b</sup>	71.1	86.4	15.3	15.7
Cry-A-4l <sub>T</sub> <sup>b</sup>	69.9	86.5	16.6	15.7
Cry-F-3l <sub>T</sub> <sup>b</sup>	68.5	81.8	13.3	20.4
Cry-F-4l <sub>T</sub> <sup>b</sup>	68.8	83.7	14.9	18.5

*Note:* TB stands for top-bottom adsorption/desorption, T for only top, A for atomonly optimization (only atomic positions free to relax, cell parameters fixed) and F for full optimization (atomic positions and cell parameters free to relax).

<sup>a</sup> Crystalline sample ‘‘adsorption’’ on surface.

<sup>b</sup> Crystalline sample ‘‘desorption’’ from surface.

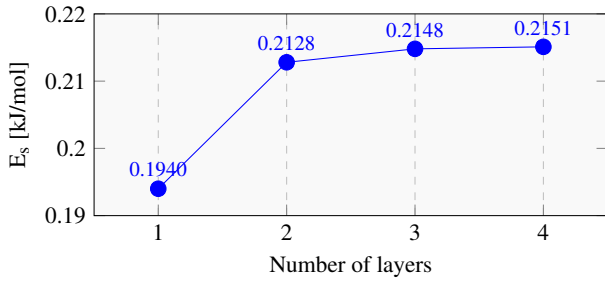


FIG. 8. Surface formation energy ( $E_s$  in kJ/mol) vs number of water layers.

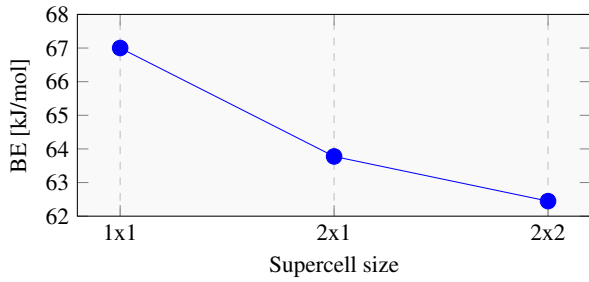


FIG. 9. Binding energy as a function of the supercell size.

### b. Full optimization vs atom only of “Cry Des” sample

As described in section A 2, in the “Cry Des” case two possible optimization strategies were adopted, by setting free to optimize only the atomic position, or both atomic positions and cell parameters. As one can see in Fig. A 2 b, the vibrational frequencies computed on the full system or only on the first coordination sphere with respect to the removed water

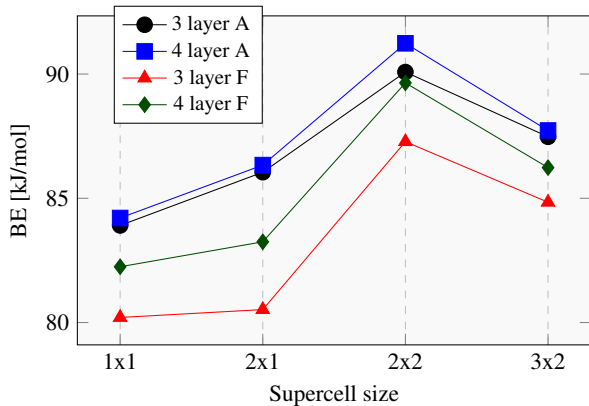


FIG. 10. “Cry Des” sample BE as a function of the supercell size. A stands for atom only optimization (only atomic positions), F stands for full optimization (atomic positions and cell parameters)

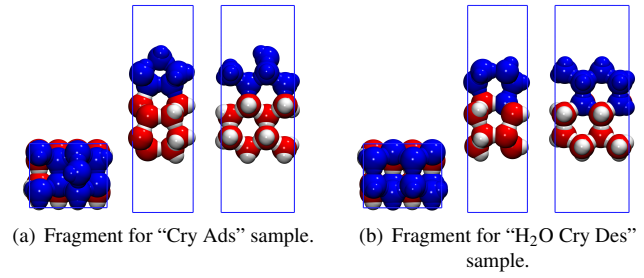


FIG. 11. Orthogonal perspectives of the periodic cell of “Cry Ads” (a) (for the water case) and “Cry Des” (b) samples. In blue the water molecules free to move in the “Fragment” frequency calculations. Atom color legend: oxygen in red, hydrogen in white. Rendering of all the molecule images have been obtained via the VMD software<sup>33</sup>.

(Fragment) give very similar  $q_{\text{vib}}^{\text{TST}}$ , while allowing or not the optimization of the cell parameters leads to different results.

In the case of the “Atom Only” optimization there is no changing in the vibrational frequencies between the reference and the complex giving an almost constant value of  $q_{\text{vib}}^{\text{TST}} \approx 1$  in the entire range of  $T$  (see Fig. A 2 b). There are two possible reasons: i) the 6HRT are at high frequency since the desorbing molecule is part of the crystalline surface, and ii) the surface does not undergo such large rearrangements after optimization and consequently the phonons do not shift. This is in line with the simulated TPD spectra of Fig. 6(b), in which also  $v_{\text{TST}}$  or  $v_{\text{TST}}^{\text{ait}}$  are very similar. The contrary for the optimization of the cell parameters, for which the  $q_{\text{vib}}^{\text{TST}}$  presents a strong dependence on  $T$ . We believe that this behavior is due to the too large concentration of the defective water because of the periodic boundary conditions. Optimizing the cell parameters produces a non-genuine effect on the surface structure, and, even if this does not change dramatically the BE, more sensitive properties like vibrational frequencies and the related partition functions do. As we aim at simulating an isolated desorption, we discuss the “Atom Only” optimization strategy in the main text.

### 3. Vibrational-Rotational partition function for low modes

In Fig. A 3 are presented the  $q_{\text{vib}}^{\text{TST}}$  ratios for desorption mechanism (Eq. 14) using the two different way to compute the vibrational partition function, *e.g.* the standard approach (Eq. 12) and the corrected low motion modes as rotations (Eq. 15, that in the above Figure is addressed as  $m-q_{\text{vib}}$ ). From the plot is clear that the  $q_{\text{vib}}^{\text{TST}}$  using the  $m-q_{\text{vib}}$  approach fail to reproduce the behaviour for  $T \rightarrow 0$  due to the importance of the 6HRT in the desorption. Overall, this correction cannot empirically correct vibrational anharmonicity. In fact, our objective was to lower the preexponential factor to a more free translator like behaviour (lowering the  $q_{\text{vib}}^{\text{TST}}$  ratio to some order of magnitude) more than a immobile particle approximation (increasing the  $q_{\text{vib}}^{\text{TST}}$ ).

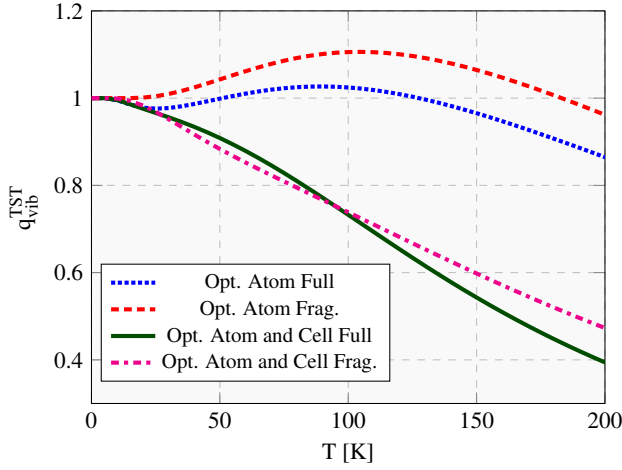


FIG. 12. Vibrational partition function ratio (left member Eq. 14) of “Cry Des” sample computed for all the normal modes (Full) and for the fragment (“Fragment”, see subsec. A 2 a for the definition) adopting both the relaxing strategies: only atomic positions and both atomic positions and cell parameters. The considered system is the periodic 3-layers 2x1 supercell.

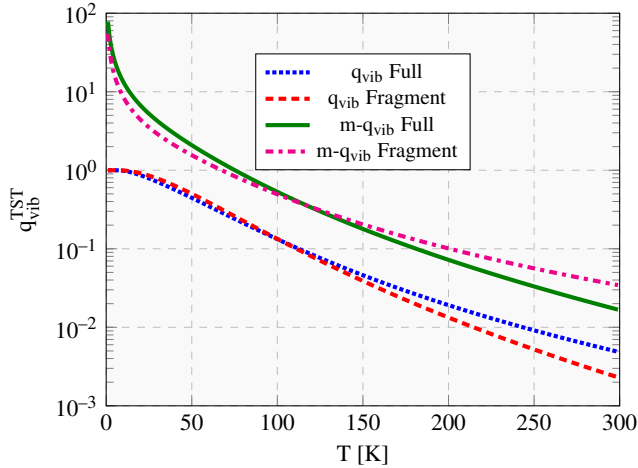


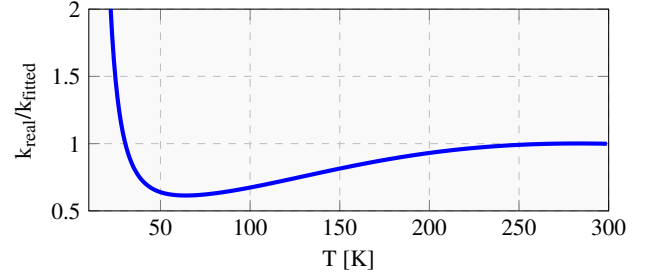
FIG. 13. Vibrational partition function ratio (Eq. 14) for methanol in the crystalline (Cry Ads) case. “Full” and “Fragment” stand respectively for complete, and desorbing molecule only vibrational frequencies calculations. m— $q_{vib}$  means that the vibrational partition function is computed with Equation 15.

#### 4. Kooji-Arrhenius fitting

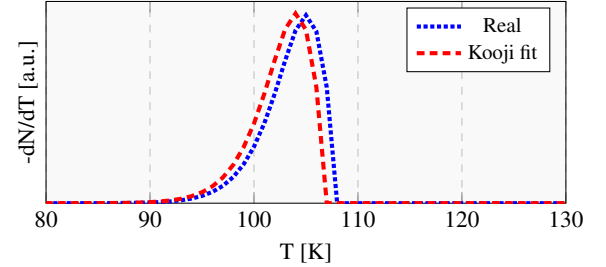
Here the desorption rate fitting from the calculated rate is presented. This allow the computational chemists to compact

easily the  $v(T)$  and the  $BH(T)$  with the Kooji-Arrhenius formalism.

In Fig. A 4 the ratio between the real desorption rate and the fitted one for the water sample 1 (Table III) is shown. The fitting returns the following parameters:  $\alpha = 1.28 \cdot 10^{15} \text{ s}^{-1}$ ,



(a) Desorption rate ratio.



(b) Synthetic TPD spectra.

FIG. 14. Comparison between the Desorption rates,  $k_{des}$ , obtain with full TST approach and the Kooji-Arrhenius fitted. The synthetic TPD spectra are computed as first-order desorption order, heating rate of  $\beta = 0.04 \text{ K/s}$  and with the surface area per adsorbed molecules  $A = 10^{-19} \text{ N}_a/\text{m}^2$ . The considered samples is the water Mol 1 sample (Table III).

$\beta = 2.22 \text{ K}$  and  $\gamma = 31.69 \text{ kJ/mol}$ . From the aforementioned plot, we can see that the Kooji-Arrhenius equation has low accuracy in describing the rate at low temperature. This can be explain in the difficulty to the Kooji-Arrhenius equation to catch the prefactor that has a strong non-linear behavior. Furthermore in Fig. A 4 are presented the two synthetic TPD, in the same condition as Fig. 5, with the fitted rate and the analytic one. The  $T_{peak}$  difference between the two approach is less then 1 K. Using Eq. 17 to infer the  $T_{peak}$ , thought the kooji-Arrhenius fitting parameters aforementioned, the difference between the real  $T_{peak}$  is less then +0.5 K.

Part IV

Conclusions & Perspectives

# Conclusions and future perspectives

## Contents

---

<b>11.1 Summary and conclusions</b>	<b>170</b>
11.1.1 Astrochemical gas-phase network	170
11.1.2 Thermal desorption of astrochemical molecules	171
<b>11.2 Future perspectives</b>	<b>172</b>
11.2.1 Astrochemical gas-phase networks, how improving it more?	172
11.2.2 Computational studies of thermal desorption: what next?	174
<b>11.3 Final remarks</b>	<b>175</b>

---



## 11.1 Summary and conclusions

In this thesis I have investigated two important aspects of the astrochemical reaction networks: the identification of erroneous reactions in the gas-phase and the extension of our knowledge on the thermal desorption of interstellar molecules from icy surface.

In Section 1.3 a general overview of the reaction networks criticalities and open questions before the start of this thesis was provided, while in Subsection 1.4.2 I described its objectives. Thus, in this last chapter, I will shortly discuss the conclusions on the light of the obtained results and their importance in astrochemistry.

### 11.1.1 Astrochemical gas-phase network

The gas-phase reaction network is a crucial element of any astrochemical model, but only a small fraction ( $< 20\%$ ) of the reactions in the network have been studied via laboratory or theoretical studies. The first main objectives this thesis aimed to achieve is a deeper understanding of the reactions and species inside the gas-phase reaction network and consequently produce a more reliable version of it.

My effort to make some order in the astrochemical gas-phase reaction networks can be summarised as follow. Using electronic structure calculations I characterized all the species present in our up-to-date reaction network (based on Wakelam et al. (2015) and with many additions and corrections), and, consequently, I built a database with the chemical properties of each species<sup>1</sup>. Almost all the species that did not have chemical information (structures, electronic state, etc..) were cations. Using this database, I computed the reaction energy of each reaction present in the above-mentioned reaction network to find and delete erroneous reactions.

I found that the  $\sim 5\%$  (342) of the network reactions with thermodynamic data (6275) are endothermic reactions not recognised as such, and almost  $2/3$  of them are reported as barrierless. After the correction for these erroneous reactions, a new gas-phase reaction network, called GRETOBAPE, was created, tested and made publicly available to the community<sup>1</sup>.

The impact of the erroneous reactions removal was studied running astrochemical models in the typical conditions of molecular clouds and warm outflows. I found that the Silicon gas-phase reaction network is the most impacted one by the “*cleaning*”. This result can be explained on the light of

---

<sup>1</sup>[https://github.com/TinacciL/GreToBaPe\\_Cleaning](https://github.com/TinacciL/GreToBaPe_Cleaning)

the number of deleted reactions that involved Si-bearing species ( $\sim 15\%$  of the total).

### 11.1.2 Thermal desorption of astrochemical molecules

Thermal desorption is of crucial importance in astrochemical models since it rules whether a molecule is in the gas- or grain-phase. However, unfortunately the chemical parameters that control this phenomenon (BE and  $\nu$ ) often are poorly known. The second main objectives this thesis aimed to achieve is a more accurate description of the thermal desorption process of astrochemical molecules from icy surface models.

My effort to provide new insights regarding thermal desorption via computational chemistry can be summarised as follow. The two main requirements for accurate adsorption data are: (i) an high level of theory and (ii) an ice grain model with a large surface area and high variability of chemically different adsorption sites. For these reasons, we choose to developed a computational methodology for obtaining not just a few adsorption site cases but a large distribution of them, in a reproducible and unbiased manner, on 200 water molecules icy grain mantle model built via our new ACO-FROST<sup>2</sup> program. Following this methodology, I obtained the adsorption distribution of ammonia and water and compared it with previous experimental and computational studies. The main possible impact of our water and ammonia BE distributions in astrochemistry is related to a low-end BE tail at low energy, which might explain the observed presence of the gas-phase of these molecules in cold regions, such as prestellar core and protoplanetary disk.

Considering the comparison with TPD experimental data, and the limits of this technique, one question rises:

*“Do TPD and computational studies measure the same quantities?”*

After all the studies I did, I think that the answer is no, and, perhaps, the computational data are closer to the process that takes place in the ISM.

With the aim to understand better the thermal desorption, I also studied the commonly used formulae to compute  $\nu$ , as provided by Hasegawa and Herbst (1993), Tait et al. (2005) and its harmonic oscillator extension. I tested their impact with different ice surface, *i.e.* ice cluster and ice periodic surface with water, and different adsorbing molecules, *i.e.* ammonia and methanol, simulating TPD experiments in laboratory conditions. The difference in computing the desorption temperature peak, causes an important difference between the Hasegawa and Herbst (1993) and the other two formulae, depending on the cases.

---

<sup>2</sup>[https://github.com/aurelegermain/ACO-FROST\\_grain\\_generator](https://github.com/aurelegermain/ACO-FROST_grain_generator)

## 11.2 Future perspectives

After the results and the knowledge achieved during this thesis, my perspective on some possible future improvement of the reaction network are here provided.

As described in Section 1.3, the uncertainties analysis techniques are one of the more important tools in the modellers hands. In fact, once key reactions (or other chemical parameters) are identified as a possible source of uncertainty, they can be reported to experimental or computational experts in order to control their feasibility. That said, these techniques are limited by the unfeasible computational cost required to sample the parameters space. I think that a possible improvement of such techniques can be obtained from the topology of the reaction network (Heyl et al., 2020). Indeed, using a graph representation for the reaction network provides new information (the degree of the nodes, centrality, etc., see Section 2.2 for more information), which can guide the analysis of uncertainties. A similar approach was applied in the reaction network dimension reduction problem (Grassi et al., 2013). Network science (*e.g.* Barabási, 2013) can also be useful since it allows to visualize and further quantify complex interactions as they unveil systemic patterns instead of addressing individual reactions only (Ruf and Danger, 2022). Another possible improvement can be obtained from the Data-Science field. In fact, a lot of attention has been addressed on the uncertainties analysis techniques applied to the Artificial Intelligent (AI) field, more specifically to Artificial Neural-Networks due to their increasing importance (Kabir et al., 2018; Gawlikowski et al., 2021). For this reason, new implementations are expected to be applied to astrochemical models in the future.

Specific arguments for gas-phase reaction networks and thermal desorption process will be discussed in details in the next Subsections.

### 11.2.1 Astrochemical gas-phase networks, how improving it more?

The main short term and straightforward prospective for improving astrochemical gas-phase network involves the Silicon reaction network. As evidenced by the impact of the cleaning process on this network (Chapter 5), the necessity to review all the reactions involving Si-bearing species has been emphasized. At the same time, the extension of this network with new reactions from literature and new experimental/computational studies is mandatory to fill in the “holes”.

Another possibility for improving the gas-phase networks reliability involves the database of chemical information, that I created during this thesis. This database can be further improved by the addition of new species and enrichment of chemical properties/information of each of them (*e.g.*, Todeschini and Consonni, 2008; Landrum, 2016). Consequently one possible prospective can be the identification of structure-properties relationship (*e.g.*, Montavon et al., 2013; Fraux et al., 2020; Heyl et al., 2020), which can provide new points of view on the relationship between species inside the reaction network.

Furthermore, the database can be used to rapidly test the endothermicity of new possible formation and destruction reactions postulated by astrochemists or guessed via dedicated programs. Finally, some of these new reactions can be further refined via time-consuming experiments or theoretical studies for estimating rate constants.

A systematic and extendable possible solution to obtain accurate rate constants, without time-consuming and case-by-case studies, is represented by automatic kinetic programs and protocols (for the exploration of the potential energy surface), the importance of which has been growing in the last decade (*e.g.*, Martínez-Núñez, 2015; Simm et al., 2018; Martínez-Núñez et al., 2021). Another possibility is represented by newly reaction network generators (more focused to empirical and data-driven approach) which can provided rate constants and new reactions, as widely applied in combustion chemistry (Vandewiele et al., 2012; Gao et al., 2016). The combination of both approaches, with the resolution to their criticalities (Cavallotti, 2022), can open the door to a new era for astrochemical reaction networks. However, the problems such programs encounter in dealing with astrochemical reactions are, firstly, the lack of a sufficiently extensive and reliable information in astrochemical database and, secondly, the harsh astrochemical physical conditions, where many rare events and quantum effects are not negligible and difficult to take into account.

A possible application of the above-mentioned improvements can be related to the insertion of new reactions. In this direction and from my point of view, the attention almost exclusively to neutral and detected species in the reaction networks, as evidenced also by the lack of information on cations, could have lead to missing reactions involving intermediate cations and isomers that did not have as much as attention in the past years. In fact, even if not detected or detectable, these species could also have a huge impact in models.

## 11.2.2 Computational studies of thermal desorption: what next?

Beside the ammonia and water distribution of adsorption, we applied also our methodology to methanol, for its crucial importance in astrochemistry. The calculations on this last molecule are finished and the article with the results will be finalized in the coming months.

The main short term prospective involves the creation of “*dirty*” icy grain model with the ACO-FROST program. In fact, an icy grain mantle model composed by only H<sub>2</sub>O molecules represents an approximation of the observed icy mantle: CO, CO<sub>2</sub>, NH<sub>3</sub> are other important components (Boogert et al., 2015). Understanding the binding energy distributions of NH<sub>3</sub> and H<sub>2</sub>O on these new mantle models and its comparison with the distributions obtained during this thesis on pure water mantle models can provide new hints on the impact of thermal desorption in the ISM.

Following the conclusions on the impact the water distribution in defining the protoplanetary disk snow-line (Chapter 9), it will be interesting to see how our conclusions, based on simple protoplanetary disk model, will change when a more complex model is used, which takes into account the many processes occurring in the protoplanetary disk (*e.g.*, dust accretion, fragmentation and migration, or magnetic field and cosmic rays, etc.. Henning and Semenov, 2013). In this direction, another interesting possibility would to consider our distribution of adsorption sites inside an astrochemical stochastic surface model.

A possible future extension to the prefactor discussion (Chapter 10), will be the computation of accurate anharmonic corrections to the vibrational normal modes (*e.g.*, Waitt et al., 2021; Rybicki and Sauer, 2022) to our peculiar systems. Indeed, these corrections have never been applied to adsorption of molecules on a “*soft*” surface like our water icy surfaces.

Many improvements over the years have increase the accuracy of computational studies keeping the computational cost low. I would expect in the forthcoming years an improvement for methods based on DLNPO (or PNO) (Guo et al., 2018) theory or new functionals with the Grimme’s 3c correction (Grimme et al., 2021), which can be applied to astrochemical problems.

In the same vein, it is worth mentioning the impact of machine learning (ML) in computational chemistry. An example is provided by the machine learning (ML) force field (Poltavsky and Tkatchenko, 2021), which can be used for MD simulation with DFT accuracy but at classical force field computational cost. Another ML example is  $\Delta$ -ML. This last method aims to add a local or non-local correction to a property obtained using an efficient and, thus, computational cheap low-level ab initio theory (*e.g.*, Chmiela et al.,

2018; Stohr et al., 2020; Nandi et al., 2021). This technique can be seen as the future evolution of the ONIOM method. Close to the astrochemical necessity for improving BE database, an example of the power of  $\Delta$ -ML is presented by Jensen et al. (2022), who carried out a massive computational study of the binding energies of atmospheric molecular clusters.

### 11.3 Final remarks

Regarding the gas-phase reaction network, with the simple endothermicity principle, my thesis revealed that many source of errors are present in the gas-phase reactions network. The astrochemical community is aware of this problem, but for the first time, we can prove it with numbers. For this reason, I think that in addition to the important (and more attractive) task of enriching the networks with new reactions to explain the newly detected species, the revision of old reactions is also of crucial importance to provide a more reliable gas phase reaction network.

On the grain-phase reaction network side, with a new computational strategy, I obtained distributions of adsorption sites which can be used to provide a more complete view of the complexity of the thermal desorption. Despite the number of approximations assumed in astrochemical surface models, I think that the extension from a single BE value to a distribution of values and the used of the new  $\nu$  formulation are needed and, maybe, can provide new clues in the understanding of many astronomical observations.

I think that my thesis has only scratched the tip of the iceberg that represents the identification of erroneous reactions in the gas-phase reaction network and the thermal desorption in the ISM. There are plenty of issues that are still waiting to be understood and tackled with new techniques, as described above.

Computational astrochemistry promises to be a very important and likely unique tool in the years to come.



## Part V

# Résumé en Français de la Thèse et des Résultats

# Résumé de thèse

La façon dont les molécules se forment, survivent et prospèrent dans le milieu interstellaire est au centre des préoccupations fascinantes de l'astrochimie. Les observations astronomiques ont détecté une multitude de molécules ( $\geq 270$ ) possédant jusqu'à 13 atomes (sans tenir compte des HAP). De plus, les modèles astrochimiques tentent de simuler les processus qui régissent l'évolution chimique interstellaire, fournissant des prédictions pour reproduire les observations. Entre les deux, il y a la compréhension de la chimie qui se produit dans les conditions interstellaires très particulières, par rapport aux conditions terrestres.

Les réseaux de réaction sont l'épine dorsale des modèles astrochimiques et l'un des éléments les plus cruciaux où se concentrent les connaissances chimiques. Cependant, en raison de leur complexité et de leur comportement non linéaire sur la modélisation astrochimique, les réseaux de réaction peuvent également être la première source d'erreur dans les simulations, si des données chimiques incorrectes sont présentes. Au cours de ma thèse, j'ai essayé de "*mettre de l'ordre*" dans les réseaux de réactions astrochimiques couramment utilisés en utilisant la chimie numérique comme outil. Le but de ma thèse peut être divisé en deux objectifs immédiats: (1) améliorer les réseaux de réactions chimiques en phase gazeuse, identifier d'éventuelles réactions incorrectes, et (2) améliorer les réseaux de réactions chimiques en surface des grains, fournissant de nouvelles informations sur le processus de désorption thermique interstellaire.

*Objectif 1:* Les réseaux en phase gazeuse actuels (*e.g.*, ceux des bases de données KIDA et UMIST) contiennent environ 8000 réactions impliquant plus de 500 espèces. Parmi celles-ci, moins de 20% ont été étudiées en laboratoire ou théoriquement, et pas seulement dans la gamme des densités ( $10^3$ – $10^{10}$ ) et des températures (10–200 K) interstellaires. Ainsi, mon premier objectif était de "*nettoyer*" les réseaux de l'erreur la plus évidente possible: la présence de réactions endothermiques. À cette fin, à l'aide de calculs de mécanique quantique (QM), j'ai d'abord créé une base de données de struc-

tures et de propriétés pour presque toutes les molécules présentes dans les réseaux de réaction en phase gazeuse couramment utilisés. Puis, avec ces données, j'ai effectué un contrôle de l'endothermicité pour chaque réaction présente dans le réseau de réaction en phase gazeuse généré par notre groupe au fil des années, qui s'appuie sur la base de données KIDA. J'ai constaté que  $\sim 5\%$  des réactions sont endothermiques non reconnues comme telles, dont les 2/3 sont signalées comme sans barrière. L'impact de la suppression des réactions incorrectes a été étudié via la modélisation des régions froides et chaudes typiques, découvrant que le réseau de réaction impliquant des espèces porteuses de Si est le plus impacté. Après la suppression des réactions incorrectes, un nouveau réseau de réaction en phase gazeuse, appelé GRETOBAPE, a été créé et rendu public. Dans la foulée, j'ai également rendu public un réseau réduit à utiliser lorsque seules les principales espèces chimiques doivent être calculées.

*Objectif 2:* Deux grandeurs majeures régissent la désorption thermique des espèces des glaces interstellaires : l'énergie de liaison des espèces (BE) et le facteur pré-exponentiel ( $\nu$ ). Habituellement, les bases de données astrochimiques ne fournissent qu'une seule ou très peu de BE et le  $\nu$  est calculé à partir de formules approximatives. Les BE et, dans de très rares cas, les  $\nu$  sont généralement obtenus via des expériences ou des études numériques qui ne considèrent qu'une ou quelques molécules d'eau pour simuler le manteau de glace des grains interstellaires. Cependant, les surfaces des glaces interstellaires sont probablement irrégulières et une multitude de sites d'adsorption différents doivent exister. Ainsi, afin de mieux simuler la désorption des molécules des glaces interstellaires, j'ai développé une nouvelle méthodologie pour calculer une distribution précise et non biaisée des sites d'adsorption et des BEs et  $\nu$  correspondants. Premièrement, un nouveau grand modèle de manteau glacé de grains a été obtenu via le nouveau programme ACO-FROST, développé par notre groupe et auquel j'ai largement contribué. Deuxièmement, j'ai obtenu une distribution des sites d'adsorption pour l'ammoniac et l'eau, où j'ai calculé le BE et  $\nu$ . L'extrémité inférieure et supérieure de la distribution BE de l'eau et de l'ammoniac pourrait avoir un impact majeur sur l'abondance gazeuse prévue de ces deux espèces. Par exemple, cela pourrait expliquer la présence observée dans la phase gazeuse de ces deux molécules dans les régions froides, telles que les noyaux préstellaires et les disques externes protoplanétaires, où la désorption thermique ne devrait pas être efficace sur la base des valeurs BE uniques de la littérature précédente. De plus, j'ai étudié l'impact des formules couramment utilisées pour  $\nu$  pour calculer le taux de désorption thermique dans les expériences de désorption à température programmée (TPD). À cette fin, j'ai simulé des expériences TPD en utilisant le modèle de glace amorphe utilisée pour dériver

les distributions BE d'ammoniac et d'eau, décrites ci-dessus, et un modèle de glace cristalline. J'ai découvert que l'utilisation de formules différentes entraînait des différences importantes dans la description du processus de désorption.

Enfin, ma thèse n'a fait qu'effleurer la pointe de l'iceberg des données astrochimiques utilisées pour simuler les processus chimiques se produisant dans le milieu interstellaire. J'espère que ma contribution a ouvert la voie à des études plus systématiques des réseaux de réactions en phase gazeuse et du taux de désorption thermique des molécules interstellaires, pour fiabiliser les prédictions des modèles astrochimiques.

**Mots clés:** Milieu Interstellaire, Astrochimie, Chimie de Gas, Chimie de Surface, Network de Réaction, Chimie Computationnelle.

## Résumée de résultats Part II

**Contexte** Les réactions en phase gazeuse sont d’une importance fondamentale pour prédire les abondances moléculaires dans les modèles astrochimiques. Le rôle et l’orientation de ce réseau de sous-réactions en astrochimie peuvent être quantifiés par les réactions ( $\sim 8000$ ) et les espèces ( $\sim 500$ ) présentes dans la base de données KIDA<sup>1</sup>, qui est de mon point de vue la base de données la plus à jour d’informations chimiques en astrochimie. Les astrochimistes ont progressivement ajouté de nouvelles réactions au fur et à mesure de la détection de nouvelles espèces, augmentant la complexité du réseau réactionnel. Dans ce cadre, je me suis rendu compte que presque toute l’attention est focalisée sur l’explication des espèces nouvellement identifiées, en se concentrant sur les réactions de formation de ces espèces ou de leurs précurseurs, tandis que l’épine dorsale du réseau réactionnel (la partie la plus ancienne) a moins de considération.

Le nombre croissant de réactions augmente collatéralement le nombre d’espèces. Ces espèces sont constituées principalement d’ions positifs et d’espèces neutres. Sur le nombre total d’espèces présentes dans la base de données (ou réseau) astrochimique, plus de la moitié sont des ions positifs, qui jouent ainsi un rôle clé dans la chimie des nuages moléculaires froids. Malgré leur nombre, avant ma thèse, de nombreux cations n’étaient pas caractérisés par des expériences ou des études théoriques, comme en témoigne le fait que, pour beaucoup d’entre eux, la seule information disponible dans la base de données KIDA est leur nom. Je pense que ce manque d’information est principalement dû à leur rôle précurseur dans le réseau de réaction en phase gazeuse et au faible nombre de détection: sur  $\sim 270$  espèces interstellaires détectées, environ 30 sont des cations (McGuire, 2022). Les cations dans les réseaux se comportent principalement comme des intermédiaires dans la formation d’espèces plus “*importantes*” détectées, et leur faible détection peut également s’expliquer par la difficulté de dériver leurs propriétés spectroscopiques. D’autre part, une grande attention est accordée aux espèces

---

<sup>1</sup><https://kida.astrochem-tools.org/>

neutres, et presque toutes ont été caractérisées théoriquement par [Woon and Herbst \(2009\)](#).

Les réactions rapportées dans les réseaux astrochimiques sont basées sur des études en laboratoire, des calculs de mécanique quantique (QM), des principes généraux (*e.g.* théorie de la capture pour estimer les coefficients de réaction), ainsi que des réactions déduites de l'intuition chimique et du principe de similarité. Cependant, parfois aucune enquête sur leur exothermicité, ou leur faisabilité dans les conditions interstellaires, n'a été réalisée. En fait, cette dernière affirmation est étayée par le fait que seulement  $\sim 20\%$  des réactions ont été correctement étudiées via des expériences ou des études informatiques. Comme déjà mentionné, je pense que le contrôle devrait être consacré aux réactions dorsales du réseau. En effet, compte tenu du comportement non linéaire du réseau de réaction, il se pourrait qu'une réaction considérée comme sans importance puisse apporter une contribution drastique aux sorties du modèle ([Wakelam et al., 2006](#)), et par conséquent valider l'importance du contrôle.

**Objectifs et innovations de la recherche** L'objectif principal de cette partie II de ma thèse est d'identifier les réactions erronées dans le réseau de réaction en phase gazeuse en effectuant un contrôle exo-/endothermique étendu des réactions et, en pratique, de fournir un nouveau réseau de réaction, qui est nettoyé des réactions endothermiques non explicitement reconnues comme telles. Ce contrôle étendu n'a jamais été appliqué aux réactions astrochimiques, également en raison du manque d'informations chimiques spécifiques. En fait, et non d'importance secondaire, la création d'une base de données de données physico-chimiques obtenues via une méthode chimique QM précise est le premier objectif de cette partie II. Il est donc crucial d'atteindre nos objectifs et d'ouvrir la voie à de nouvelles approches.

**Résumé des résultats** Le premier objectif était de créer une base de données de données physico-chimiques précises pour les cations, comparables en termes de méthodologie à celles disponibles pour les espèces neutres ([Woon and Herbst, 2009](#)). Pour atteindre cet objectif, j'ai développé une méthodologie computationnelle pour lier une structure à chaque nom d'espèce<sup>2</sup> et, par conséquent, toutes les informations physico-chimiques au niveau CCSD(T) /aug-cc-pVTZ//M06-2X/cc-pVTZ. Pour chaque cation, j'ai calculé le groupe de symétrie, l'état électronique, la multiplicité de spin électronique, les constantes de rotation, le moment dipolaire, l'énergie électronique absolue, l'énergie du point zéro. Au total, j'ai effectué des calculs ab

---

<sup>2</sup><https://github.com/TinacciL/FromGraphToXYZ>



initio de 262 cations. Toutes ces données ont été fusionnées avec les propriétés des espèces neutres pour créer la base de données GreToBaPe<sup>3</sup>.

Le second et majeur objectif de la partie II de cette thèse est de produire un réseau réactionnel en phase gazeuse “*nettoyé*” à partir de réactions endothermiques non rapportées comme telles. Le nouveau réseau de réaction s’appelle GRETOBAPE, construit à partir du KIDA2014 accessible au public (Wakelam et al., 2015) avec l’ajout de plusieurs nouvelles réactions et corrections d’études plus récentes. Avec les informations provenant de la base de données susmentionnée, j’ai créé une procédure basée sur la théorie des graphes pour gérer facilement le nettoyage du réseau de réaction et visualiser/étudier ses propriétés<sup>3</sup>. J’ai trouvé que les  $\sim 5\%$  (342) des réactions du réseau original avec des données thermodynamiques (6275) sont des réactions endothermiques non reconnues comme telles, et près des 2/3 d’entre elles sont signalées comme sans barrière. Le réseau de réaction au silicium est le plus affecté par le processus de nettoyage du réseau. Comme le montre le nombre de réactions supprimées impliquant des espèces contenant du silicium ( $\sim 15\%$  du total des réactions supprimées), les espèces supprimées (6 sur un total de 11) et la comparaison des simulations astrochimiques utilisant le réseau de réaction avant et après nettoyage, dans des conditions de nuage moléculaire et de choc d’écoulement moléculaire chaud. De plus, un examen des éventuelles réactions de destruction manquantes entre les espèces neutres et les cations les plus abondants et réactifs (*e.g.*  $\text{H}^+$ ,  $\text{He}^+$ ,  $\text{H}_3^+$ ,  $\text{H}_3\text{O}^+$  et  $\text{HCO}^+$ ) a été fait.

Une version réduite du réseau de réaction nettoyé a également été fournie pour les cas où un grand réseau de réaction n’est pas nécessaire tout en augmentant le temps de calcul, des modèles astrochimiques *e.g.* couplés à des simulations hydrodynamiques de l’effondrement des nuages ou de la chimie des disques protoplanétaires.

**Structure de cette partie de la thèse** Afin d’atteindre les objectifs mentionnés ci-dessus, un article a été publié (Tinacci et al., 2021) et un autre soumis en octobre 2022. Chacun de ces articles constitue les deux chapitres suivants de la partie de cette thèse. Le premier Chapitre (Chap. 4) porte sur la création de la base de données chimiques pour les cations, tandis que le deuxième Chapitre (Chap. 5) traite de l’identification des réactions exothermiques, leur retrait du réseau, l’étude de leur impact sur les modèles astrochimiques et, enfin, la publication d’un nouveau réseau de réaction “*propre*”.

---

<sup>3</sup>[https://github.com/TinacciL/GreToBaPe\\_Cleaning](https://github.com/TinacciL/GreToBaPe_Cleaning)

## Résumée de résultats Part III

**Contexte** Dans tous les processus considérés à l'intérieur du réseau de réaction en phase granulaire dans cette thèse, la désorption thermique est d'une importance cruciale dans les modèles astrochimiques car elle détermine si une molécule est en phase gazeuse ou en phase granulaire. La constante de vitesse de désorption thermique est normalement décrite mathématiquement comme un processus d'Arrhenius dans lequel le facteur pré-exponentiel ( $\nu$ ) et l'énergie de liaison (BE) sont les paramètres qui collectent toutes les informations chimiques du processus, *i.e.*

$$k_{des} = \nu \exp\left(\frac{BE}{RT}\right),$$

où T est la température, R la constante des gaz parfaits. Le BE a un rôle clé dans les modèles astrochimiques puisqu'il est utilisé pour modéliser non seulement la désorption thermique mais aussi la diffusion thermique, puisque les barrières d'énergie de diffusion sont approchées comme étant une fraction du BE (voir Par. 1.2.2). Pour cette raison, de nombreuses études se sont concentrées, principalement, sur la détermination des BE pendant que le  $\nu$  était normalement considéré comme une valeur constante de  $\sim 1 \times 10^{12} \text{ s}^{-1}$  ou calculé via la formule suivante fournie par Hasegawa and Herbst (1993):

$$\nu = \sqrt{\frac{2 BE}{\pi^2 m A}},$$

où A est la surface par molécule adsorbée (normalement supposée être de  $10^{-19} \text{ m}^2$ ) et m est la masse de l'espèce adsorbée. Je pense que le rôle de cette dernière quantité a été sous-estimé par les astrochimistes, surtout quand on l'a comparé à l'attention portée par les scientifiques de la science des matériaux/catalyse.

Quoi qu'il en soit, de nombreuses incertitudes et informations fragmentaires concernant BE et  $\nu$  sont toujours présentes dans les bases de données

astrochimiques, malgré au fil des années, un grand effort de la communauté astrochimique a été fait pour fournir les quantités mentionnées ci-dessus via des études expérimentales ou informatiques. Si les expériences de désorption à température programmée (TPD) (la seule technique utilisée en astrochimie jusqu'à présent) ne peuvent pas fournir d'informations à l'échelle atomique, en raison des méthodes de détection, la chimie computationnelle peut fournir une image plus détaillée et précise du processus de désorption thermique et surmonter certaines des limites expérimentales. Cependant, afin de fournir des données précises via la chimie computationnelle, un modèle de surface fiable et un niveau théorique précis doivent être utilisés.

À mon avis, de nombreux travaux de calcul astrochimiques sur les BE présentent des limitations importantes. Dans de nombreux cas, le modèle de surface de grains glacés ne reçoit pas l'attention qu'il mérite, comme en témoigne le fait que ses propriétés physicochimiques ne sont presque toujours pas fournies. Récemment, quelques modèles de glace avec 1 à 20 molécules d'eau ont été considérés pour calculer les BE de différentes espèces. Cependant, premièrement, ces modèles ne peuvent pas décrire toutes les interactions chimiques entre la molécule absorbée et une "réelle" grande surface d'eau glacée et, deuxièmement, ils n'ont que peu de sites d'absorption.

Pour ces raisons, il est important de fournir une distribution d'absorption pour les espèces astrochimiques importantes, sans oublier d'utiliser un niveau théorique précis et un modèle de surface fiable.

**Objectifs et innovations de la recherche** Du point de vue de la chimie computationnelle pure, l'objectif principal de cette partie est de fournir une image plus précise et complète du processus de désorption thermique des molécules astrochimiques à partir de modèles de surface glacée. Cet objectif peut être atteint avec le développement de: (i) un modèle de manteau à gros grains de glace construit avec des calculs corrects et sa caractérisation physico-chimique, (ii) une méthodologie précise et impartiale pour calculer une distribution de sites de liaison sur le grain glacé et son application sur certaines espèces importantes en astrochimie et (iii) une comparaison détaillée sur les facteurs pré-exponentiels couramment utilisés en astrochimie.

**Résumé des résultats** La première étape pour atteindre nos objectifs était de développer un modèle d'amas glacé d'eau amorphe qui peut imiter le manteau de grain glacé. Ce résultat a été principalement obtenu par Aurèle Germain, doctorant ITN-ACO à Turin, avec qui j'ai collaboré à ce projet. Pour obtenir un modèle précis, nous avons utilisé le nouveau programme semi-empirique GFN-xTB utilisant principalement l'hamiltonien électron-

ique semi-empirique à liaisons fortes GFN2, qui montre une précision et des performances incroyables. À l'aide de ce dernier programme, un script python appelé ACO-FROST<sup>1</sup> a été développé pour construire un modèle de manteau de grains de glace via un ajout étape par étape de molécules d'eau orientées au hasard. Des vérifications minutieuses des propriétés de certains modèles construits ont été effectuées pour évaluer leurs caractéristiques physicochimiques.

Un modèle de manteau de grains glacés de 200 molécules d'eau a finalement été créé et je l'ai utilisé comme point de départ pour le développement d'un nouveau système pour calculer une distribution des sites d'adsorption (et donc leur distribution BE) de toute espèce interstellaire pertinente, de manière reproductible et impartiale. Ce système a une applicabilité générale puisque les calculs de structure électronique reposent sur la méthode ONIOM (voir Subsection 2.1.5), qui permet de choisir le meilleur compromis entre précision et coût de calcul. Toutes les structures et les fréquences harmoniques ont été obtenues au niveau théorique ONIOM(B97-D3/aug-cc-pVTZ:xTB:GFN2), tandis que les raffinements ponctuels (et les corrections BSSE) ont été effectués au niveau ONIOM(DLPNO-CCSD(T)/aug-cc-pVTZ:xTB:GFN2). Nous avons d'abord appliqué cette méthodologie à l'ammoniac et à l'eau<sup>2</sup> ensuite. Ces résultats ont été comparés et discutés par rapport aux résultats antérieurs de la littérature expérimentale et computationnelle. De cette comparaison, j'ai réalisé que les études informatiques ne mesurent pas vraiment les mêmes quantités que les expériences TPD et, peut-être, les données informatiques sont plus proches du processus qui se déroule dans l'ISM. Les distributions BE de l'eau et de l'ammoniac montrent une extrémité inférieure à basse énergie, ce qui pourrait expliquer la présence observée de la phase gazeuse de ces molécules dans les régions froides, telles que le noyau préstellaire et le disque protoplanétaire.

Outre l'amélioration du calcul des BEs précis, moins d'attention a été portée sur *nu* bien qu'il joue un rôle dans l'obtention de la constante de désorption thermique. En utilisant des cas d'adsorption d'eau, d'ammoniac et de méthanol, comme molécules de test, sur des amas de glace et des surfaces périodiques de glace, j'ai testé l'impact de différents facteurs pré-exponentiels sur des expériences simulées de TPD. J'ai comparé la formule issue de Hasegawa and Herbst (1993), Tait et al. (2005) et son extension d'oscillateur harmonique. La différence dans le calcul du pic de température de désorption, dans des conditions expérimentales, en adoptant le préfac-

<sup>1</sup>[https://github.com/aurelegermain/ACO-FROST\\_grain\\_generator](https://github.com/aurelegermain/ACO-FROST_grain_generator)

<sup>2</sup>Dans ce cas, la fonctionnelle DFT utilisée pour les calculs de structure et de fréquence était la B97-3c pour son rapport haute précision/coût de calcul.

teur [Tait et al. \(2005\)](#) et son extension d'oscillateur harmonique est minime. En revanche, la différence entre ces deux dernières approches et la formule [Hasegawa and Herbst \(1993\)](#) va de 20 à 60 K selon les cas.

**Structure de cette partie de la thèse** Afin d'atteindre les objectifs mentionnés ci-dessus, deux articles ont été publiés ([Germain et al., 2022](#); [Tinacci et al., 2022](#)) et deux autres seront soumis prochainement. Chacun de ces articles constitue les quatre chapitres suivants de la partie de cette thèse. Le premier Chapitre (Chap. 7) porte sur la création et la caractérisation d'un modèle de manteau de grains glacés, tout en utilisant ce dernier modèle, le deuxième Chapitre (Chap. 8) sur le développement d'une méthodologie précise de distribution BE et son application sur le cas de l'ammoniac. Dans le troisième chapitre (Chap. 9), l'application de la méthodologie construite au cas de l'absorption d'eau et son impact possible sur la ligne de neige protoplanétaire est présentée. Enfin, le dernier chapitre (Chap. 10) traite d'une comparaison détaillée sur les facteurs pré-exponentiels couramment utilisés en astrochimie.

Part VI  
Appendix



# ACO-VR project: The journey of water to Earth

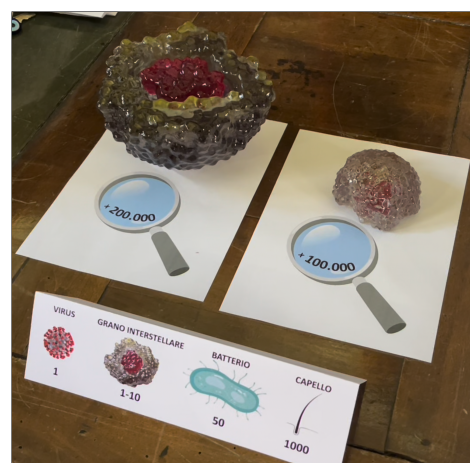
During my thesis, I collaborated with my ACO colleagues to organise and build an astrochemical virtual reality project about the journey of water to Earth. The project was created for outreach purposes, and it has already been displayed at Torino University's Department of chemistry open days (May 2022), the UNIGHT (United citizens for research - Notte Europea delle Ricercatrici e dei Ricercatori) 2022 in Torino (October 2022) and at the Genoa Science Festival (October 2022)<sup>1</sup>. The project will also be presented at the ACO Congress in Toulouse (June 2023). In the following section, I present the project document, which includes the goals, objectives and the different steps of the virtual reality experience. The latter will be offered for free download on the Oculus AppLab store under the name "ACO Water Experience", meanwhile the project source code can be found at the following link:[https://github.com/Marcus-Keil/ACO\\_VR\\_Project](https://github.com/Marcus-Keil/ACO_VR_Project). Other outreach products can be found in the ACO website (<https://aco-itn.oapd.inaf.it/outreach-products>). Some of the activities in which I participated are illustrated in Figure A.1.

---

<sup>1</sup>The INAF outreach group upload a youtube video on our project, at the following link:<https://www.youtube.com/watch?v=vgzibE3QOVg>.



(a) Torino University's Department of chemistry open days (May 2022).



(b) Torino University's Department of chemistry open days (May 2022).



(c) Torino UNIGHT (October 2022).



(d) Torino UNIGHT (October 2022).



(e) Genoa Science Festival (October 2022).

Figure A.1: Some of the activities I carried out during my PhD in the ACO outreach context.

## The journey of water to Earth

*Do you know where our water originally came from?*

*Water is an essential part of life, but where it comes from is still a mystery.*

*This is one of the puzzles that Astrochemistry is trying to solve.*

*With this VR experience, we will walk you through our origins and try to solve this riddle.*

### Context

Astrochemistry is a relatively new science, characterized by its highly interdisciplinary nature. Sciences that are often perceived to be separate from one another collaborate in order to achieve a common goal: understanding the formation of the Universe using chemistry.

The public does not really know about the existence of this science; it cannot be part of the school curriculum due to its cutting edge nature. Our goal as ACO students is to shed light on astrochemistry and to fill the gap between society and this science. Science festivals and similar events need more modern content and methods of outreach.

Virtual Reality is one of the most powerful and recent outreach tools that is growing in interest around the world. As part of the ACO work agreement, it had been decided to conceive and develop an astrochemical VR experience.

Among the astronomical fields, astrochemistry is among the newest. Thus, VR outreach projects like the one proposed here are very rare. In this way, we aim to enlighten the public about astrochemistry, its importance and its goals.

### Goals

- Make people aware of the existence of the ACO project.
- Introduce the audience to our scientific field: astrochemistry.
- Explain, simply and effectively, the cosmic history of water.
- Simply explain the process of formation of a Sun-like star system.
- Trigger the curiosity of the target audience towards the subject in question.
- Enable the audience to experience VR.
- Introduce the audience to the different disciplines within astrochemistry.

## Objectives and Impact

The main target audience will be: young people (for Europe) 15-34 years old that we can interact with at big science festivals, during proposed school activities and other public occasions.

A second level target audience will be scientists - peers - that will be targeted during science congresses. Our qualitative impact studies will focus on our major actions, namely participating in science festivals, European Researchers Night (ERN) and presenting the final product in conferences and congresses, as well as in secondary schools. These fields represent a range of event types and groups of people and as such need to be treated differently in terms of impact measuring.

The events are divided in two categories: firstly, the “drop-in” events gathering an important number of people, like science festivals and ERN; secondly, the presentation of the project to a limited number of people (a few dozen) in specific environments such as congresses, conferences and secondary schools. For the second event type, the groups of people targeted is also important, since we expect the impact made those with high school-level education to be different compared to those at a higher education level.

## Description of the final product

The participant will arrive at the ACO corner at the scientific festivals. She/he will find a desk with ACO flyers and gadgets required for the VR experience. The participant will be welcomed by one of the ESRs. The ESRs will explain the experience and will help the participant to put on the Oculus and to start the VR journey. At the same time, while the participant is playing, other members of the audience will be able to see the experience on a big monitor (also in the ACO corner).

The VR experience consists of three different activities with short explanatory videos. It will be in English. The different steps of the VR experience will be as follows:

1. When the participant puts on the Oculus, she/he will choose one of the available languages in subtitles. An introductory video will explain the context of the upcoming experience via an avatar. *Estimated duration of the video: maximum 1,5 minutes.*
2. First activity: We start with the formation of water, as a VR activity on a first stage. Here, the user can build up the water molecules, within a molecular cloud, as an interactive experience. Pop-ups to show controls during this stage. *Estimated duration of the experience: maximum 3 minutes.*

3. Then, an intermediate video plays, showing how the water molecules are accreted into the ices, along with the evolution of the star. This second video will be focused on the build-up of amorphous water ices. This will range in time from the collapse of the prestellar core until the formation of a protoplanetary disk. *Estimated duration of the video: 30-60 seconds.*
4. Second activity: In this activity, the user will build the “small rocky bodies” together. Here, gravitational forces start to play an important role, with this part of the experience dedicated to emphasizing this concept and bringing the experience closer to reality. *Estimated duration of the experience: maximum 3 minutes.*
5. After this, a third video will show a zoomed-out view of the grains with the ice mantles formed on them (another minute or so), plus a time-lapse of the evolution of the stellar system with the bodies formed, from the protoplanetary disk phase until the formation of a young version of a Sun-like stellar system. *Estimated duration of the video: maximum 1 minute.*
6. A final interactive VR activity will focus on throwing the assembled bodies towards a dry-looking Earth, to create the oceans and explain the context. *Estimated duration of the experience: 2-3 minutes.*

# Bibliography

- Fred C Adams, Charles J Lada, and Frank H Shu. Spectral evolution of young stellar objects. *The Astrophysical Journal*, 312:788–806, 1987.
- Marcelino Agúndez and Valentine Wakelam. Chemistry of dark clouds: databases, networks, and models. *Chemical Reviews*, 113(12):8710–8737, 2013.
- Y Aikawa and E Herbst. Two-dimensional distributions and column densities of gaseous molecules in protoplanetary disks-ii. deuterated species and uv shielding by ambient clouds. *Astronomy & Astrophysics*, 371(3):1107–1117, 2001.
- Ahmet Altun, Frank Neese, and Giovanni Bistoni. Extrapolation to the limit of a complete pair natural orbital space in local coupled-cluster calculations. *Journal of Chemical Theory and Computation*, 16(10):6142–6149, 2020.
- P. Andre, D. Ward-Thompson, and M. Barsony. From Prestellar Cores to Protostars: the Initial Conditions of Star Formation. In V. Mannings, A. P. Boss, and S. S. Russell, editors, *Protostars and Planets IV*, page 59, May 2000.
- Vincent G Anicich. An index of the literature for bimolecular gas phase cation-molecule reaction kinetics. *JPL-Publication-03-19*, 2003.
- Daniela Ascenzi, Andrea Cernuto, Nadia Balucani, Paolo Tosi, Cecilia Ceccarelli, Luca Matteo Martini, and Fernando Pirani. Destruction of dimethyl ether and methyl formate by collisions with  $\text{He}^+$ . *Astronomy & Astrophysics*, 625:A72, May 2019.
- Nadia Balucani, Cecilia Ceccarelli, and Vianney Taquet. Formation of complex organic molecules in cold objects: the role of gas-phase reactions. *Monthly Notices of the Royal Astronomical Society*, 449:L16–L20, April 2015. ISSN 0035-8711.
- Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.
- CA Beichman, PC Myers, JP Emerson, S Harris, R Mathieu, PJ Benson, and RE Jennings. Candidate solar-type protostars in nearby molecular cloud cores. *The Astrophysical Journal*, 307:337–349, 1986.
- Edwin A Bergin and Mario Tafalla. Cold dark clouds: The initial conditions for star formation. *Annu. Rev. Astron. Astrophys.*, 45:339–396, 2007.



- A. C. Adwin Boogert, Perry A. Gerakines, and Douglas C.B. Whittet. Observations of the icy universe. *Annual Review of Astronomy and Astrophysics*, 53:541–581, 2015. ISSN 00664146.
- Giulia Bovolenta, Stefano Bovino, Esteban Vöhringer-Martinez, David A Saez, Tommaso Grassi, and Stefan Vogt-Geisse. High level ab initio binding energy distribution of molecules on interstellar ices: Hydrogen fluoride. *Molecular Astrophysics*, 21:100095, 2020.
- Giulia M Bovolenta, Stefan Vogt-Geisse, Stefano Bovino, and Tommaso Grassi. Binding energy evaluation platform: A database of quantum chemical binding energy distributions for the astrochemical community. *The Astrophysical Journal Supplement Series*, 262(1):17, 2022.
- Christopher J Burrows, Karl R Stapelfeldt, Alan M Watson, John E Krist, Gilda E Ballester, John T Clarke, David Crisp, John S Gallagher III, Richard E Griffiths, J Jeff Hester, et al. Hubble space telescope observations of the disk and jet of hh 30. *The Astrophysical Journal*, 473(1):437, 1996.
- Paola Caselli and Cecilia Ceccarelli. Our astrochemical heritage. *The Astronomy and Astrophysics Review*, 20(1):1–68, 2012.
- Carlo Cavallotti. Automation of chemical kinetics: status and challenges. *Proceedings of the Combustion Institute*, 2022.
- C. Ceccarelli. The Hot Corinos of Solar Type Protostars. In D. Johnstone, F. C. Adams, D. N. C. Lin, D. A. Neufeld, and E. C. Ostriker, editors, *Star Formation in the Interstellar Medium: In Honor of David Hollenbach*, volume 323 of *Astronomical Society of the Pacific Conference Series*, page 195, December 2004.
- C Ceccarelli, L Loinard, A Castets, AGGM Tielens, and E Caux. The hot core of the solar-type protostar iras 16293-2422: H<sub>2</sub>CO emission. *Astronomy and Astrophysics*, 357:L9–L12, 2000.
- C Ceccarelli, P Caselli, F Fontani, R Neri, A López-Sepulcre, C Codella, S Feng, I Jiménez-Serra, B Lefloch, JE Pineda, et al. Seeds of life in space (solis): the organic composition diversity at 300–1000 au scale in solar-type star-forming regions. *The Astrophysical Journal*, 850(2):176, 2017.
- Cecilia Ceccarelli, Claudio Codella, Nadia Balucani, Dominique Bockelée-Morvan, Eric Herbst, Charlotte Vastel, Paola Caselli, Cécile Favre, Bertrand Lefloch, and K Öberg. Organic chemistry in the first phases of solar-type protostars. *arXiv preprint arXiv:2206.13270*, 2022.
- Q Chang, HM Cuppen, and E Herbst. Gas-grain chemistry in cold interstellar cloud cores with a microscopic monte carlo approach to surface chemistry. *Astronomy & Astrophysics*, 469(3):973–983, 2007.
- SB Charnley. Stochastic astrochemical kinetics. *The Astrophysical Journal*, 509(2):L121, 1998.

- SB Charnley. Stochastic theory of molecule formation on dust. *The Astrophysical Journal*, 562(1):L99, 2001.
- Junbo Chen, Jin Kato, Jason B Harper, Yihan Shao, and Junming Ho. On the accuracy of qm/mm models: A systematic study of intramolecular proton transfer reactions of amino acids in water. *The Journal of Physical Chemistry B*, 125(32):9304–9316, 2021.
- Stefan Chmiela, Huziel E Sauceda, Klaus-Robert Muller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature communications*, 9(1):1–10, 2018.
- K-J Chuang, GLEB Fedoseev, C Scirè, GA Baratta, C Jäger, Th Henning, H Linnartz, and Maria Elisabetta Palumbo. Formation of complex organic molecules in molecular clouds: acetaldehyde, vinyl alcohol, ketene, and ethanol via the “energetic” processing of c2h2 ice. *Astronomy & Astrophysics*, 650:A85, 2021.
- Lung Wa Chung, WMC Sameera, Romain Ramozzi, Alister J Page, Miho Hatanaka, Galina P Petrova, Travis V Harris, Xin Li, Zhuofeng Ke, Fengyi Liu, Hai-Bei Li, Lina Ding, and Keiji Morokuma. The oniom method and its applications. *Chemical reviews*, 115(12):5678–5796, 2015.
- Connor W Coley, Wengong Jin, Luke Rogers, Timothy F Jamison, Tommi S Jaakkola, William H Green, Regina Barzilay, and Klavs F Jensen. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chemical science*, 10(2):370–377, 2019.
- HM Cuppen, EF Van Dishoeck, E Herbst, and AGGM Tielens. Microscopic simulation of methanol and formaldehyde ice formation in cold dense cores. *Astronomy & Astrophysics*, 508(1):275–287, 2009.
- HM Cuppen, LJ Karssemeijer, and T Lamberts. The kinetic monte carlo method as a way to solve the master equation for interstellar grain chemistry. *Chemical reviews*, 113(12):8840–8871, 2013.
- Susanta Das, Kwangho Nam, and Dan Thomas Major. Rapid convergence of energy and free energy profiles with quantum mechanical size in quantum mechanical–molecular mechanical simulations of proton transfer in dna. *Journal of Chemical Theory and Computation*, 14(3):1695–1705, 2018.
- Marta De Simone. *Hot Corinos: the early organic molecular enrichment of the planet formation zones*. PhD thesis, Université Grenoble Alpes, January 2022. URL <https://tel.archives-ouvertes.fr/tel-03791884>.
- Marta De Simone, Cecilia Ceccarelli, Claudio Codella, Brian E Svoboda, Claire J Chandler, Mathilde Bouvier, Satoshi Yamamoto, Nami Sakai, Yao-Lun Yang, Paola Caselli, et al. Tracking the ice mantle history in the solar-type protostars of ngc 1333 iras 4. *The Astrophysical Journal Letters*, 935(1):L14, 2022.
- Denis Duflot, Céline Toubin, and Maurice Monnerville. Theoretical Determination of Binding Energies of Small Molecules on Interstellar Ice Surfaces. *Frontiers in Astronomy and Space Sciences*, 8:24, mar 2021. ISSN 2296-987X.

- F Dulieu, L Amiaud, E Congiu, J-H Fillion, E Matar, A Momeni, V Pirronello, and JL Lemaire. Experimental evidence for water formation on interstellar dust grains by hydrogen and oxygen atoms. *Astronomy & Astrophysics*, 512:A30, 2010.
- François Dulieu, Emanuele Congiu, Jennifer Noble, Saoud Baouche, Henda Chaabouni, Audrey Moudens, Marco Minissale, and Stéphanie Cazaux. How micron-sized dust particles determine the chemistry of our Universe. *Scientific Reports*, 3(1):1338, dec 2013. ISSN 2045-2322.
- Theodore Dunham Jr. Interstellar neutral potassium and neutral calcium. *Publications of the Astronomical Society of the Pacific*, 49:26–28, 1937.
- Joan Enrique-Romero. *Surface chemistry of molecules of astrophysical interest : theory and simulations*. PhD thesis, Université Grenoble Alpes & Universitat Autònoma de Barcelona, October 2021. URL <https://tel.archives-ouvertes.fr/tel-03579109>.
- Joan Enrique-Romero, Cecilia Ceccarelli, Albert Rimola, Dimitrios Skouteris, Nadia Balucani, and Piero Ugliengo. Theoretical computations on the efficiency of acetaldehyde formation on interstellar icy grains. *Astronomy & Astrophysics*, 655:A9, 2021.
- Joan Enrique-Romero, Albert Rimola, Cecilia Ceccarelli, Piero Ugliengo, Nadia Balucani, and Dimitrios Skouteris. Quantum mechanical simulations of the radical–radical chemistry on icy surfaces. *The Astrophysical Journal Supplement Series*, 259(2):39, 2022.
- James Fair and Robert J Madix. Low and high coverage determinations of the rate of carbon monoxide adsorption and desorption from pt (110). *The Journal of Chemical Physics*, 73(7):3480–3485, 1980.
- Gleb Fedoseev, Sergio Ioppolo, Dongfeng Zhao, Thanja Lamberts, and Harold Linnartz. Low-temperature surface formation of nh<sub>3</sub> and h<sub>2</sub>co: hydrogenation of nitrogen atoms in co-rich interstellar ice analogues. *Monthly Notices of the Royal Astronomical Society*, 446(1):439–448, 2015.
- Martin Feinberg. *Foundations of chemical reaction network theory*. Springer, 2019.
- S Ferrero, F Grieco, AS Ibrahim Mohamed, F Dulieu, A Rimola, C Ceccarelli, C Nervi, M Minissale, and P Ugliengo. Acetaldehyde binding energies: a coupled experimental and theoretical study. *Monthly Notices of the Royal Astronomical Society*, 516(2):2586–2596, 2022.
- Stefano Ferrero, Lorenzo Zamirri, Cecilia Ceccarelli, Arezu Witzel, Albert Rimola, and Piero Ugliengo. Binding energies of interstellar molecules on crystalline and amorphous models of water ice by ab initio calculations. *The Astrophysical Journal*, 904(1):11, 2020.
- Guillaume Fraux, Rose K Cersonsky, and Michele Ceriotti. Chemiscope: Interactive structure-property explorer for materials and molecules. *Journal of Open Source Software*, 5(51):2117, 2020. URL <https://chemiscope.org/>.
- Connie W Gao, Joshua W Allen, William H Green, and Richard H West. Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications*, 203:212–225, 2016.

- Diego Garay Ruiz, Moises Alvarez Moreno, Carles Bo, and Emilio Martinez Nunez. New tools for taming complex reaction networks: the unimolecular decomposition of indole revisited. *ACS Physical Chemistry Au*, 2022.
- Crispin Gardiner. *Stochastic Methods*. Springer Berlin, fourth edition, 2009.
- Robin T Garrod, Susanna L Widicus Weaver, and Eric Herbst. Complex chemistry in star-forming regions: An expanded gas-grain warm-up chemical model. *The Astrophysical Journal*, 682(1):283, 2008.
- Robin T Garrod, Mihwa Jin, Kayla A Matis, Dylan Jones, Eric R Willis, and Eric Herbst. Formation of complex organic molecules in hot molecular cores through nondiffusive grain-surface and ice-mantle chemistry. *The Astrophysical Journal Supplement Series*, 259(1):1, 2022.
- RT Garrod and E Herbst. Formation of methyl formate and other organic species in the warm-up phase of hot molecular cores. *Astronomy & Astrophysics*, 457(3):927–936, 2006.
- Antonio Garufi, Linda Podio, Claudio Codella, Davide Fedele, Eleonora Bianchi, Cecile Favre, Francesca Bacciotti, Cecilia Ceccarelli, Seyma Mercimek, Kazi Rygl, et al. Alma chemical survey of disk-outflow sources in taurus (alma-dot)-v. sample, overview, and demography of disk molecular emission. *Astronomy & Astrophysics*, 645:A145, 2021.
- Jakob Gawlikowski, Cedrique Rovile Njietcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*, 2021.
- Aurele Germain, Lorenzo Tinacci, Stefano Pantaleone, Cecilia Ceccarelli, and Piero Ugliengo. Computer generated realistic interstellar icy grain models: Physicochemical properties and interaction with nh<sub>3</sub>. *ACS Earth and Space Chemistry*, 6(5):1286–1298, 2022. doi: [10.1021/acsearthspacechem.2c00004](https://doi.org/10.1021/acsearthspacechem.2c00004).
- T Grassi, S Bovino, D Schleicher, and FA Gianturco. Chemical complexity in astrophysical simulations: optimization and reduction techniques. *Monthly Notices of the Royal Astronomical Society*, 431(2):1659–1668, 2013.
- T Grassi, S Bovino, P Caselli, G Bovolenta, S Vogt-Geisse, and B Ercolano. A novel framework for studying the impact of binding energy distributions on the chemistry of dust grains. *Astronomy & Astrophysics*, 643:A155, 2020.
- Stefan Grimme, Andreas Hansen, Sebastian Ehlert, and Jan-Michael Mewes. r2scan-3c: A “swiss army knife” composite electronic-structure method. *The Journal of Chemical Physics*, 154(6):064103, 2021.
- Waage P. Guldberg CM. Studier i affiniteten. *Forhandlinger: Videnskabs Selskabet i Christiania*, 35, 1864.

- Yang Guo, Christoph Riplinger, Ute Becker, Dimitrios G Liakos, Yury Minenkov, Luigi Cavallo, and Frank Neese. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [dlpno-ccsd (t)]. *The Journal of chemical physics*, 148(1):011101, 2018.
- Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.
- Tatsuhiko I Hasegawa and Eric Herbst. Three-phase chemical models of dense interstellar clouds-gas dust particle mantles and dust particle surfaces. *Monthly Notices of the Royal Astronomical Society*, 263:589, 1993.
- Tatsuhiko I Hasegawa, Eric Herbst, and Chun M. Leung. Models of gas-grain chemistry in dense interstellar clouds with complex organic molecules. *The Astrophysical Journal Supplement Series*, 82:167, sep 1992. ISSN 0067-0049.
- William M Haynes, David R Lide, and Thomas J Bruno. *CRC handbook of chemistry and physics*. CRC press, 2016.
- Jiao He and Gianfranco Vidali. Application of a diffusion-desorption rate equation model in astrochemistry. *Faraday Discussions*, 168:517–532, 2014.
- Jiao He, Kinsuk Acharyya, and Gianfranco Vidali. Binding energy of molecules on water ice: Laboratory measurements and modeling. *The Astrophysical Journal*, 825(2):89, 2016.
- Jiao He, SM Emtiaz, and Gianfranco Vidali. Measurements of diffusion of volatiles in amorphous solid water: application to interstellar medium environments. *The Astrophysical Journal*, 863(2):156, 2018.
- Thomas Henning and Dmitry Semenov. Chemistry in protoplanetary disks. *Chemical Reviews*, 113(12):9016–9042, 2013.
- E Herbst. Radiative association rate coefficients under shocked conditions in interstellar clouds-the case of  $\text{CH}_3^+ + \text{H}_2$ . *Astronomy and Astrophysics*, 153:151–156, 1985.
- E Herbst, TJ Millar, S Wlodek, and DK Bohme. The chemistry of silicon in dense interstellar clouds. *Astronomy and Astrophysics*, 222:205–210, 1989.
- Eric Herbst. The astrochemistry of  $\text{H}_3^+$ . *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 358(1774):2523–2534, 2000.
- Eric Herbst. *Gas Phase Reactions*, pages 561–574. Springer New York, New York, NY, 2006.
- Eric Herbst and William Klemperer. The formation and depletion of molecules in dense interstellar clouds. *The Astrophysical Journal*, 185:505–534, 1973.

- Eric Herbst and Ewine F Van Dishoeck. Complex organic interstellar molecules. *Annual Review of Astronomy and Astrophysics*, 47:427–480, 2009.
- Johannes Heyl, Serena Viti, Jonathan Holdship, and Stephen M Feeney. Exploiting network topology for accelerated bayesian inference of grain surface reaction networks. *The Astrophysical Journal*, 904(2):197, 2020.
- U Hincelin, B Commerçon, V Wakelam, F Hersant, S Guilloteau, and E Herbst. Chemical and physical characterization of collapsing low-mass prestellar dense cores. *The Astrophysical Journal*, 822(1):12, 2016.
- Kenzo Hiraoka, Nagayuki Ohashi, Yoshihide Kihara, Kazuyosi Yamamoto, Tetsuya Sato, and Akihiro Yamashita. Formation of formaldehyde and methanol from the reactions of h atoms with solid co at 10–20 k. *Chemical Physics Letters*, 229(4-5):408–414, 1994.
- David Hollenbach and EE Salpeter. Surface recombination of hydrogen molecules. *The Astrophysical Journal*, 163:155, 1971.
- Stéphane Humbel, Stefan Sieber, and Keiji Morokuma. The imomo method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and sn 2 reaction: Rcl+ cl-. *The Journal of chemical physics*, 105(5):1959–1967, 1996.
- P Jenniskens, DF Blake, MA Wilson, and A Pohorille. High-density amorphous ice, the frost on interstellar grains. *The Astrophysical Journal*, 455:389, 1995.
- Andreas Buchgraitz Jensen, Jakub Kubečka, Gunnar Schmitz, Ove Christiansen, and Jonas Elm. Massive assessment of the binding energies of atmospheric molecular clusters. *Journal of Chemical Theory and Computation*, 2022.
- Mihwa Jin and Robin T Garrod. Formation of complex organic molecules in cold interstellar environments through nondiffusive grain-surface and ice-mantle chemistry. *The Astrophysical Journal Supplement Series*, 249(2):26, 2020.
- HM Dipu Kabir, Abbas Khosravi, Mohammad Anwar Hosen, and Saeid Nahavandi. Neural network-based uncertainty quantification: A survey of methodologies and applications. *IEEE access*, 6:36218–36234, 2018.
- D A King. Thermal desorption from metal surfaces: A review. *Surface Science*, 47:384–402, 1975.
- Stephen J Klippenstein, Yu-Chuan Yang, Victor Ryzhov, and Robert C Dunbar. Theory and modeling of ion–molecule radiative association kinetics. *The Journal of chemical physics*, 104(12):4502–4516, 1996.
- F. Kruczkiewicz, J. Vitorino, E. Congiu, P. Theulé, and F. Dulieu. Ammonia snow lines and ammonium salts desorption. *Astronomy & Astrophysics*, 652:A29, aug 2021.
- G Landrum. Rdkit: Open–source cheminformatics software. *GitHub and SourceForge*, 10:3592822, 2016.



- Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Duřak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.
- Jacopo Lupi. *Computational strategies for the accurate thermochemistry and kinetics of gas-phase reactions*. PhD thesis, Scuola Normale Superiore, September 2022. URL <https://doi.org/10.5281/zenodo.7097381>.
- Antonios Makrymallis and Serena Viti. Understanding the formation and evolution of interstellar ices: a bayesian approach. *The Astrophysical Journal*, 794(1):45, 2014.
- Emilio Martínez-Núñez. An automated method to find transition states using chemical dynamics simulations. *Journal of computational chemistry*, 36(4):222–234, 2015.
- Emilio Martínez-Núñez, George L Barnes, David R Glowacki, Sabine Kopec, Daniel Peláez, Aurelio Rodríguez, Roberto Rodríguez-Fernández, Robin J Shannon, James JP Stewart, Pablo G Tahoces, et al. Automekin2021: An open-source program for automated reaction discovery. *Journal of computational chemistry*, 42(28):2036–2048, 2021.
- Matthew J McDermott, Shyam S Dwaraknath, and Kristin A Persson. A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. *Nature communications*, 12(1):1–12, 2021.
- Brett A McGuire. 2021 census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules. *The Astrophysical Journal Supplement Series*, 259(2):30, 2022.
- Andrew McKellar. Evidence for the molecular origin of some hitherto unidentified interstellar lines. *Publications of the Astronomical Society of the Pacific*, 52(307):187–192, 1940.
- Marco Minissale, Emanuele Congiu, and François Dulieu. Direct measurement of desorption and diffusion energies of o and n atoms physisorbed on amorphous surfaces. *Astronomy & Astrophysics*, 585:A146, 2016.
- Marco Minissale, Yuri Aikawa, Edwin Bergin, Mathieu Bertin, Wendy A Brown, Stephanie Cazaux, Steven B Charnley, Audrey Coutens, Herma M Cuppen, Victoria Guzman, et al. Thermal desorption of interstellar ices: A review on the controlling parameters and their implications from snowlines to chemical complexity. *ACS Earth and Space Chemistry*, 2022.
- GF Mitchell, JL Ginsburg, and PJ Kuntz. A steady-state calculation of molecule abundances in interstellar clouds. *The Astrophysical Journal Supplement Series*, 38:39–68, 1978.
- G Molpeceres, J Kaestner, VJ Herrero, RJ Peláez, and B Maté. Desorption of organic molecules from interstellar ices, combining experiments and computer simulations: Acetaldehyde as a case study. *Astronomy & Astrophysics*, 664:A169, 2022.

- German Molpeceres, Johannes Kastner, Gleb Fedoseev, Danna Qasim, Richard Schoomig, Harold Linnartz, and Thanja Lamberts. Carbon atom reactivity with amorphous solid water: H<sub>2</sub>O-catalyzed formation of h<sub>2</sub>co. *The journal of physical chemistry letters*, 12(44):10854–10860, 2021.
- Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15(9):095003, 2013.
- Keiji Morokuma, Qingfang Wang, and Thom Vreven. Performance evaluation of the three-layer oniom method: Case study for a zwitterionic peptide. *Journal of chemical theory and computation*, 2(5):1317–1324, 2006.
- Apurba Nandi, Chen Qu, Paul L Houston, Riccardo Conte, and Joel M Bowman.  $\delta$ -machine learning for potential energy surfaces: A pip approach to bring a dft-based pes to ccSD (t) level of theory. *The Journal of Chemical Physics*, 154(5):051102, 2021.
- AH Narten, C-G\_ Venkatesh, and SA Rice. Diffraction pattern and structure of amorphous solid water at 10 and 77 K. *The Journal of Chemical Physics*, 64(3):1106–1121, 1976.
- Frank Neese. Software update: the orca program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1327, 2018.
- Mark Newman. *Networks*. Oxford university press, 2018.
- Karin I Öberg and Edwin A Bergin. Astrochemistry and compositions of planetary systems. *Physics Reports*, 893:1–48, 2021.
- Stefano Pantaleone, Joan Enrique-Romero, Cecilia Ceccarelli, Stefano Ferrero, Nadia Balucani, Albert Rimola, and Piero Ugliengo. H<sub>2</sub> formation on interstellar grains and the fate of reaction energy. *The Astrophysical Journal*, 917(1):49, aug 2021.
- EM Penteado, C Walsh, and HM Cuppen. Sensitivity analysis of grain surface chemistry to binding energies of ice species. *The Astrophysical Journal*, 844(1):71, 2017.
- Jessica Perrero, Joan Enrique-Romero, Stefano Ferrero, Cecilia Ceccarelli, Linda Podio, Claudio Codella, Albert Rimola, and Piero Ugliengo. Binding energies of interstellar relevant s-bearing species on water ice mantles: A quantum mechanical investigation. *arXiv preprint arXiv:2209.07255*, 2022a.
- Jessica Perrero, Joan Enrique-Romero, Berta Martínez-Bachs, Cecilia Ceccarelli, Nadia Balucani, Piero Ugliengo, and Albert Rimola. Non-energetic formation of ethanol via cch reaction with interstellar h<sub>2</sub>o ices. a computational chemistry study. *ACS Earth and Space Chemistry*, 6(3):496–511, 2022b.
- Magnus Wilhelm Persson. Current view of protostellar evolution (ENG), 8 2014. URL [https://figshare.com/articles/figure/Current\\_view\\_of\\_protostellar\\_evolution/654555](https://figshare.com/articles/figure/Current_view_of_protostellar_evolution/654555).

- Elettra L. Piacentino and Karin I. Öberg. Computational estimation of the binding energies of pox and hpox ( $x = 2, 3$ ) species. *The Astrophysical Journal*, 939(2):93, nov 2022.
- Igor Poltavsky and Alexandre Tkatchenko. Machine learning force fields: Recent advances and remaining challenges. *The Journal of Physical Chemistry Letters*, 12(28):6551–6564, 2021.
- SS Prasad and WT Huntress Jr. A model for gas phase chemistry in interstellar clouds. i—the basic model, library of chemical reactions, and chemistry among c, n, and o compounds. *The Astrophysical Journal Supplement Series*, 43:1–35, 1980.
- Albert Rimola, Vianney Taquet, Piero Ugliengo, Nadia Balucani, and Cecilia Ceccarelli. Combined quantum chemical and modeling study of co hydrogenation on water ice. *Astronomy & Astrophysics*, 572:A70, 2014.
- Albert Rimola, Dimitrios Skouteris, Nadia Balucani, Cecilia Ceccarelli, Joan Enrique-Romero, Vianney Taquet, and Piero Ugliengo. Can formamide be formed on interstellar ice? an atomistic perspective. *ACS Earth and Space Chemistry*, 2(7):720–734, 2018.
- Christoph Riplinger, Peter Pinski, Ute Becker, Edward F Valeev, and Frank Neese. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. ii. linear scaling domain based pair natural orbital coupled cluster theory. *The Journal of chemical physics*, 144(2):024109, 2016.
- Alexander Ruf and Grégoire Danger. Network analysis reveals spatial clustering and annotation of complex chemical spaces: Application to astrochemistry. *Analytical Chemistry*, 2022.
- Marcin Rybicki and Joachim Sauer. Rigid body approximation for the anharmonic description of molecule–surface vibrations. *Journal of Chemical Theory and Computation*, 18(9):5618–5635, 2022.
- Bethmini Senevirathne, Stefan Andersson, Francois Dulieu, and Gunnar Nyman. Hydrogen atom mobility, kinetic isotope effects and tunneling on interstellar ices (ih and asw). *Molecular Astrophysics*, 6:59–69, 2017.
- Gregor N Simm, Alain C Vaucher, and Markus Reiher. Exploration of reaction pathways and chemical transformation networks. *The Journal of Physical Chemistry A*, 123(2):385–399, 2018.
- Dimitrios Skouteris, Nadia Balucani, Cecilia Ceccarelli, Noelia Faginas Lago, Claudio Codella, Stefano Falcinelli, and Marzio Rosi. Interstellar dimethyl ether gas-phase formation: a quantum chemistry and kinetics study. *Monthly Notices of the Royal Astronomical Society*, 482(3):3567–3575, January 2019.
- Ian WM Smith and Bertrand R Rowe. Reaction kinetics at very low temperatures: laboratory studies and interstellar chemistry. *Accounts of Chemical Research*, 33(5):261–268, 2000.

- Lynza H Sprowl, Charles T Campbell, and Liney Arnadottir. Hindered translator and hindered rotor models for adsorbates: Partition functions and entropies. *The Journal of Physical Chemistry C*, 120(18):9719–9731, 2016.
- T Stantcheva and E Herbst. Models of gas-grain chemistry in interstellar cloud cores with a stochastic approach to surface chemistry. *Astronomy & Astrophysics*, 423(1):241–251, 2004.
- Martin Stohr, Leonardo Medrano Sandonas, and Alexandre Tkatchenko. Accurate many-body repulsive potentials for density-functional tight binding from deep tensor neural networks. *The Journal of Physical Chemistry Letters*, 11(16):6835–6843, 2020.
- Steven Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, 1994.
- Timothy Su and Walter J Chesnavich. Parametrization of the ion–polar molecule collision rate constant by trajectory calculations. *The Journal of Chemical Physics*, 76(10):5183–5185, 1982.
- Mats Svensson, Stephane Humbel, Robert DJ Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. Oniom: a multilayered integrated mo+ mm method for geometry optimizations and single point energy predictions. a test for diels- alder reactions and pt (p (t-bu) 3) 2+ h2 oxidative addition. *The Journal of Physical Chemistry*, 100(50):19357–19363, 1996.
- Polydore Swings and Léon Rosenfeld. Considerations regarding interstellar molecules. *Astrophysical Journal*, 86:483–486, 1937.
- Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.
- Steven L Tait, Zdenek Dohnálek, Charles T Campbell, and Bruce D Kay. n-alkanes on mgo (100). ii. chain length dependence of kinetic desorption parameters for small n-alkanes. *The Journal of chemical physics*, 122(16):164708, 2005.
- Vianney Taquet, Cecilia Ceccarelli, and Claudine Kahane. Multilayer modeling of porous grain surface chemistry-i. the grainoble model. *Astronomy & Astrophysics*, 538:A42, 2012.
- Jessica Tennis, Jean-Christophe Loison, and Eric Herbst. Radiative association between neutral radicals in the interstellar medium:  $\text{CH}_3^+ + \text{CH}_3\text{O}$ . *The Astrophysical Journal*, 922(2):133, 2021.
- AGGM Tielens and W Hagen. Model calculations of the molecular composition of interstellar grain mantles. *Astronomy and Astrophysics*, 114:245–260, 1982.
- Alexander GGM Tielens. *The physics and chemistry of the interstellar medium*. Cambridge University Press, 2005.

- Lorenzo Tinacci, Stefano Pantaleone, Andrea Maranzana, Nadia Balucani, Cecilia Ceccarelli, and Piero Ugliengo. Structures and properties of known and postulated interstellar cations. *The Astrophysical Journal Supplement Series*, 256(2):35, 2021. doi: [10.3847/1538-4365/ac194c](https://doi.org/10.3847/1538-4365/ac194c).
- Lorenzo Tinacci, Aurele Germain, Stefano Pantaleone, Stefano Ferrero, Cecilia Ceccarelli, and Piero Ugliengo. Theoretical distribution of the ammonia binding energy at interstellar icy grains: A new computational framework. *ACS Earth and Space Chemistry*, 6(6):1514–1526, 2022. doi: [10.1021/acsearthspacechem.2c00040](https://doi.org/10.1021/acsearthspacechem.2c00040).
- Lorenzo Tinacci, Simon Ferrada-Chamorro, Cecilia Ceccarelli, Stefano Pantaleone, Daniela Ascenzi, Andrea Maranzana, Nadia Balucani, and Piero Ugliengo. The gretobape gas-phase reaction network: the importance of being exothermic. *The Astrophysical Journal Supplement Series*, Accepted for publication and currently under review.
- Lorenzo Tinacci, Aurele Germain, Stefano Pantaleone, Cecilia Ceccarelli, Nadia Balucani, and Piero Ugliengo. Theoretical water binding energy distribution and snowline in protoplanetary disks. *The Astrophysical Journal*, Submitted.
- Roberto Todeschini and Viviana Consonni. *Handbook of molecular descriptors*. John Wiley & Sons, 2008.
- Nenad Trinajstić. *Chemical graph theory*. Routledge, 2018.
- AG Tsvetkov and VI Shematovich. Kinetic monte carlo method for simulating astrochemical kinetics: Hydrogen chemistry in diffuse clouds. *Solar System Research*, 44(3):177–188, 2010.
- HC Van de Hulst. Evolution and physics of solid particles. *Harvard Observatory Monographs*, 7:73, 1948.
- Ewine F. van Dishoeck, Eric Herbst, and David A. Neufeld. Interstellar water chemistry: From laboratory to observations. *Chemical Reviews*, 113(12):9043–9085, 2013.
- Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.
- Nick M Vandewiele, Kevin M Van Geem, Marie-Françoise Reyniers, and Guy B Marin. Genesys: Kinetic model construction using chemo-informatics. *Chemical Engineering Journal*, 207:526–538, 2012.
- AI Vasyunin and Eric Herbst. Reactive desorption and radiative association as possible drivers of complex molecule formation in the cold interstellar medium. *The Astrophysical Journal*, 769(1):34, 2013.
- AI Vasyunin, AM Sobolev, DS Wiebe, and Dmitry A Semenov. Influence of uncertainties in the rate constants of chemical reactions on astrochemical modeling results. *Astronomy Letters*, 30(8):566–576, 2004.
- Gianfranco Vidali. H<sub>2</sub> formation on interstellar grains. *Chemical reviews*, 113(12):8762–8782, 2013.

- Torben Villadsen, Niels FW Ligterink, and Mie Andersen. Predicting binding energies of astrochemically relevant molecules via machine learning. *Astronomy & Astrophysics*, 666:A45, 2022.
- Serena Viti, Mark P Collings, John W Dever, Martin RS McCoustra, and David A Williams. Evaporation of ices near massive stars: models based on laboratory temperature programmed desorption data. *Monthly Notices of the Royal Astronomical Society*, 354(4):1141–1145, 2004.
- Craig Waitt, Audrey R Miles, and William F Schneider. Adsorbate free energies from dft-derived translational energy landscapes. *The Journal of Physical Chemistry C*, 125(37):20331–20342, 2021.
- V. Wakelam, J. C. Loison, R. Mereau, and M. Ruaud. Binding energies: New values and impact on the efficiency of chemical desorption. *Molecular Astrophysics*, 6:22–35, 2017.
- Valentine Wakelam, Eric Herbst, and Franck Selsis. The effect of uncertainties on chemical models of dark clouds. *Astronomy & Astrophysics*, 451(2):551–562, 2006.
- Valentine Wakelam, E Herbst, J Le Bourlot, F Hersant, Franck Selsis, and S Guilloteau. Sensitivity analyses of dense cloud chemical models. *Astronomy & Astrophysics*, 517:A21, 2010a.
- Valentine Wakelam, IWM Smith, E Herbst, Juergen Troe, W Geppert, H Linnartz, K Öberg, E Roueff, M Agúndez, P Pernot, et al. Reaction networks for interstellar chemical modelling: improvements and challenges. *Space science reviews*, 156(1):13–72, 2010b.
- Valentine Wakelam, J-C Loison, E Herbst, B Pavone, A Bergeat, K Béroff, M Chabot, A Faure, Daniele Galli, Wolf D Geppert, et al. The 2014 kida network for interstellar chemistry. *The Astrophysical Journal Supplement Series*, 217(2):20, 2015.
- Naoki Watanabe and Akira Kouchi. Efficient formation of formaldehyde and methanol by the addition of hydrogen atoms to co in h<sub>2</sub>o-co ice at 10 k. *The Astrophysical Journal*, 571(2):L173, 2002.
- David E Woon and Eric Herbst. Quantum chemical predictions of the properties of known and postulated neutral interstellar molecules. *The Astrophysical Journal Supplement Series*, 185(2):273, 2009.
- Satoshi Yamamoto. Introduction to astrochemistry. *Editorial: Springer*, 2017.



Grazie a tutta la famiglia astrochimica DOC/ACO che hanno reso i giorni di questo progetto bellissimi e leggeri.

Un ringraziamento speciale va però a Piero e Cecilia che mi hanno insegnato più di quello che sarebbe richiesto dall'università. Sono esempi di dedizione all'insegnamento, al lavoro ed alla scienza. Ma soprattutto esempi di professori la cui conoscenza non si limita solo a quella accademica.

*"L'Io solitario che gira intorno a se stesso e si nutre solo di sé finisce strozzato da un gran pianto o da un gran riso."*(Fellini, 8<sup>1/2</sup>)