THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Hydrogen Cyanide at the Onset of Prebiotic Chemical Reactivity

MARCO CAPPELLETTI

Department of Chemistry and Chemical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2026

Hydrogen Cyanide at the Onset of Prebiotic Chemical Reactivity

MARCO CAPPELLETTI

ISBN 978-91-8103-365-6

Doktorsavhandlingar vid Chalmers tekniska högskola,
Ny serie nr 5822
ISSN 0346-718X
https://doi.org/10.63959/chalmers.dt/5822

Department of Chemistry and Chemical Engineering
Chalmers University of Technology
SE-412 96 Göteborg,
Sweden
Phone: +46(0)31 772 1000

Cover: A lost HCN molecule consulting a map at a sterile crossroads, representing the onset of prebiotic chemistry.

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2026.

Hydrogen Cyanide at the Onset of Prebiotic Chemical Reactivity

MARCO CAPPELLETTI

Department of Chemistry and Chemical Engineering
Chalmers University of Technology

Abstract

Hydrogen cyanide (HCN) is a unique molecule. It is ubiquitous in the universe, highly reactive, and capable of forming a diverse range of life's building blocks through self-reaction. Its solid phase is equally intriguing, as it has been observed to glow and exhibit electric discharges, likely driven by its strong polarity. While investigating these exceptional properties is crucial for astrochemistry and the origin of life, HCN's high toxicity and tendency to form intractable polymers have left its rich chemistry relatively unexplored. In this thesis, I employ quantum chemical methods to unravel the reactivity and solid-state properties of HCN.

The first part of this thesis focuses on the reactivity of HCN. An exploration of the thermodynamic landscape of its self-reaction reveals that while most products are thermodynamically favorable, several proposed polymerization pathways are endergonic. Among the most favored products are highly conjugated polymers and the nucleobase adenine. The kinetics of adenine formation is further investigated by comparing proposed base-catalyzed pathways in an HCN-rich environment. These results provide a kinetic explanation for the low adenine yields typically observed in experiments.

The second part concerns the HCN crystal, with a particular focus on Titan, where HCN ice is abundant. In this part of the thesis, crystal morphology and polar surface properties are investigated. The results suggest a high anisotropy of the crystal, making it grows into needle-shaped structures, whose tips can exert strong electric fields. These electric fields arise from the uncompensated electrostatic potential at the polar surfaces and are suggested to facilitate chemical transformations. Furthermore, the electronic structure of HCN polar surfaces shows the emergence of localized metallic surface states. This metallicity is transferred to chemisorbed water molecules, suggesting enhanced reactivity at the interface. This thesis provides insights into the exceptional properties of HCN in the context of prebiotic chemistry and astrochemistry.

Keywords

hydrogen cyanide, diaminomaleonitrile, adenine, HCN-derived polymers, origin of life, Titan, astrochemistry, Computational chemistry.

List of Appended Papers

This thesis is based on the following papers:

I **A Thermodynamic Landscape of Hydrogen Cyanide-Derived Molecules and Polymers**
Hilda Sandström, Fernando Izquierdo-Ruiz, Marco Cappelletti, Rana Dogan, Siddhant Sharma, Clara Bailey, and Martin Rahm
ACS Earth and Space Chemistry 2024, 8 (6) 1272–1280.

II **How Does Adenine Form from Hydrogen Cyanide?**
Marco Cappelletti and Martin Rahm
Accepted, *Journal of the American Chemical Society*

III **Electric Fields Can Assist Prebiotic Reactivity on Hydrogen Cyanide Surfaces**
Marco Cappelletti, Hilda Sandström, and Martin Rahm
Accepted, *ACS Central Science.*

IV **Hydrogen Cyanide Surfaces Can Turn Water Metallic**
Marco Cappelletti and Martin Rahm,
Manuscript.

Contribution report

This author's contributions to the appended papers are summarized below:

I Performed quantum chemical calculations together with H.S., F.I-R., R.D., and S.S. Wrote the manuscript together with H.S., F.I-R., and M.R.

II Performed quantum chemical calculations. Designed the study and wrote the manuscript together with M.R.

III Performed quantum chemical calculations together with H.S. Designed the study and wrote the manuscript together with H.S. and M.R.

IV Performed quantum chemical calculations. Designed the study and wrote the manuscript together with M.R.

List of Abbreviations

AICN 4-amino-1H-imidazole-5-carbonitrile

AIMD *ab initio* molecular dynamics (MD)

AISN aminoiminosuccinonitrile

AMN aminomalononitrile

B3LYP Becke, 3-parameter, Lee-Yang-Parr

BO Born-Oppenheimer

CC coupled-cluster

CCSD CC including single and double excitations

CCSD(T) CCSD including perturbative triples

DAMN diaminomaleonitrile

DFT density functional theory

DLPNO domain-based local pair natural orbital

GGA generalized gradient approximation

HB hydrogen bond

HF Hartree-Fock

IAN iminoacetonitrile

MD molecular dynamics

MEP minimum energy path

MO molecular orbital

PBE Perdew-Burke-Ernzerhof

PCM Polarizable Continuum Model

PES potential energy surface

RNA ribonucleic acid

RT room temperature

SE Schrödinger equation

TS transition state

UV ultraviolet

VS Voet and Schwartz

"The energy from ultraviolet light probably would be greater than that from electrical discharges. Because of the difficulties of working with ultraviolet light in the region in which these reduced gases would absorb, I decided to work with electrical discharges."

– Stanley L. Miller, 1957

Acknowledgment

This work presented in this thesis was made possible by financial support from the Swedish Research Council (grant 2020-04305 and 2024-05049). The research relied on computational resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at the Chalmers Centre for Computation and Engineering (C3SE) and National Supercomputer Centre (NSC), funded by the Swedish Research Council through grant agreement No. 2022-06725.

It has been a four-and-a-half-year journey that changed me in many ways. Very few experiences in life let you meet outstanding people, share ideas and values, travel around the world, and receive a salary for being a nerd. The price to pay, however, is scattered gray hair and the realization of how little we actually know and can contribute to the world. Nonetheless, the opportunity to be part of science has been invaluable.

I wish to thank my supervisor, Martin Rahm, for his guidance and pedagogical support, and for believing in me. To all my colleagues, who were there for every high and every low. Special thanks to my past and present group members, and almost-so: Hilda, Mårten, Ruslan, Jona, Eve, Erika, Rui, Hanna, Chris, Mark, Ely. Thank you for the breaks, the laughs, the vents, but especially for all the crucial support that pushed me up to here. Hopefully I helped you on your journeys too.

To my friends and all the people who in any way helped me get to where I am today. As I'm writing this few hours before printing, I do not have time to list you all. Honorable mentions: the Polpettorcia (enhanced) group, my deepest thanks for being such incredible friends, and for making me experience the Italian chaos from time to time; the Irriducibili, Fede and Abe, for the best vacations of my life; Kernel tmtm, Conte, you contributed to my path into science, but soon we will launch our startup; Mattia, thank you for inspiring me being a better person, see you in Minecraft; Paolo, without you I would *literally* not be here, but for many reasons; Ana, you gave me the tools to walk this journey, you deserve everything; and many thanks my team LaBamba, who made me survive this last year.

To my family: Luca, my twin sister Maria Giulia, mom and dad, who truly shaped me into who I am now. Grazie di tutto.

Marco Cappelletti, Göteborg, January 2026

Contents

CHAPTER 1

Introduction

Physics shapes our Universe, but chemistry brings it to life. The most sublime achievement accomplished by chemistry is, indeed, creating life. While the transition from inanimate matter to self-replicating organisms remains an enigma, the scientific consensus posits that this profound metamorphosis entailed a cascade of naturally-occurring chemical transformations. Such a path was likely tortuous and obscured by the disorder of the primordial chemical chaos. The overwhelming task to trace this road pertains to the domain of **prebiotic chemistry**. The breadth of this inherently speculative yet fascinating field spans from the early formation of life's "building blocks", to the assembly of the first functional polymers [1]. In this thesis, I address crucial questions in prebiotic chemistry by means of **computational chemistry** methods.

Throughout almost a century, a plenitude of different feedstocks, processes, and scenarios have been proposed to have contributed to the emergence of life, and may contribute to it elsewhere. Among the central precursors, **Hydrogen Cyanide** (HCN), also known as prussic acid, occupies a unique position. This molecule is both abundant and ubiquitous in the universe, full of chemical energy, and, crucially, it is able to originate an ample diversity of life's building blocks. These properties make HCN a widely recognized prebiotic precursor, both in the early Earth and elsewhere.

At the bottom of HCN's significance in prebiotic chemistry lies its reactivity: HCN can undergo violent reactions, forming a complex mixture of polymers and molecules called **HCN-derived products**. When hydrolyzed, these products have been shown to release prebiotically-relevant molecules, such as amino acids, purines, and pyrimidines. However, because of the extraordinary complexity of HCN self-reaction chemistry, their composition remains largely unresolved. This prompts the question:

Q1. Which molecules and polymers can form spontaneously from HCN?

The heterogeneous material that HCN forms is largely intractable. Decades of experimental effort to characterize its composition have yielded only a glimpse into its complexity, leaving most structural details hidden within the tar. Computational chemistry represents an essential tool that allows to tackle this problem from a different angle. **Paper I** takes a first step in this direction by elucidating the thermodynamic landscape of HCN-derived products.

One of the most interesting HCN-derived molecules is, arguably, adenine **adenine**. Adenine is omnipresent in all modern life forms: it is a component of DNA, RNA, and ATP, and either part of or the precursor to plenty of different coenzymes and other biologically active molecules [2]. The discovery of adenine formation from HCN represents a milestone in the research on the origin of life [3]. However, despite the interest, the mechanistic steps that connect such a toxic chemical with such a pivotal molecule for life remain elusive. Clarifying this uncertainty is the focus of **Paper II**, which investigates the question:

Q2. Through which reaction mechanism(s) does adenine form from HCN?

The exceptional nature of HCN is not limited to its chemical reactivity. This molecule presents a sizable dipole moment [4], which contributes to its solid-state properties. For example, electrical breakdowns and glow have been observed during HCN crystallization experiments [5], clear evidence of strong electrostatic forces at play. HCN plausibly exists in the solid state in many places where temperatures are low enough. The presence HCN ice has been established in the atmosphere of Saturn's moon Titan, a fascinating world that is rendered inhospitable by cryogenic temperatures, below −73 °C. Hence, an aspiring research question is evoked:

Q3. Can HCN crystal surfaces promote prebiotic chemistry in cold environments?

Surfaces are often invoked in prebiotic chemistry, as they often allow otherwise inaccessible chemistry. While intriguing, the physicochemical properties of solid-state HCN remain underexplored. Acknowledging this knowledge gap, **Papers III–IV** focus on the HCN crystal, particularly on the potential catalytic properties of its polar surfaces.

CHAPTER 2

Background

"There must have been a time when inorganic elements alone existed on our planet: [...] Now is there a fact, or a shadow of a fact, supporting the belief that these elements, without the presence of any organic compounds, and acted on only by known forces, could produce a living creature? At present it is to us a result absolutely inconceivable."

– Charles Darwin, 1863

The research presented in this thesis builds upon decades of investigation into the origins of life. In this chapter, I provide the necessary background to place my work in context.

2.1 The Puzzle of Abiogenesis

Although the concept is debated, life is most commonly defined as "a self-sustaining chemical system capable of Darwinian evolution" [6]. The general consensus holds that the emergence of life involves a gradual transition from inanimate matter. This process, also known as **abiogenesis**, is among the most profound questions of humankind and one of the greatest unresolved mysteries in science. The field of the origin of life is inherently broad and cross-disciplinary. Studying such an immense puzzle requires the interplay of physics, biology, geology, astronomy, and chemistry. Although the ultimate goal is to understand how life can generally arise, we are limited to a single example: "life as we know it". Therefore, most research necessarily focuses on how life emerged here on Earth.

2.1.1 Early Earth geological setting

Geological records suggest that life was already present on Earth as far as 3.7 billion years ago (Gya) [7, 8]. Although any direct evidence from the first living organisms is lost forever, these records can provide an upper limit for how long it took for life to emerge. Recent studies are pushing this limit even further in the past, up to 4.1 Gya, during the Hadean eon [9, 10]. Nevertheless, the general consensus is that life emerged rather rapidly: only $\sim$ 200-400 million years after liquid water is believed to have formed on the surface ($\sim$4.4-4.2 Gya) [11, 12].

Little is known about the environmental conditions of the Hadean Earth.[1] Early hypotheses, based on astrochemical considerations, proposed that the atmosphere was strongly reducing and rich in CH_4 and H_2 [13]. However, while the precise redox state remains debated [14], recent evidence points toward a neutral atmosphere dominated by CO_2 and N_2 [15]. The temperature at the surface of the early Earth is also largely unknown, although there is evidence of liquid water as far back as 4.4 Gya [9, 12, 16]. The surface was likely irradiated by intense ultraviolet (UV) radiation [17], due to the young Sun and the absence of an ozone layer [18].

2.1.2 The blurred origin of life on Earth

To study how abiogenesis occurred on Earth, two main strategies are generally adopted: "top-down" and "bottom-up". Top-down approaches use common features in modern biology to trace back how our Last Universal Common Ancestor (LUCA) might have looked like. Research suggests that LUCA was a thermophilic, anaerobic, CO_2-fixing organism living in a hydrothermal setting around $\sim$ 4.2 Gya [19, 20]. Notably, these studies indicate that LUCA was a highly developed form of life – with a complexity comparable to modern prokaryotes – and was part of an established ecosystem [20]. This evidence, while valuable, implies that analyzing LUCA may provide limited insight on the actual first spark of life. Therefore, as top-down approaches inherently struggle to look far beyond this point, the mystery of abiogenesis remains.

On the other hand, bottom-up approaches start from first-principle chemistry, integrating knowledge from physics, geology, and astronomy. This is the domain of **prebiotic chemistry**. Its scope spans from the synthesis of the first simple precursors to the abiotic assembly of complex polymers, such as ribonucleic acid (RNA). Of course, as we know little about the exact conditions of the early Earth, prebiotic chemistry also relies on a significant number of assumptions. Furthermore, bottom-up approaches are afflicted by multiple simultaneous chicken-and-egg problems. Modern life depends on interdependent systems: genes storing instructions, proteins reading and executing

[1] Earth changed drastically during the Hadean eon ($\sim$4.6 Gya up to 4.03 Gya), transitioning from a magma world to an ocean world. Here, I refer as "Hadean" or "early Earth" to, technically, mid- to late-Hadean, whose conditions favored the emergence of life.

them, metabolic cycles harnessing chemical energy to power these mechanisms, and membranes providing an enclosed environment that allows all the above to happen. Since none of them can function without the others, which came first?

2.2 Prebiotic Chemistry

Until the early 20th century, the origin of life was largely a subject of philosophical debate rather than testable scientific hypotheses. The first breakthrough came with Oparin (1924) [21] and Haldane (1929) [22], who independently proposed the **primordial soup** hypothesis: a scenario where organic molecules in the early oceans polymerized and self-assembled over vast timescales to produce the first life form. These ideas culminated in the famous Miller-Urey experiment in 1953 [23]. In this experiment, Stanley Miller and Harold Urey demonstrated that simple reducing gases subjected to electric discharges yield amino acids. Although the assumptions behind the experiment are now considered outdated, the experiment triggered a paradigm shift and launched the field of prebiotic chemistry.

These early speculations also carried the idea that the emergence of life obeyed the laws of chemistry. Achieving it appears nearly impossible: without the machinery of life, the synthesis of large molecules (such as RNA and proteins) is strongly endergonic in aqueous solutions, as hydrolysis is more favored (the so-called "water paradox") [24, 25]. An external input of energy is thus essential. This energy could be provided by UV light, or by gradients in thermal energy, concentration, or pH. Additionally, solid surfaces are often invoked to facilitate these reactions by locally increasing concentration and acting as catalysts [26–29].

2.2.1 Prebiotic scenarios

On the early Earth

The primordial soup hypothesis laid the ground for modern hypotheses that accommodated new scientific discoveries. Specifically, current research focuses on plausible[2] environments where reactants can concentrate and some form of energy can be harnessed. One prominent example is **warm little ponds**.

Originally described by Charles Darwin, a warm little pond is an hypothetical shallow lake on the early Earth [31]. The main source of disequilibrium in these environments is the wet-dry cycle, which can promote condensation reactions [29]. UV light provides an additional energy source; while potentially disruptive, it could also serve as a selection pressure, favoring the survival of UV-stable building blocks [32]. These scenarios generally fall under the umbrella of "genetics-first" theories, which suggest that the emergence of

[2]Which meaning is of course loose [30].

life was driven by self-replicating polymers such as RNA. The "RNA World" hypothesis [33] is currently a widely accepted model in the field [34–37].

An alternative idea suggests that life began in **hydrothermal vent** systems on the ocean floor [38]. In contrast to "genetic-first" theories, these scenarios often imply "metabolism-first", and they focus on self-sustaining metabolic pathways. The appeal of hydrothermal vents stems from the striking resemblance between chemical reactions occurring at those sites and modern life's metabolic routes [38, 39]. Furthermore, the thermal, concentration, and pH gradients maintained between the ocean and the vents have been argued to be more "life-like" disequilibria (sources of energy), as opposed to "nonspecific" energy inputs as UV light or wet-dry cycles [40, 41]. Mineral surfaces can also create a local environment with reduced water activity, which has been proposed to solve the hydrolysis problem [24]. However, the typically high temperatures are considered incompatible with the fragile nature of RNA, which is why the RNA-world hypothesis seldom focuses on these scenarios [42, 43]. The debate[3] continues.

Elsewhere

Other hypotheses look beyond Earth, which is the domain of the emerging field of **astrobiology**. These theories focus on how life might emerge elsewhere in the Universe, in environments that are drastically different from the early Earth, exploiting diverse energy sources and chemical species. Current research concerns Mars,[4], icy moons like Jupiter's Europa and Saturn's Enceladus, and also the atmosphere of Venus [46]. However, Saturn's largest moon, Titan, stands out as a unique world.

Titan is considered a natural prebiotic laboratory [47] that simultaneously could not be further from Earth. The average surface temperature is 90 K, and the atmosphere goes as "high" as 200 K [48]. Its characteristic thick, yellow atmosphere is mainly composed of nitrogen (N_2) and methane (CH_4), and is filled with nitrogen-bearing organic molecules [49, 50]. Finally, scattered on the surface are lakes made of liquid methane and ethane, which render Titan such a familiar yet alien world. This anoxic environment has inspired numerous alternative prebiotic chemistry scenarios [51–54]. The upcoming Dragonfly mission (2028) will put these theories to the test and hopefully provide new data that will stimulate further research.

[3] "Debate" is perhaps an understatement, as it is not uncommon to encounter heated discussions. For example, John Sutherland stated that "life originated at vents should, like the vents themselves, remain 'In the deep bosom of the ocean buried'" [44]. In response, Elbert Branscomb and Micheal J. Russell called "Frankenstein models" and "energized soups" the "energized assemblages of building blocks" proposed in genetics-first theories, and reassured that "this discouraging view of Frankenstein theories need not unsettle us too much, because the 'bosom of the deep' is at hand to save the day" [40, 41].

[4] Where we might have found traces of life! [45]

2.2.2 The precursors

The general assumption in genetics-first theories is that modern life's building blocks must have been formed abiotically – somehow and somewhere – on the early Earth [1]. Understanding the synthesis and assembly of these species in a plausible prebiotic scenario is therefore fundamental. This challenge represents the primary concern of current prebiotic chemistry research. Luckily, we are not fumbling in the dark. As a matter of fact, quite the opposite: as Leslie Orgel wrote, the "crime" of the origin of life on Earth has "no shortage of clues", "on the contrary, there are far too many clues and far too many suspects" [13].

To be considered a plausible prebiotic precursor, a species should generally satisfy three criteria: (i) be common or readily available in prebiotic scenarios, (ii) possess high chemical energy, and (iii) promote the formation of biologically relevant species. These precursors could have been synthesized in situ (endogenous production), or delivered from space (exogenous delivery) [55]. The discovery of amino acids, sugars, and nucleobases on comets and carbonaceous chondrites strongly supports the view that exogenous delivery contributed to the availability of organics in the early Earth [56–60].

A multitude of chemical species has been proposed over the last century (see, e.g., ref.s [57, 61]). Typically, specific routes are investigated for each family of biomolecules (e.g., sugars, amino acids, purines). For instance, formaldehyde (H_2CO) is a likely precursor to sugars, whereas cyanoacetylene (HC_3N) is often invoked for the synthesis of pyrimidines. Among these candidates, HCN occupies a central role.

CHAPTER 3

Hydrogen Cyanide in Prebiotic Chemistry

"To assume that life arose from inorganic matter presents overwhelming difficulties."
– Stanley Miller, 1957

Hydrogen cyanide (HCN) is a prominent molecule in prebiotic chemistry. In this chapter, I describe its physicochemical properties and the fundamental reasons for its central role in the origin of life.

3.1 The HCN Molecule

HCN is the simplest nitrile (Figure 3.1a). Its properties stem from the carbon–nitrogen triple bond, which dictates a linear geometry and makes its chemistry largely governed by π-symmetry frontier molecular orbitals (MOs) (Figure 3.1b). This molecule is a weak acid ($pK_a = 9.2$ in water) and its deprotonation yields the cyanide anion (CN^-), a strong nucleophile. HCN possesses a sizable dipole moment, 2.99 D in the gas-phase [62], making it an excellent hydrogen bond (HB) donor and acceptor. The gas-phase dissociation energy of the dimer is $\sim$3.8 kcal mol^{-1}, slightly larger than that of the water dimer [63, 64]. Furthermore, strong cooperative effects arise when the molecules align in linear chains [65–70]. These interactions are primary electrostatic in nature [67].

At ambient pressure, HCN condenses at 299 K (26 °C) into a volatile, low-density liquid (704 g cm^{-3} at 5 °C). The high polarity of the molecule governs much of its physical behavior in the condensed phase. Liquid HCN displays an outstanding dielectric constant of 144.8 at 5 °C, markedly larger than that of water at the same temperature (85.8) [71, 72]. Due to the linear geometry of the molecule, the liquid phase is composed of small and transient hydrogen-

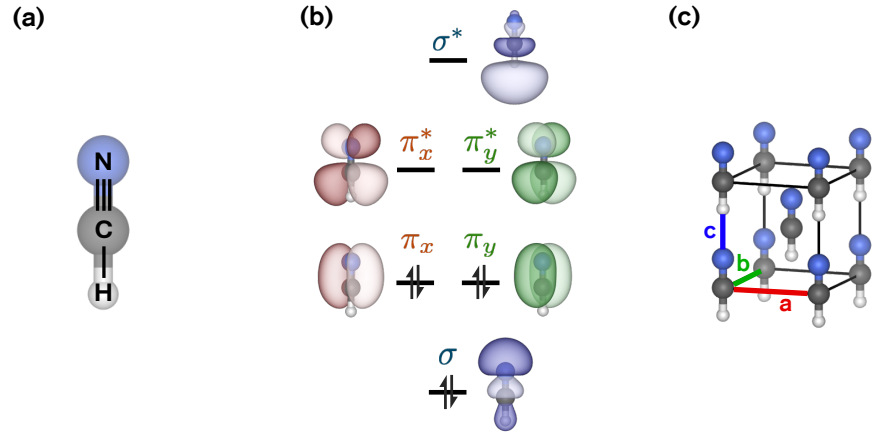

Figure 3.1: (a) Lewis structure of HCN overlaid on the ball-and-stick representation that will be used throughout this thesis, where hydrogen, carbon and nitrogen atoms are depicted in white, dark gray, and blue, respectively. (b) Molecular orbitals (MOs) of HCN, ordered by energy (not to scale). (c) Orthorhombic *Imm2* phase of the HCN crystal.

bonded linear chains [73].

HCN solidifies below 260 K (−13 °C), forming needle-like crystals that are connected by each other in a "cobweb" framework (Figure 3.2). Two crystalline phases exist [74–76]: a high-temperature *I4mm* tetragonal phase and a low-temperature *Imm2* orthorhombic phase, separated by a ferroelastic transition occurring at 170 K (−103 °C) [77, 78]. Both crystal structures are polar and composed of hydrogen-bonded linear chains aligned parallel to the polar axis. This topology gives rise to substantial cooperative effects and a macroscopic electric field that is exerted at the polar facets. Remarkably, peculiar electrostatic phenomena, such as visible glow, strong pyroelectricity, and electrical breakdowns following crystal fractures, have been observed during low-temperature HCN crystallization experiments. [5]. However, mainly because of the high toxicity of HCN, experimental data on its solid-state properties remains scarce.

The structural isomer of HCN is **hydrogen isocyanide** (HNC), which lies considerably higher in energy ($\sim$15 kcal mol^{-1}) above HCN. Their interconversion in the gas-phase is prohibitive, as it entails a barrier of $\sim$48 kcal mol^{-1} [79, 80].

3.2 HCN in the Universe

HCN is ubiquitous in our Universe. It has been detected in the gas-phase in a diverse range of environments, such as the interstellar medium (ISM) [81], in star-forming regions [82], on carbonaceous chondrites [83], in the coma

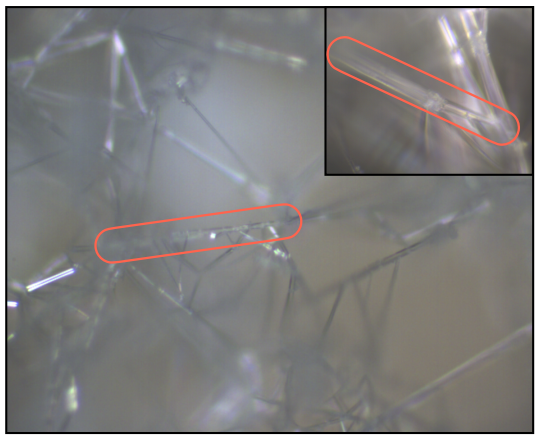

Figure 3.2: Microscope images of the needle-like structures of solid HCN. Figure reproduced from **Paper III**, licensed under CC BY 4.0.

of comets [84–86], and in the atmosphere of several planets [87, 88], dwarf planets [89], and moons [90–92]. This simple nitrile is also abundant. For example, HCN is found in comets with abundances $\sim$0.1–0.2% with respect to water [93]. In the solid-state, HCN has been tentatively detected on Neptune's moon Triton [94], and is known to be abundant in the atmosphere of Saturn's moon Titan, where clouds of pure HCN ice are observed [95]. Conversely, the HCN crystal is yet to identified in comets, possibly due to chemical reactions with other species that destroy the solid [96]. HCN-water ice mixtures, and possibly pockets of pure HCN ice, may be present in protoplanetary disks [97].

The HCN isomer, HNC, is also ubiquitous in the universe and is often found with abundances that are comparable to that of HCN. In particular, typical HNC/HCN ratios range within 0.01-1.0 [98–102], in striking contrast with expectations based on mere thermodynamics. In comets, the ratio increases with decreasing heliocentric distance [100, 103–105], which indicates an in-situ production [106, 107]. HNC may derive from the isomerization of HCN by H atoms, the photodissociation of a parent molecule, or various chemical reactions [85, 107, 108].

3.2.1 HCN in Titan's atmosphere

Titan's atmospheric chemistry is driven by solar UV photons and energetic particles (e.g., from Saturn's magnetosphere and galactic cosmic rays). While ground-state chemical reactions are expected to be slow due to the low temperatures (below 200 K [48]), Titan hosts a rich and complex environment that is far from being fully understood [50, 109]. Titan's most characteristic feature is the pervasive yellow haze, composed by **tholins**, a complex mixture of nitrogen-rich organic solids [110]. The composition of tholins resembles that

of the HCN polymerization products, although the former exhibit a higher degree of complexity [111, 112].

HCN is one of the primary reaction products of Titan's atmospheric photochemistry [113]. It forms in the upper atmosphere and freezes below ∼300 km of altitude, forming clouds [95, 114, 115]. Moreover, solid HCN and its reaction products are believed to deposit at the surface [49, 116], where it may form co-crystals with hydrocarbons [117, 118]. Notably, nonequilibrium amounts of HNC are observed also in Titan's atmosphere: the observed HNC/HCN goes as high as 0.3 in the upper atmosphere [102]. Chemical models suggest that most HNC forms in the ionosphere, mainly through $HCNH^+$[49, 119, 120]. The HNC abundance anomaly on Titan remains largely unexplained; ion chemistry, photochemistry, and cosmic ray-induced chemistry represent likely explanations.

3.3 HCN in Prebiotic Chemistry

The central role of HCN in prebiotic chemistry has a long history. As early as 1854, this molecule was known to react with ammonia and acetaldehyde to yield amino acids, a reaction now known as the Strecker synthesis [121]. A century later, in 1957, Stanley Miller demonstrated that HCN was indeed a key intermediate in the formation of amino acids from reducing gases subjected to electric discharges [122]. However, what really sparked the research on HCN in the origin of life field was Juan Oró's groundbreaking detection of **adenine** from the hydrolyzed HCN polymerization products, in 1960 [3]. This discovery triggered a multitude of experiments on HCN polymerization in the subsequent decades, which reported the synthesis of other essential building blocks of life, such as several amino acids and other nucleobases.

The plausibility of HCN as a precursor to biomolecules on the Hadean Earth depends on its concentration. Endogenous production could be achieved from photochemistry [123], volcanic activity [124] and impact events [125]. The composition of the primordial atmosphere – likely rich in CO_2 and N_2 – makes HCN synthesis via electric discharges unfeasible, contrary to previous beliefs. Nevertheless, impacts might have created a transient reducing environment capable of forming HCN [126]. Cometary impacts might have been a source of HCN themselves [55], albeit the survival rate might have been very low [127]. The often-cited wet-dry cycles in warm little ponds would be ineffective, as HCN cannot concentrate through evaporation due to its high volatility. In essence, recent estimates suggest that HCN production in the early Earth was likely insufficient for it to accumulate in the oceans and polymerize [128, 129]. Consequently, several concentration mechanisms have been proposed. Two frequently discussed hypothesis are concentration by eutectic freezing [130, 131] and HCN formation from formamide, its (less volatile) hydrolysis product [132–134].

3.4 HCN Self-Reaction Chemistry

The polymerization of HCN has been first observed by Joseph L. Proust in 1806 [135]. In this process, HCN readily undergoes violent self-reaction and polymerization, forming a heterogeneous family of yellow-to-black solids. Although they are commonly referred to as "HCN-derived polymers", they are more correctly described as macromolecular organic solids. Nevertheless, I will adopt the standard nomenclature and refer to them as polymers.

3.4.1 Reaction conditions

HCN polymerization occurs under a broad range of conditions (see Ref.s [136, 137] for detailed reviews). The reaction is typically initiated by the formation of the cyanide anion (CN^-) [138, 139] or the cyano radical ($CN^\bullet$) [140]. These species can be generated in various ways: by thermal activation, by the presence of a base (e.g., NH_3) or minerals, or by ionizing radiation [3, 141–144]. Moreover, this polymerization can occur in a wide range of temperatures, and in both gas and liquid-phase, starting from either pure HCN or various solutions [145–147]. Polymerization is generally not observed in the solid state in the dark. However, irradiation can generate reactive species and polymerization may proceed upon melting [142].

Experiments using pure liquid HCN typically involve the addition of 1-10% of ammonia [138, 145, 148]. In aqueous solutions, ammonia or other bases are used to adjust the pH to 9.2 – the pK_a of HCN – where the polymerization is fastest [149]. The hydrolysis of HCN to formamide competes with polymer formation and becomes faster at HCN concentrations below 0.01 M [129]. Eutectic freezing of water-HCN mixtures, whose eutectic temperature is −21 °C, can be used to concentrate HCN and induce polymerization [130, 131, 150]. As a matter of fact, polymerization from aqueous NH_4CN has been observed at temperatures as low as −78 °C [151]. Since the eutectic temperature of NH_3-H_2O is −100 °C, part of the sample likely remained unfrozen and contained highly concentrated NH_4CN.

The synthesis of HCN polymers can be also achieved from precursors other than HCN itself, including the HCN trimer and tetramer [152–155], as well as derivatives such as formamide [156, 157]. However, the discussion of these alternative synthetic routes lies outside the scope of this thesis.

3.4.2 Mechanisms and intermediates

How the HCN self-reaction proceeds is largely unknown. In this section, I discuss base-catalyzed (non-radical) reaction mechanisms from HCN, which will be the focus of this thesis. The first step is the nucleophilic attack by CN^- to the nitrile group ($-C\equiv N$) of HCN, which forms the HCN dimer, iminoacetonitrile (IAN), **2** (Figure 3.3) [139, 158]. Crucially, the cyanide anion

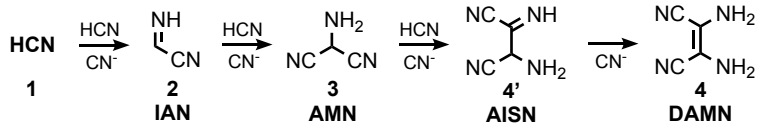

Figure 3.3: First steps of base-catalyzed HCN self-reaction, up to the tetramer (**4**).

regenerates, making it a catalyst.

This step marks the beginning of HCN polymerization. The resulting dimer presents two electrophilic carbon atoms, both susceptible to further attack by CN^-. For instance, successive nucleophilic attacks on the so-formed nitrile groups leads to polyimine, whereas an attack on the imine group (=NH) yields the trimer, aminomalononitrile (AMN), **3** (Figure 3.3) [159, 160]. This latter species features two nitrile groups and a primary amine ($-NH_2$), a new nucleophilic group in the reaction medium, which can then attack electrophiles along with CN^-. Both the dimer and the trimer are highly reactive transient species that have been only indirectly observed [149].

These mechanisms illustrate well the origin of the "**combinatorial explosion**" in HCN chemistry [161]: each reaction step generates new reactive functional groups that, crucially, can be either electrophiles or nucleophiles. Consequently, the number of possible reactions increases "combinatorially" with each addition. Moreover, this complexity is further increased by other reaction types. Processes such as cyclization, cross-linking [162], and redox steps [163] significantly expand the accessible (H,C,N) chemical space. On top of that, the presence of other reactants like water or ammonia, or non-stoichiometric steps such as deamination (loss of NH_3), further fuel this combinatorial explosion. Since the competition between these reactions is likely sensitive to reaction conditions (e.g., temperature, pH, concentration), it is easy to understand why the exact composition of HCN-derived products is so difficult to reproduce experimentally [148].

Despite the chaotic context, other intermediates beyond IAN and AMN have been established. The most prominent example is the tetramer, (cis-) diaminomaleonitrile (DAMN), **4**, which forms via CN^- attack on AMN to produce aminoiminosuccinonitrile (AISN), **4'**,[1] followed by tautomerization. DAMN is strongly thermodynamically favored and forms rapidly [159, 164, 165]. Whether DAMN is the intermediate to most HCN-derived products or a dead-end is still debated. Ferris and co-workers demonstrated that DAMN formation from AMN is virtually irreversible, and suggested that DAMN must be the precursor to all polymers [164]. They also showed that addition of tetramer to diluted NH_4CN solutions accelerated the polymerization [149]. On the other hand, He et al. reported that, from neat HCN, only $\sim 10\%$ of the

[1]Following the notation in **Paper II**, the imine form of imine-enamine tautomers is denoted by adding a prime symbol (').

polymers comes from DAMN [138]. Most likely, multiple pathways compete, and DAMN may be one of the several possible intermediates.

3.4.3 Polymer characterization

The considerable experimental challenges have not deterred researchers from attempting to characterize these HCN-derived products. However, despite decades of study, their exact structure remains elusive. In what follows, I provide a brief overview of the results reported in the literature (see Ref. [136] for a detailed review).

Both aqueous and anhydrous polymerization experiments typically yield two distinct phases: a yellow-brown water-soluble solid and an insoluble black residue [136]. The soluble polymers are considered an intermediate product as their yield decreases with temperature and with the degree of polymerization [166].

Due to the intractability, characterization techniques performed on the black polymers are generally limited to broad-scope techniques, capable of no more than measuring molecular weight distributions and detecting functional groups. Molecular weight estimations reveal wide ranges, spanning at least one order of magnitude. Reported distributions vary from 100 up to 250 kDa [141, 143, 166]. Typical chemical characterization techniques performed on the HCN-derived polymers include Nuclear Magnetic Resonance (NMR), IR spectroscopy, and Mass Spectrometry. When ammonia is added to neat HCN, the H:C:N ratio differs only slightly (4-20%) from the 1:1:1 stoichiometric ratio [111, 148]. However, polymers formed in aqueous environments tend to incorporate oxygen, especially at high temperatures [167, 168]. Several functional groups are frequently detected, including amines, imines, nitriles, olefins (C=C) [138, 148], as well as hydroxyl groups and carbonyls when water is present [167, 169]. Most experiments show evidence of highly unsaturated structures [138, 148, 169, 170].

Consequently, a multitude of different structures have been suggested to be part of the polymeric material, some of which are depicted in Figure 3.4. This diversity comes as no surprise, as the extreme variability of the experimental results inevitably leads to different – often contradictory – conclusions. For example, Völker suggested the "double-ladder" structure (**5**) [171], whereas Umemoto saw evidence of a "single-ladder" one (**6**) [172]; Mamajanov and Herzfeld proposed nitrogen-substituted polyacetylene (−C(H)=N−, **7**) as the main constituent of the polymer [170]; in contrast, He et al. suggested that the majority of the polymer is formed by polyimine (**8**) [138]. Other (albeit different) experiments did not find the same features, and suggested that the polymer is in fact much more complex [162, 173]. To date, there is no consensus regarding the structure or polymerization pathways of these solids. It is plausible that multiple reaction routes occur simultaneously, leading to a vast diversity of heterogeneous polymers that lack a defined repeating unit. Understanding the thermodynamics these various proposed polymers is

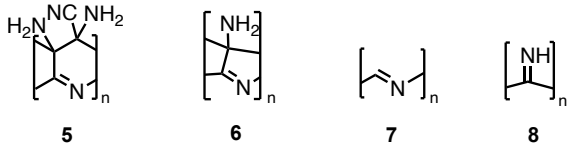

Figure 3.4: Some of the proposed HCN-derived polymers.

therefore crucial. This is the focus of **Paper I**.

3.4.4 Prebiotic chemistry experiments

Different experimental approaches are adopted when the objective is the identification of prebiotically relevant molecules. In these studies, the focus shifts to specific compounds and, more recently, involve powerful analytical techniques (e.g., HPLC-UV, GC-MS).

Prebiotic experiments on HCN commonly employ dilute aqueous solutions (e.g., 0.1 M NH_4CN) at various temperatures (from -78 to 100 °C) and spanning diverse time scales (from hours to years) [137]. Only the water-soluble fraction is retained for analysis, while the insoluble polymer is discarded.[2] Crucially, direct analysis of the supernatant typically yields only trace amounts of biomolecules. For this reason, most experiments involve a harsh acid hydrolysis step (commonly 6 M HCl at 100 °C for 24 h), which significantly increases the yields [137]. It is unclear whether this treatment converts precursors into their final forms, liberates species trapped within the polymer matrix, or both.

3.4.5 Building blocks of life from HCN

To date, a multitude of prebiotically-relevant molecules has been successfully detected in HCN self-reaction experiments (see Ref. [137] for an exhaustive review of the experiments). For example, a total of twenty-four amino acids have been detected, particularly glycine, alanine, and aspartic acid. Furthermore, the detection of six purines and nine pyrimidines has been reported, including all five standard nucleobases (adenine, guanine, cytosine, thymine, and uracil). Other families of molecules identified include hydantoins and pteridines. Unfortunately, the formation mechanisms for virtually all of these compounds remain mostly unknown.

Formation of adenine

Among the biomolecules synthesized from HCN, adenine occupies a unique position. This nucleobases is formally a HCN pentamer ($C_5H_5N_5$) and is

[2]Early experiments usually performed hydrolysis on both fractions indiscriminately, and the insoluble part was discarded afterwards.

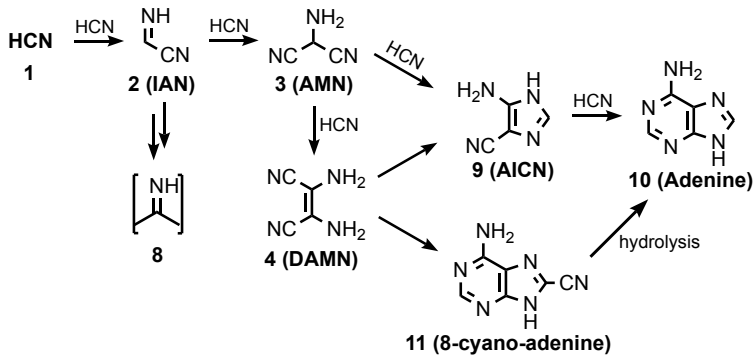

Figure 3.5: Summary of the main proposed base-catalyzed liquid phase reaction pathways from HCN to adenine [57, 183]. Also depicted is polyimine formation. Arrows can represent multiple mechanisms.

synthesizable through a simple experimental procedure. As Leslie Orgel wrote, these features make it "hard to believe that it is a coincidence" [150].

Typical reported yields of adenine from the hydrolyzed supernatant are a fraction of a percent ($\sim$0.001-0.1%) of the initial HCN [137]. Hydrolysis substantially increases the yield: only trace amounts are found from the unhydrolyzed medium [131, 145, 147, 174]. While specific reaction conditions can increase selectivity (e.g., HCN in excess liquid ammonia at high temperature yields 22% of adenine [175]), the typical observed rates of formation reveal that the pathway to adenine is likely kinetically disfavored compared to other routes.

Decades of experiments have contributed to delineate a (blurred) picture of the base-catalyzed mechanism in liquid phase [57, 137]. The main proposed pathways (Figure 3.5) involve the formation of the trimer (AMN) and the tetramer (DAMN). Both species are supposed to convert to 4-amino-1H-imidazole-5-carbonitrile (AICN), **9**, a known HCN-derived molecule [149, 176, 177] and a likely intermediate to adenine (**10**) [178–181]. While AICN formation in diluted aqueous solutions appears slow [149, 176], its synthesis seems to be enhanced in ammonia-rich environments [149, 175]. Depending on the conditions, other HCN-carriers are believed to mediate AICN formation. In particular, water-rich environment promote formamide formation, which may react with AMN to yield AICN [157]; conversely, concentrated ammonia leads to formamidine, the amination product of HCN, which may react with the tetramer [149, 177]. Formamidine is also invoked to react with AICN to give adenine. Finally, AICN can also form through a UV-induced isomerization of DAMN, which represents a compelling mechanism for the prebiotic formation of adenine from HCN [149, 175, 178, 182].

An alternative mechanism to adenine, that does not involve AICN, has been proposed by Voet and Schwartz [183]. Their suggested pathway derives

from several indirect observations, including that of substituted imidazoles and 8-substituted purines. The route leads to **8-cyano-adenine** (**11**, Figure 3.5), which is claimed to quantitatively convert to adenine upon acid hydrolysis, via 8-carboxamide-adenine. This mechanism is suggested to explain the increase in adenine yield after hydrolysis.

More recently, many computational studies helped elucidate the mechanistic details. For example, Sandström and Rahm investigated the initial steps up to DAMN and a polyimine fragment [159]; Wang et al. computed the AMN path in full using formamide as a catalyst [184]. Furthermore, other computational works focused on certain steps, such as HCN addition to AICN [185, 186], or the photochemical conversion of DAMN to AICN [182]. However, a comprehensive picture is still missing. **Paper II** provides a thorough computational exploration of these pathways.

Chapter 4

Theory and Computational Methods

"[...] the mathematical treatment of a large part of physics and the the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble"

– Paul A. M. Dirac, 1929

Prebiotic chemistry scenarios are particularly challenging to investigate experimentally. Tasks as isolating transient intermediates, resolving specific mechanisms, or maintaining cryogenic temperatures are often arduous. Computational chemistry can circumvent these difficulties and serves as a crucial complement to experiments. This is particularly important for HCN, given the toxicity and the complex reactivity discussed in the previous chapter.

This chapter outlines the theoretical framework and computational techniques underlying the research. The goal is to provide the necessary background to interpret the results presented in the following chapters. For a more in-depth discussion, the interested reader is referred to Ref. [187].

4.1 Quantum Mechanics for Chemists

Quantum chemistry is the branch of quantum physics that focuses on chemical systems. Its purpose is to determine chemical properties *ab initio*, i.e. solely using fundamental physics, although, as we shall see, some manual tweaking will sometimes be necessary.

4.1.1 The electronic structure problem

In quantum chemistry, the goal is typically to solve the time-independent Schrödinger equation (SE), an eigenvector equation:

$$\hat{H}\,|\Psi\rangle = E\,|\Psi\rangle\,, \tag{4.1}$$

where $\hat{H}$, $|\Psi\rangle$, and E are the system's Hamiltonian, wavefunction, and total energy, respectively. We can think of a chemical system as a collection of M nuclei and N electrons. In this context, it is more convenient to express the SE in position space:[1]

$$\hat{H}\Psi(\mathbf{r},\mathbf{R}) = E\Psi(\mathbf{r},\mathbf{R}), \tag{4.2}$$

where $\Psi(\mathbf{r},\mathbf{R}) \equiv \Psi(\mathbf{r}_1,...,\mathbf{r}_N,\mathbf{R}_1,...,\mathbf{R}_M)$ is a function of electrons with coordinates $\mathbf{r} \equiv \{\mathbf{r}_i\}_{i=1}^{N}$ and nuclei with coordinates $\mathbf{R} \equiv \{\mathbf{R}_\alpha\}_{\alpha=1}^{M}$.

The energy of a system corresponds to the sum of the kinetic energy and the potential energy of all particles. In short, the Hamiltonian is given by:

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn}, \tag{4.3}$$

where $\hat{T}$ denotes the single-particle kinetic energy operator, and $\hat{V}$ stands for the two-particle potential energy operator, which for our purposes is the classical Coulomb electrostatic interaction. The subscript e stands for electrons, whereas n for nuclei; for example, V_{en} indicates the (sum of) electron-nucleus interactions.

Unfortunately, such problem is a tough nut to crack, which forces us to resort to approximations. The typical starting point in quantum chemistry is the **Born-Oppenheimer (BO) approximation** [188]. This approach assumes that the motion of nuclei and electrons is uncorrelated due to their difference in mass. From the perspective of the electrons, the nuclei appear frozen in space. This approximation allows us to split Equation (4.2) into two: a nuclear SE and an electronic SE. Our main focus is the latter, also known as the **electronic structure problem**:

$$\hat{H}_e\Psi_e(\mathbf{x}) = E_e\,\Psi_e(\mathbf{x}), \tag{4.4}$$

where $\Psi_e(\mathbf{x})$ is the **electronic wavefunction**. This function also depends on the **spin** (σ), which is an intrinsic property of electrons and can be "up" or "down". We use the notation $\mathbf{x} = (\mathbf{r},\sigma)$. $\hat{H}_e$ is the electronic Hamiltonian, which also contains the nuclei-nuclei interaction term $\hat{V}_{nn}$, a mere constant. Lastly, $E_e(\mathbf{R})$ is the (ground-state) **electronic energy**. The zero of energy ($E_e = 0$) corresponds to all particles at infinite separation. Thus, the electronic energy is the energy change associated with bringing all nuclei and electrons together. Equation (4.4) implicitly depends on the nuclear coordinates ($\mathbf{R}$), which are treated as parameters. The function $E_e(\mathbf{R})$ is called potential energy surface (PES), because it the potential energy that the nuclei "feel" while moving, within the BO approximation.

[1] For simplicity, I assume we are interested in the ground state only.

Despite this approximation, solving the electronic SE remains an arduous task. In fact, Equation (4.4) cannot be solved analytically except for the simplest one-electron systems (the hydrogen-like atoms). The "elephant in the room" is the electron-electron interaction term, $\hat{V}_{ee}$. In a nutshell, because this Coulomb term depends on all instantaneous electron-electron distances, the position of an electron depends on the position of all the other electrons. This interdependence makes it impossible to obtain an exact analytical solution. Thus, quantum chemistry is essentially the quest to accurately model (or circumvent) this interaction.

4.1.2 Slater determinants

We are now left with the challenge of mathematically representing the electronic wavefunction $\Psi_e(\mathbf{x})$. In many-body systems, the practical approach is to build the total wavefunction in terms of single-particle wavefunctions $\varphi_i(\mathbf{x})$, called **spin-orbitals**. For example, we can write a two-electron wavefunction as:

$$\Psi_e(\mathbf{x}_1, \mathbf{x}_2) = \varphi_1(\mathbf{x}_1)\varphi_2(\mathbf{x}_2). \tag{4.5}$$

While mathematically convenient, this expression, known as **Hartree product**, introduces a strong approximation, because it effectively decouples the motion of the two particles, i.e., they don't "feel" each other.

We continue with this "spin-orbital approach" and account for another quantum property: antisymmetry. According to the symmetrization postulate, we must ensure that the wavefunction is antisymmetric with respect to exchange of any two particles. As the Hartree product is not antisymmetric, we need to "arrange" the spin-orbitals in a different way. As it turns out, the mathematical construction that makes the wavefunction antisymmetric is the **Slater determinant**:

$$\Psi_e(\mathbf{x}) \approx \frac{1}{\sqrt{N}} \begin{vmatrix} \varphi_1(\mathbf{x_1}) & \varphi_2(\mathbf{x_1}) & \dots & \varphi_N(\mathbf{x_1}) \\ \varphi_1(\mathbf{x_2}) & \varphi_2(\mathbf{x_2}) & \dots & \varphi_N(\mathbf{x_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{x_N}) & \varphi_2(\mathbf{x_N}) & \dots & \varphi_N(\mathbf{x_N}) \end{vmatrix} \equiv |\varphi_1\varphi_2 \dots \varphi_N\rangle \,. \tag{4.6}$$

In this notation, $\varphi_i(\mathbf{x}_j)$ means "the electron j in the spin orbital i". The ket representation (the rightmost term) is a compact notation of the determinant.

The use of Slater determinant, partially solves the "uncorrelated motion" problem. In fact, this construction correctly accounts for the repulsion between electrons of the same spin, a phenomenon known as the **exchange interaction**. However, the motion of electrons with opposite spins remains uncorrelated.

4.1.3 Hartree-Fock approximation

The introduction of Slater determinants naturally leads to the Hartree-Fock (HF) method, which is the first and most important approach to solve the

electronic structure problem. In HF, electron-electron interaction is averaged out, and each electron "feels" the mean electrostatic potential generated by the other $N-1$ electrons, instead of the true instantaneous potential. This approach is called **mean-field** approximation.

The construction of the HF wavefunction relies on the **variational principle**, which states that the energy of any trial wavefunction is an upper bound to the exact ground state energy. Therefore, the HF solution Ψ_e^{HF} is the set of spin-orbitals φ_i, arranged in a single Slater determinant, that minimizes the electronic energy. In mathematical terms, the equation is simply:

$$\hat{H}_e \Psi_e^{\mathrm{HF}} = E_e^{\mathrm{HF}} \Psi_e^{\mathrm{HF}} \tag{4.7}$$

or, making each term explicit:

$$\left[\hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn}\right] |\varphi_1 \varphi_2 ... \varphi_N\rangle = E_e^{\mathrm{HF}} |\varphi_1 \varphi_2 ... \varphi_N\rangle \,. \tag{4.8}$$

The minimization is carried out with the constraint of keeping φ_i normalized and orthogonal (orthonormal) to each other. While a thorough description of the theory lies outside the scope of this thesis, it is important to mention that this orthonormality constraint imposed to spin orbitals is a fundamental requirement of HF theory. This requirement simplifies a N-particle problem into a sum of more easily solvable 1 and 2-particle terms, called one-electron and two-electron integrals.

The solution of these HF equations is a new set of molecular orbitals $\varphi_i(\mathbf{x})$ with energy ε_i. If the initial set of spin orbitals is chosen larger than the number of electrons in the system (N), only the first N orbitals will be occupied, whereas the rest, called **virtual** orbitals, remain unoccupied (Figure 4.1). Importantly, expanding the set of initial spin-orbitals improves the HF solution; however, it does not fix the lack of electron correlation, as the limitations imposed by the single Slater determinant remains.

Although crude, this approach can provide rather good results, especially for equilibrium geometries. However, when quantitative results are needed, HF is often unsatisfactory. Nevertheless, this effort is not wasted. Since obtaining the HF wavefunction is relatively cheap, a common approach is to use the HF solution as a starting point and build upon it to improve the accuracy. These are the so-called **post-HF methods**. In practice, we need to get free from the constraint imposed by a single Slater determinant, typically by making use of the virtual orbitals. With that, the description of electron-electron interactions can be improved. As an Italian would say: we made electron correlation walk out the door, and now we let it come back in through the window.

4.1.4 Coupled-cluster theory

Coupled-cluster (CC) theory is a powerful post-HF method, which introduces the electron correlation by including electronic excitations in the wavefunction.

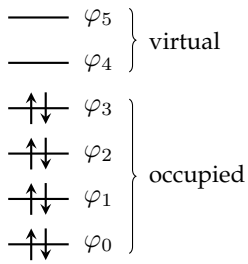

Figure 4.1: Schematic representation of HF energy levels (in order of energy, not to scale), where a Slater determinant composed of six spin-orbitals is used to describe a eight-electron system. The orbitals are labeled and filled with increasing energy and can be occupied up to two electrons: one spin up and one spin down (represented by arrows).

As the theory is rather complicated I will limit myself to explaining the basic concepts.

In the CC framework, the *exact* wavefunction is constructed by applying an excitation operator ($\hat{T}$) to a reference wavefunction. This operator can be written in the form:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... + \hat{T}_N \,, \tag{4.9}$$

where $\hat{T}_1$, $\hat{T}_2$, and $\hat{T}_3$ correspond to all single, double, and triple excitations, respectively. What these operators do is swapping electrons from occupied orbitals to virtual orbitals, thus effectively describing electron excitations. For example, $\hat{T}_1$ takes one electron from an occupied orbital i and places it into a virtual orbital a, for all combinations of i and a (describing all possible single excitations):

$$\hat{T}_1 \Psi_e^{\mathrm{HF}}(\mathbf{x}) = \hat{T}_1 \left|\varphi_1 ... \varphi_N\right\rangle = \sum_{i,a} t_i^a \left|\varphi_1 ... \varphi_{i-1} \varphi_a \varphi_{i+1} ... \varphi_N\right\rangle \tag{4.10}$$

where t_i^a are unknown coefficients. Similarly, $\hat{T}_2$ excites all pairs of electrons, with coefficients t_{ij}^{ab}. The sum in Equation (4.9) goes up to the number of electrons, meaning that all possible electron excitations are considered. It is important to specify that these are not physical electronic excitations, but rather a mathematical tool to go beyond the single determinant. To describe the exact wavefunction, these excitations are introduced by means of an exponential:

$$\Psi(\mathbf{x}) = e^{\hat{T}} \, \Psi_e^{\mathrm{HF}}(\mathbf{x}). \tag{4.11}$$

The exponential can be expanded in Taylor series, giving:

$$\begin{aligned} e^{\hat{T}} &= 1 + \hat{T} + \frac{1}{2!} \hat{T}^2 + ... \\ &= 1 + \left(\hat{T}_1 + \hat{T}_2 + ...\right) + \frac{1}{2!} \left(\hat{T}_1 + \hat{T}_2 + ...\right)^2 + ... \,, \end{aligned} \tag{4.12}$$

which, of course, needs to be truncated somewhere. Usually, the truncation is chosen at the second order, i.e., up to two-electron excitations. This approach is called CC including single and double excitations (CCSD). The most widely adopted approach is however an extension of it: CCSD including perturbative triples (CCSD(T)) [189], in which triple excitation coefficients t_{ijk}^{abc} are found in an approximate – less expensive – way from the CCSD solution. CCSD(T) is highly accurate, and is widely considered as the gold standard in quantum chemistry [190].

4.1.5 Density functional theory

Density functional theory (DFT) is a widely used approach in quantum chemistry and solid-state physics. DFT takes a different path compared to post-HF methods: instead of striving to find an accurate representation of the wavefunction, it uses the **electron density**, $\rho(\mathbf{r})$. The electron density is the probability of finding electrons in a given volume of space, and is defined as the integral of the electronic wavefunction over all coordinates, including spins (a sum), except one spatial coordinate:[2]

$$\rho(\mathbf{r}) \coloneqq N \sum_{\sigma_1} ... \sum_{\sigma_N} \int d\mathbf{r}_2 ... \int d\mathbf{r}_N \left|\Psi_e(\mathbf{r}_1, \sigma_1, \mathbf{r}_2, \sigma_2, ..., \mathbf{r}_N, \sigma_N)\right|^2 . \quad (4.13)$$

This approach offers a significant advantage: it replaces the wavefunction, which depends on $3N$ spatial variables, with the electron density, a function of only 3 spatial variables.

Fundamentals

While the revolutionary idea upon which DFT is built dates back to 1927, the formalization was set forth in 1964 by Pierre Hohenberg and Walter Kohn. Hohenberg and Kohn (HK) formulated two theorems [191]. The first HK theorem proves that the electron density uniquely determines an electronic Hamiltonian, and thus an electronic wavefunction, i.e.:

$$\rho(\mathbf{r}) \implies \hat{H}_e \implies \Psi_e(\mathbf{x}). \quad (4.14)$$

In other words, there cannot exist multiple electron densities that generate the same Hamiltonian and, consequently, the same wavefunction. This first theorem also implies that the ground-state electronic energy E_0 is a **functional** of the ground-state electron density, $\rho_0(\mathbf{r})$, so that we can write:

$$E_0[\rho_0] = T_e[\rho_0] + E_{ee}[\rho_0] + E_{en}[\rho_0], \quad (4.15)$$

where T_e is the kinetic energy of the electrons, E_{ee} is the electron-electron repulsion, and E_{en} is the electron-nucleus attraction. The second HK theorem

[2] Since electrons are indistinguishable, it is irrelevant which $\mathbf{r}_i$ is taken out.

proves that the variational principle works for the wavefunction as well. In short, the two HK theorems establish that the electron density can be used instead of the wavefunction to solve quantum chemical problems. However, not only is the electron-electron interaction still unsolvable, but we created another problem: the kinetic energy term, $T_e[\rho_0]$, is unknown.

A recurring philosophy in DFT is to separate unknown terms from what is analytically solvable. The idea is to calculate exactly anything that we can, and dispose the rest in a "garbage bin" term, which contains everything that we did not account for. Following this approach, we can think of the electron-electron interaction energy ($E_{ee}[\rho_0]$) as being composed of a (known) classical electrostatic repulsion within an electron cloud, called the Coulomb term ($J[\rho_0]$), and the non-classical rest.

In 1965, Kohn and Lu Jeu Sham [192] applied this same logic to the kinetic energy. The underlying assumption of the Kohn-Sham (KS) approach is that, for a system of N (interacting) electrons, there exists a fictitious system of N non-interacting electrons subjected to an effective potential $V_s(\mathbf{r})$ such that it generates the same electron density $\rho_0(\mathbf{r})$. Thus, the true kinetic energy ($T_e[\rho_0]$) is split into the kinetic energy of this fictitious system ($T_s[\rho_0]$) and the non-classical rest. By combining the two unknown terms (from the electron-electron repulsion and from the kinetic energy) we obtain the **exchange-correlation (XC) functional**, $E_{\text{XC}}[\rho_0]$, a term that contains everything we don't know how to calculate. Specifically, $E_{\text{XC}}[\rho_0]$ includes all quantum effects. The total energy functional is rewritten as:

$$E_0[\rho_0] = T_s[\rho_0] + J[\rho_0] + E_{en}[\rho_0] + E_{\text{XC}}[\rho_0]. \tag{4.16}$$

Although there is no known form for $T_s[\rho_0]$, Kohn and Sham proposed a method to calculate it. Because the electrons in this fictitious system are non-interacting, they are perfectly described by a set of single-particle wavefunctions arranged in a Slater determinant. Yes: the orbitals were pushed out of the door, only to be let in through the window.

These single-particle spatial functions ($\varphi(\mathbf{r})$), known as KS orbitals, are determined by solving N single-particle equations, the **KS equations**, which contain the kinetic energy operator and the effective potential:

$$\left[\hat{T} + V_s(\mathbf{r})\right] \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) \qquad \text{such that } \sum_{i=1}^{N} \left|\varphi_i(\mathbf{r})\right|^2 = \rho_0(\mathbf{r})\,. \tag{4.17}$$

The effective potential is composed of the terms in Equation (4.16), but in form of potential. The only missing task is to find an explicit expression of the XC functional.

Exchange-correlation energy approximations

Over the decades, *hundreds* of XC functionals (from now on, simply "functional") have been developed, and the number increases each year [193].

To complicate things further, their performance vary wildly across different systems, making it difficult to establish a proper hierarchy among the various approaches. This mess is a well-known weak point of DFT, and source of strong criticism [194]. However, some order does exist. John P. Perdew proposed the famous **Jacob's Ladder**, [195] which groups the functionals in various rungs, spanning from the "Earth" of Hartree-Fock to the "Heaven" of chemical accuracy.

The first rung of Jacob's ladder, as well as the simplest approach, was devised by Kohn and Sham themselves. They suggested an expression of the inherently *non-local* exchange-correlation effects with a function of the *local* electron density. This approach is called **local density approximation (LDA)**, and gives fair results in materials where the electron density varies slowly, as in metals, but the approximations are generally considered too crude [196].

The second rung of the ladder is occupied by **generalized gradient approximation (GGA)** functionals, in which a dependence on the electron density gradient is added. This addition provides a better description of non-locality, as the derivative of the density contains some information on its surroundings. GGAs significantly improve the results for bond lengths and energies, but it remains not indicated for reaction barriers [197]. The most popular GGA functional has been devised by Perdew, Burke, and Ernzerhof in 1996, and it takes the name of Perdew-Burke-Ernzerhof (PBE) [198]. PBE is widely used in solid-state systems thanks to its good general performance [199]. **Papers III–IV** employ PBE to study the HCN crystal. Continuing the GGA approach, a dependence on the second derivative of the density can also be included, giving **meta-GGA** functionals, the third rung. Both GGA and meta-GGA functionals are called **semilocal**.

The major deficiency of DFT functionals up to the third rung is **self-interaction error** [200]. As the Coulomb potential is computed from the total density, each electron "feels" the repulsion by the entire cloud, including itself! To minimize this problem, a "cheat" becomes necessary: incorporate into the XC functional a portion of the HF exchange interaction. The functionals that contain the HF exchange are said to be **hybrid**, either hybrid-GGAs or hybrid-meta-GGAs, and occupy the fourth rung of Jacob's ladder. Importantly, while this addiction makes the exchange part non-local, the correlation part remains semilocal. The most recognized hybrid-GGA functional is indubitably Becke, 3-parameter, Lee-Yang-Parr (B3LYP), which success stems from its outstanding performance for many types of system, including organic molecules [196, 201]. For these reasons, B3LYP is the the primary computational method employed in this thesis (**Papers I–III**).

Finally, the fifth rung belongs to **double-hybrid** functionals, which include a perturbation term that depends on unoccupied orbitals and makes them fully non-local.

The selection of an appropriate density functional is a non-trivial task. To carry out a full-scale benchmark of all available functionals for every new system is impractical. Therefore, the standard practice, also adopted in this

thesis, is to choose functionals that have been already validated in the literature for similar systems, and subsequently validate them on a subset of cases.

Going empirical

XC functionals can be further distinguished by whether data (from experiment or highly accurate calculations) is used to parametrize them. In particular, **non-empirical** functionals (e.g., PBE) are purely derived from theory, while **empirical** functionals (e.g., B3LYP) are obtained from fitting reference data. The use of experimental data is source of a heated debate between "purists" and "pragmatists" [202].

The second major problem of DFT, besides the self-interaction error, is its failure in describing long-range dispersion interactions [200]. This flaw affects all functionals with local and semilocal correlation, and thus it affects all rungs up to hybrids. For these functionals, **dispersion corrections** should always be used [200]. The idea is simple: since DFT does not describe long-range interactions, these can be included *a posteriori* as an additional energy term:

$$E_{e,\text{total}}^{\text{DFT}} = E_e^{\text{DFT}} + E_{\text{disp}}. \tag{4.18}$$

The dispersion correction E_{disp} is computed by summing contributions from each pair of atoms at a distance longer than a certain threshold, and are often empirical. Throughout this thesis, the Becke-Johnson (BJ) damping variant of D3 (D3(BJ)) dispersion correction is used [203, 204].

4.1.6 Handling extended systems

Solid-state systems are often computed exploiting their periodicity in the bulk. This is introduced with periodic boundary conditions (PBC).

A crystals lattice is defined by a unit cell, the smallest repeating unit. In other words, the entire solid can be reproduced by translating the unit cell by all possible lattice vectors $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. This is very convenient, because a region of space with just a handful of atoms[3] contains all the information we need to describe the bulk behavior. We only need a mathematical construction to represent the periodicity.

In a crystal lattice, the potential operator ($\hat{V}(\mathbf{r})$) possesses the same periodicity as the lattice, such that:

$$\hat{V}(\mathbf{r} + \mathbf{R}_\mathbf{a}) = \hat{V}(\mathbf{r}) \qquad \forall \mathbf{R}_\mathbf{a} = \mathbf{n} \cdot \mathbf{a} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3\,, \tag{4.19}$$

where $\mathbf{R}_\mathbf{a}$ is a linear combination of the lattice vectors. The periodicity of the potential implies that the single-particle solutions of the Hamiltonian must satisfy **Bloch's Theorem**. This theorem states that the wavefunction for a single electron ($\psi(\mathbf{r})$) in a periodic potential can be written as the product of

[3] Hand size may vary.

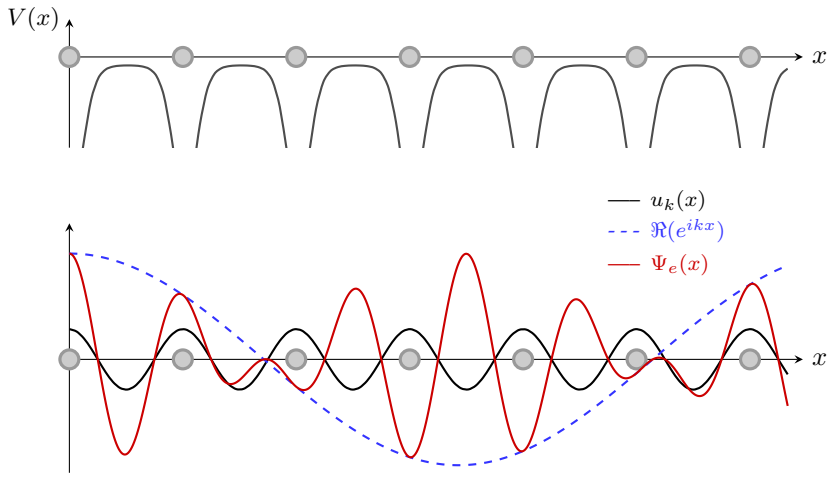

Figure 4.2: Example of Bloch solution for a one-dimensional lattice with one atom (gray circles) per cell. The upper plot shows the periodic potential $V(x)$,.

two functions: a periodic function $u_{n,\mathbf{k}}(\mathbf{r})$, with the same periodicity of the lattice and that reflects the local environment, and an envelope function $e^{i\mathbf{k}\cdot\mathbf{r}}$, a plane-wave phase factor describes a phase change upon translation:[4]

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\, u_{n,\mathbf{k}}(\mathbf{r}), \tag{4.20}$$

where n is the energy level, and $\mathbf{k}$ is a reciprocal-space vector. The reciprocal space is a mathematical construct that is perfectly suited for describing periodic functions. By exploiting the periodicity of the system, the calculation on the infinite bulk solid can be reduced to a finite domain: it can be shown that all unique electronic states are contained within the first **Brillouin Zone**. Geometrically, this zone is defined as the set of points in the reciprocal space that are closer to the origin than to any other reciprocal lattice point.

Electronic band structure: a physicist's view

Solving the Schrödinger equation in a periodic potential for a specific $\mathbf{k}$-vector yields a set of periodic functions $u_{n,\mathbf{k}}(\mathbf{r})$ and their corresponding energies $\varepsilon_n(\mathbf{k})$. In contrast to finite system, characterized by well-spaced discrete energy levels, the energies in an extended system vary smoothly across reciprocal space (as $\mathbf{k}$ is continuous) and form continuous energy surfaces called **electronic bands**.

The electronic behavior of a material is determined by how the electrons in the unit cell fill these bands. In particular, a system is a **metal** if the highest occupied band is partially filled (until the **Fermi level**, at 0 K), which allows

[4]Wavefunctions are defined up to a phase factor, which means that their phase is arbitrary and not physically meaningful. A phase change does not alter the physical state.

electrons to be excited into unoccupied states and move freely at any temperature greater than 0 K. Conversely, if there is an energy gap between the valence band (from valence states) and the conduction band (the lowest unoccupied states), the solid can be an electronic either a **semiconductor** or an **insulator**. The distinction is made based on the magnitude of the **band gap**; if the gap is sufficiently small that electrons can be excited to the conduction band, the solid is a semiconductor, otherwise it is an insulators. In case of a gap, the Fermi level lies in the middle.

Electronic band structure: a chemist's view

The chemist's view of the band structure involves picturing interacting atomic orbitals. As Roald Hoffmann stated, "chemists feel themselves uncomfortable in reciprocal space, and would rather return to, and think in, real space" [205]. Furthermore, we are used to deal with atoms and orbitals rather than states and waves.

In an isolated atom, electrons occupy discrete energy levels. In crystals, atoms are arranged into a periodic lattice and their atomic orbitals overlap and interact. The interaction of 2 orbitals brings 2 new orbitals: the in-phase and the out-of-phase combination of the starting ones. Analogously, N interacting orbitals lead to the formation of N electronic states. Since N in a macroscopic solid approaches infinity, the energy spacing between these states becomes infinitesimal. Thus, the discrete levels gradually merge into a continuum of states, the electronic bands.

Each $\mathbf{k}$-vector represents a certain collection of interacting orbitals of the atoms in the crystal. For example, $\mathbf{k} = \mathbf{0}$ describes a wave that never changes phase, which translates to orbitals that have all the same phase. Similarly, different values of $\mathbf{k}$ leads to different phase combinations between orbitals. The constructive and destructive interference determines the energy of that (crystal) orbital.

4.2 Practical Quantum Mechanics for Chemists

We now shift from abstract formulas to their numerical implementation, the realm of **computational chemistry**. This section describes the essentials of quantum chemical calculations from a computational chemist's pragmatic perspective.

4.2.1 Basis sets

To solve the equations numerically, we need a way to describe the single-particle wavefunctions, whether orbitals φ_i or periodic functions $u_{n,\mathbf{k}}$. These functions are complicated and are not easily handled by a computer. Thus, they are often described in terms of known and more easily manageable

mathematical functions. The general idea is to expand a generic function $\psi(\mathbf{r})$ as a linear combination of basis functions $\chi_j(\mathbf{r})$:

$$\psi(\mathbf{r}) = \sum_{j=1}^{M} c_j \chi_j(\mathbf{r}), \tag{4.21}$$

where M the total number of basis functions. These basis functions form the **basis set**. The larger M, the better ψ is represented, but the more functions the computer has to handle. The combination of method and basis set determines the accuracy and the cost of a calculation; it is referred to as the **level of theory**, and denoted as "method/basis set".

Gaussian basis sets

Gaussian basis sets are typically employed for finite systems, and, less frequently, for extended systems. In this approach, the basis functions $\chi_j(\mathbf{r})$ are **Gaussian type orbitals (GTO)**, and are given, in turn, by a linear combination gaussian functions (gaussians) centered on the nuclei. These gaussians can be of s-type (spherical gaussian), p-type, d-type, and so on (cartesian gaussians), to mimic the symmetry of the various atomic orbitals.

The advantage of GTOs is twofold: the integral of a gaussian is analytically known, and the product of two gaussians is a linear combination of gaussians. Thus, both one- and two-electron integrals can be computed rapidly. The drawback is that gaussian functions poorly describe the exponential ($e^{-\zeta r}$) wavefunction both near and far away from the nuclei. To overcome this latter problem, **diffuse functions**, which are very broad gaussian functions, can be added. Diffuse functions should be included when it is important to accurately describe the electron density away from the nuclei, such as when computing anions, complexes, or excited states.

A plethora of different Gaussian basis sets has been developed. In this thesis, two basis sets are used. For DFT calculations, Pople's 6-31+G(d,p) is used [206]. This basis set is widely adopted as it represents a balanced choice between accuracy and feasibility; it includes polarization functions (higher angular momentum) for hydrogen atoms (p-type) and for second-row atoms (d-type), while also adding diffuse functions (s-type) for second-row atoms. Conversely, for single-point energy corrections (see below), the larger Dunning's correlation-consistent aug-cc-pVDZ and aug-cc-pVTZ basis set are used [207].

Plane-wave basis sets

For extended systems, the periodicity of $u_{n,\mathbf{k}}(\mathbf{r})$ makes plane-waves (PW) the natural choice for a basis set. In this case, the basis function are plane waves $e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}$, where $\mathbf{G}$ is a reciprocal lattice vector. This sum is truncated at a certain kinetic energy cut-off, which is why the cutoff criterion is typically expressed in eV.

Plane-waves offer significant advantages over GTOs for extended systems: they work well in reciprocal space, they are independent of atomic positions, and their convergence is controlled by a single parameter. As plane-waves span the entire space, the basis set quality is uniform throughout the unit cell, as opposed to the atom-centered GTOs. However, plane-waves make vacuum computationally expensive. Another drawback of plane-waves is that they are inefficient for describing all electrons. Since the wavefunction of core electrons oscillates rapidly, to correctly represent it would require very large cutoffs. A commonly employed solution is the Projector Augmented Wave (PAW) method [208], which uses computationally efficient pseudo-wavefunction for the core region.

4.2.2 Self-consistent field

Once the electrons are described by a basis set, the eigenvalue equations can be solved. These equations are non-linear: the solution depends on itself. Thus, they cannot be solved directly, but rather iteratively, a process called **Self-Consistent Field (SCF)**.

After an initial guess, the electron-electron interactions are calculated to obtain a new set of wavefunctions and energy levels. These new results are then used to update the interaction terms, and the cycle repeats until convergence. The convergence criteria are typically based on the changes in energy and/or electron density. Solving the SCF cycle for a fixed set of atomic positions is known as a **single-point** calculation. This yields the final electronic energy and the converged wavefunction, or electron density.

4.2.3 Computational cost

The "computational cost" measures the resources needed to perform a calculation. It represents the major limiting factor in computational chemistry.

The cost of a calculation is largely determined by the method and the number of basis functions (M). In particular, the scaling is often polynomial and is denoted as $O(M^n)$. The value of n depends on the chosen method and is dictated by the computational bottleneck. Two types of scaling can be defined: in CPU time and in memory. The allocated memory sometimes represents the major limitation of a calculation: while longer times can be tolerated by just waiting longer, running out of RAM makes the calculation crash.

In addition to this scaling, there exists an often hidden prefactor (CM^n) that governs the computational costs for small to medium systems. For instance, two methods can scale equally (same n) but may presents very different prefactors, resulting in two tangibly different effective performances.

DFT scaling

In a DFT calculation, the limiting resource is typically the computational time. The CPU scaling strongly depends on the functional, and is roughly determined by the rung in the Jacob's ladder. For example, GGA functionals scale as $O(M^3)$, which is considered "cheap". In contrast, hybrid-DFT functionals scale as $O(M^4)$ due to the evaluation of HF exchange integrals.

CCSD(T) scaling

Post-HF methods are notoriously costly, due to the large number of integrals. The CCSD(T) method scales as an astonishing $O(M^7)$ [190]; this means that the computational time doubles upon increasing M by a mere $\approx 10\%$.

However, the major issue for the applicability of CC methods in general is the memory. While in DFT integrals are calculated "on the fly", in these correlated methods all integrals must be stored somewhere, which requires an incredible amount of memory. The required memory for CCSD(T) scales with $O(M^4)$. This scaling often renders calculations impractical, as the used RAM *doubles* each time M increases by $\approx 19\%$.[5]

Domain-Based Local Pair Natural Orbital approximation

The domain-based local pair natural orbital (DLPNO) approach is a **local correlation** method devised to reduce the computational cost of post-HF methods, while retaining most of their accuracy [209]. The intuition behind local correlation methods is that electron correlation effects decay rapidly with distance, and most of the electron-electron interaction is "local". Thus, the treatment of correlation can be safely truncated after a certain orbital-orbital distance.

Applied to CC theory, the DLPNO-CCSD(T) method typically recovers more than 99.9% of the correlation energy at a fraction of the cost [209]. Remarkably, the CPU scaling becomes linear ($O(M)$) for large systems. For this reason, in this work I use DLPNO-CCSD(T).[6]

4.2.4 Geometry optimization

When computing chemical systems, chemists are typically interested in specific geometries, and in particular in two points in the PES:

- Minima: the equilibrium geometries, where all forces are zero.
- First-order saddle points: transition state (TS), the maximum along the minimum energy path (MEP) of a reaction (Figure 4.3).

[5] I noticed this tremendous increase firsthand: the memory use of CCSD(T) was small for a $H_3C_3N_3$, limited for a $H_4C_4N_4$ system, and gigantic for a $H_5C_5N_5$ system!

[6] Which I apply to systems up to $H_{21}C_{21}N_{21}$.

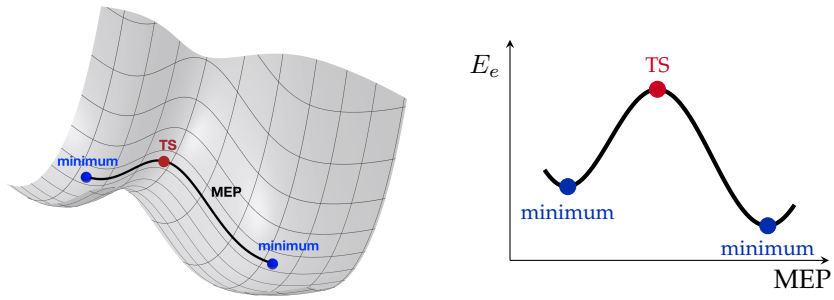

Figure 4.3: Representation of a MEP across a 2D PES.

To reach these points, a **geometry optimization** needs to be performed. A geometry optimization run requires the computation of the first derivative of the electronic energy with respect to the nuclear position, i.e., the forces that the nuclei are subjected to. The, the nuclei are moved according to these forces, and the SCF cycle re-computed on the new geometry (Figure 4.4). Geometry optimizations are more computationally demanding that single-point calculations, because they involve solving the SCF cycle many times and computing the forces on the atoms.

It is important to note that the PES is an unknown function. The only way to evaluate it is to "explore" it. Thus, common geometry optimizations algorithms can only lead to **local minima** (Figure 4.4). Finding the **global** (or absolute) minimum of a PES necessitates a complete exploration of the entire PES, which can be done only for very small systems. While global optimization algorithms exist, they are feasible to implement only with low levels of theory [210].

The same argument applies to TSs. However, this search is significantly more challenging: not only is locating first-order saddle points mathematically non-trivial, but multiple transition states often exist between reactants and products. As with minima, the only way to determine the lowest TS is to compute them all. It is therefore clear that finding the actual MEP is an incredibly challenging task.

In **Papers I–III**, geometry optimizations on molecular systems are performed with Gaussian 16 revision B.01 [211]. In **Papers III–IV**, geometry optimizations on extended systems are performed with the Vienna Ab initio Simulation Package (VASP) version 6.4.1 [212].

4.2.5 Single-point energy correction

A very common approach to increase the accuracy while containing the computational costs is to carry out an **single-point energy correction**. In this approach, a geometry optimization is performed using a "low-level" (LL),

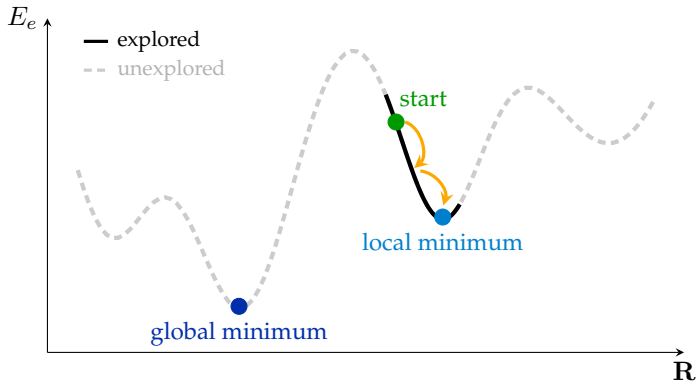

Figure 4.4: Representation of a geometry optimization on a PES. Orange arrows represent geometry optimization steps from a starting geometry to a local minimum. Only the black part of the PES is explored during the optimization, whereas the gray dashed region remains unknown. There is no mathematical way to tell the existence of the global minimum without exploring the whole space.

e.g., HF or DFT with a moderate basis set size. Once the minimum or the TS is found, a high-accuracy method and/or a larger basis set ("high-level, HL") can be employed for a single-point calculations, to obtain a more accurate electronic energy. This refinement relies on the assumption that the two PES are reasonably similar (Figure 4.5). The common nomenclature is "HL-method/HL-basis-set//LL-method/LL-basis-set".

In **Papers II–III**, single-point energy corrections with DLPNO-CCSD(T) are performed with ORCA 6.1 [213].

4.3 Modeling Real-World Phenomena

The computational methods described thus far return the electronic energy and geometry of an ideal system at 0 K. However, real-world chemistry often occurs in complex environments (e.g, a solution) and at finite temperatures. To properly describe the macroscopic reality, taking additional steps is necessary. In this section, I briefly detail some of the computational methods I have employed in my research to model real-world phenomena.

4.3.1 Macroscopic quantities from microscopic particles

In the macroscopic world, collective properties emerge from a system of interacting molecules. Strictly speaking, a single molecule does not possess a temperature, nor can a single reaction be considered spontaneous. This

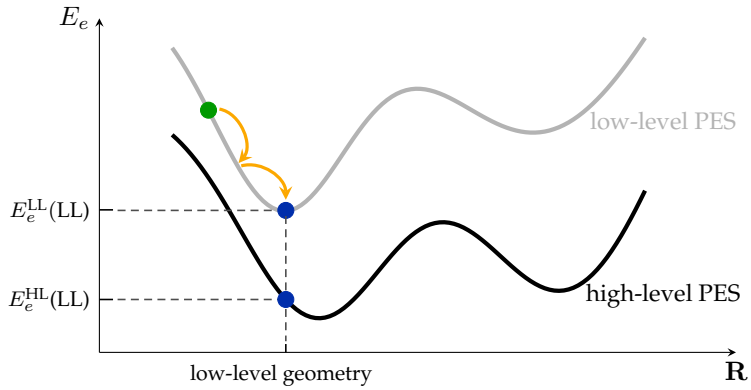

Figure 4.5: Example of an energy refinement. An optimization to a minimum (orange arrows) is performed with a low-level (LL) of theory (e.g., DFT), giving E_e^{LL}(LL). Then, a high-level (HL, e.g., DFT with a larger basis set, or CCSD(T)) single point calculation is performed on the LL minimum geometry, giving a more accurate E_e^{HL}(LL), which approximates the true HL energy on the HL geometry, E_e^{HL}(HL). The LL minimum geometry is generally not a minimum on the HL PES. The relative energy between local minima also likely changes.

represents a challenge for computational methods, as they need to extrapolate emergent properties from single particles. Statistical mechanics comes to the rescue, bridging the gap between quantum mechanics and thermodynamic observables.

The central quantity in statistical thermodynamics is the **partition function** q, which measures the number of thermally accessible microstates in a system. In the canonical ensemble, q is defined as:

$$q = \sum_i g_i e^{-\beta \varepsilon_i}, \tag{4.22}$$

where ε_i is the energy of the microstate i, and g_i the degeneracy. Crucially, the partition function is related to thermodynamic quantities, such as the internal energy, U, and the entropy, S. This quantity is therefore the link between the microscopic and the macroscopic world.

In computational chemistry, the degrees of freedom of a molecule (translation, rotation, vibration, and electronic motion) are assumed to be separable, i.e., they do not influence each other. The partition function is then given by the product of each contributions:

$$q_{tot} = q_{trans} \cdot q_{rot} \cdot q_{vib} \cdot q_{elec}. \tag{4.23}$$

These quantities are not attainable in practice, and need to be computed through various approximations. The electronic partition function q_{elec} is normally set to 1, as it is assumed that the ground state is non-degenerate

and that the first excited state is thermally inaccessible. The other terms are computed using the ideal-gas model. In particular, q_{trans} is obtained from the particle-in-a-box model, q_{rot} from the rigid-rotor approximation, and q_{vib} from the harmonic oscillator approximation.

From these simple models, the contributions to the internal energy and entropy can be computed. These so-called thermal corrections U_{corr} and S_{corr} are then used to calculate the standard-state Gibbs energy in the gas phase, $G^{\circ}_{(g)}$:

$$G^{\circ}_{(g)} = E_e + U_{\text{corr}} + k_B T - T S_{\text{corr}}, \tag{4.24}$$

where k_B is the Boltzmann constant, T is the temperature, and E_e is the electronic energy (from the SCF). We also exploited the ideal gas relation $pV = Nk_BT$ with $N = 1$.

The computational bottleneck of thermal corrections is the evaluation of contributions from vibrational motion. Furthermore, the vibrational modes can only be determined for a system that lies in a stationary point, which requires a prior geometry optimization. Thus, thermal corrections can only be evaluated with a method for which geometry optimizations are feasible. Then, a single-point (electronic) energy correction can be applied to estimate the Gibbs energy. For example:

$$G^{\circ,\text{CCSD(T)}} \approx E_e^{\text{CCSD(T)}} + G_{\text{corr}}^{\text{DFT}}, \tag{4.25}$$

where $E_e^{\text{CCSD(T)}}$ is the DLPNO-CCSD(T) energy on the DFT geometry, and $G_{\text{corr}}^{\text{DFT}}$ is the thermal correction to the Gibbs energy obtained by DFT. Since thermal corrections represent a fraction of a percent of the Gibbs energy, this approach nonetheless provides accurate estimates [200].

Finally, a note on pressure and concentration. The gas-phase standard state is defined as a pressure $p^{\circ} = 1\,\text{atm}$. The Gibbs energy of an ideal gas depends on the pressure p as:

$$G(p) = G^{\circ} + RT \ln\left(\frac{p}{p^{\circ}}\right). \tag{4.26}$$

As the standard-state concentration of a solute is $c^{\circ} = 1\,\text{M}$, to compute its standard-state Gibbs energy one needs to convert this concentration into a pressure, and calculate the correction to the Gibbs energy, ΔG_{conc}. At room temperature (RT), $1\,\text{M} \implies 24.5\,\text{atm}$, which gives $\Delta G_{\text{conc}} \approx 1.93\,\text{kcal mol}^{-1}$. Furthermore, the standard state of a solvent molecules is its concentration. In this thesis, all species in solution are considered in their standard-state. In **Paper II**, liquid HCN is modeled at 278 K, which corresponds to a concentration of HCN of 26.05 M.

4.3.2 Solvation effects

Solvents are ubiquitous in chemistry. They actively affect reaction rates, equilibria, geometries, and reaction selectivity, among other properties. Solvent

molecules are indeed not mere spectators: they establish intermolecular interactions with both themselves and the solute. However, to properly simulate the bulk liquid phase necessitates explicitly include tens or hundreds of solvent molecules, which is computationally prohibitive. To address this, **implicit solvation** (or continuum) models are often utilized as an efficient alternative.

The **Polarizable Continuum Model (PCM)** [214] is a common implicit solvation approach in computational chemistry. In the PCM framework, the explicit solvent molecules are replaced by a continuous dielectric medium, and the solute is placed inside a cavity of this polarizable continuum. This medium aims to capture the dynamics of solvent molecules around the solute, by describing a sort-of Boltzmann average of all accessible microstates [215]. The dielectric constant and the shape of the cavity are determined by solvent-dependent parameters. The solvation energy, ΔG_{solv}, is then computed self-consistently by evaluating the solute-continuum interaction. Since this interaction directly affects the wavefunction, the solvation energy is typically incorporated into the electronic energy. PCM is used in **Papers I–II**.

Implicit solvent models provide a good description of the electrostatic effect of the solvent, but they do not capture solvent-solute specific interactions. To address this, a certain number of **explicit** solvent molecules need to be included.

4.3.3 Reaction kinetics: transition state theory

Thermodynamics determines whether a reaction is favorable, but it does not indicate the rate at which it occurs. Transition state theory (TST) is a common approach used in computational chemistry to find rate of reactions, as it provides a framework to calculate the reaction rate constants k from the standard Gibbs energy barrier, $\Delta G^{\ddagger,\circ}$, of a reaction.

TST is based on the assumption that the TS is in **quasi-equilibrium** with the reactants, which leads to the **Eyring equation**:

$$k_{\text{TST}} = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger,\circ}}{RT}\right) \tag{4.27}$$

where h is the Planck constant, and R the gas constant. It is important to note that the Gibbs energy of the TS is given by all contributions (as seen in the previous section) except the vibrational mode along the reaction coordinate, which is imaginary. The contribution of the this imaginary mode is contained in the pre-factor, extrapolated to the classical limit.

4.3.4 Surface energies and crystal morphology

Surface energies of solids are commonly computed using slab models. As solid-state programs enforce 3D periodicity, to simulate a semi-infinite surface, a vacuum region is added in the direction normal to the surface. This

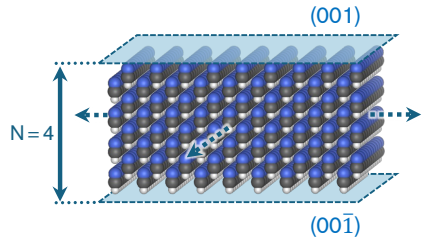

Figure 4.6: Slab model of the $(001)/(00\bar{1})$ HCN crystal surfaces. Dashed arrows represent the directions in which the system is periodic, while the two opposite surfaces are highlighted by light blue planes. The thickness of a slab depends on the number of unit cells N (here 4) in the non-periodic direction.

vacuum must be sufficiently large (typically > 10 Å) to ensure that there is no interaction between adjacent periodic images.

The surface energy γ at 0 K is defined as the energy required to create the surface from the bulk:

$$\gamma = \frac{E_{\text{slab}} - N E_{\text{bulk}}}{2A}, \tag{4.28}$$

where E_{slab} is the total energy of the surface slab, N is the number of unit cells in the slab, E_{bulk} is the energy of the bulk unit cell, and A is the surface area of one side of the slab. The factor of 2 accounts for the two identical surfaces (top and bottom) in the slab.

In case of polar surfaces, slabs are asymmetric and composed by two different terminations. For the HCN crystal (**Papers III–IV**), the polar surfaces can be either N-terminated, (hkl), or H-terminated, $(hk\bar{l})$ (Figure 4.6). Surface energies from Equation 4.28 correspond to the average between the two opposite facets of the slabs, and thus represent an approximation for polar surfaces.

The equilibrium crystal morphology is determined using the **Wulff construction**, which relies on the Gibbs-Curie-Wulff theorem [216]. According to this theorem, the equilibrium shape of a crystal is the one that minimizes the total surface free energy, $G_{\text{surf}} = \sum \gamma_i A_i$. Therefore, facets with high surface energies occupy smaller surface areas, while low-energy facets are predominantly exposed. In **Paper III**, the WulffPack python package [217] is employed to obtain the Wulff construction of the HCN crystal.

CHAPTER 5

The Art of Chemical Modeling

"Gives us insight, not numbers"
– Charles Coulson, 1959

"Garbage in, garbage out."

The previous chapter offered a peek into the world behind computational chemistry programs. While the underlying theory is complex, obtaining a result is often deceptively simple for a user. However, this simplicity should not be taken lightly. Getting an output is merely the culmination of a tortuous journey: **chemical modeling**, the bridge between theory and physical reality.

This chapter outlines some of the key aspects of chemical modeling relevant to this thesis, based on seminal papers [200, 215, 218]. The aim is to provide the necessary tools to evaluate the methodologies and results presented in the following chapters.

5.1 Modeling for Reliable Predictions

One of the main purposes of computational chemistry is making **predictions**. The goal is to gain insight into the real-world, such as how a chemical reaction proceeds, or the properties that a certain material has. Unfortunately, the rapid scaling of computational costs (see Section 4.2.3) represents a major limitation, creating an inevitable trade-off between predictive power and feasibility. This balance is delicate, as the result yielded by a really cheap method may be even deleterious if the underlying physics is so simplified that it leads to the wrong conclusion. Thus, it is crucial to devise a model that captures the nature of the system of interest, providing *reliable* predictions. This requires not only

technical expertise but also creativity and intuition. In this sense, chemical modeling is a form of *art*.

5.1.1 Defining the scope

Modeling requires a clearly defined research question, or **scope**. This involves selecting what to model and what to omit. While factors "outside the scope" could potentially alter the results, this selectivity allow to isolate specific effects. For example, **Paper II** considers a model environment of liquid HCN. The exclusion of water from the model serves to isolate HCN reactivity, providing a reference for future investigations on the role of water. Furthermore, in **Papers III–IV**, the HCN crystal is modeled in a vacuum at 0 K. While approximate, this model permit a characterization of the electrostatic effects of the ideal crystal, from which more complex models can be built on. In essence, the challenge lies in identifying a scope that captures the essential physics of the problem.

5.1.2 Choosing a model

Once the research question is defined, the next challenge is determining how to represent the system. A model is a simplified version of reality, designed to be computable, as well as tunable and interpretable. This simplification entails more than just selecting a level of theory; it also requires making the **choice** of how to build such virtual reality. Since it is unfeasible to describe everything equally, the modeler must decide which aspects of the system deserve more focus and computational resources, and, consequently, which to leave blurred. In the following, I present two examples of such choices from my work.

Static versus dynamic

At finite temperatures, atoms and molecules are in constant motion. Therefore, the ideal approach is to explicitly include the time variable, and describe the dynamics of the system. This is the domain of **Molecular Dynamics (MD)**. When the forces are calculated at the DFT-level (or above), this method is called *ab initio* MD (AIMD).

AIMD provides an in principle robust representation of chemical environments, as the system is allowed to evolve in time by solving the DFT equations at each step. This approach captures thermal and solvation effects explicitly, making it well suited for liquid-phase environments. In principle, Gibbs energies can be extracted from the sampling of the phase space. However, these advantages come with high computational costs, which typically necessitate the use of GGA functionals. Even then, AIMD remains limited to relatively small systems and short time scales.

Conversely, static approaches describes atoms and molecules as static entities at 0 K. The dynamics is replaced by a frozen representation of the

most probable state, and thermal properties are then derived using statistical mechanics. Thanks to this simplified representation, static calculations allow for the use of much higher levels of theory compared to AIMD, such as hybrid-DFT or post-HF methods, which provide significantly more accurate electronic energies than GGAs.

Thus, the choice boils down to the question: is it better to obtain accurate energies for representative static structures, or perform dynamical sampling of an ensemble with approximate methods? In **Papers I–II**, as the accuracy of energies was a critical aspect, the former option was chosen.

Periodic boundary conditions vs cluster modeling

To study crystal surfaces, slab models represent the standard approach. However, similarly to AIMD, extended-system calculations necessitate the use of GGA functionals. While these functionals perform well in describing bulk properties, they sometimes struggle with surface reactivity [197]. A potential alternative is a **cluster model**, where a finite section of the material is treated as an isolated molecule [219]. This approach allows for the use of more expensive methods, though at the cost of losing effects from periodicity.

In **Paper III**, both strategies are employed to focus on different aspects of HCN crystal surfaces. To determine surface energies, slab models are used (see Section 4.3.4). Conversely, to investigate surface reactivity, a cluster models is considered; in this model, polar surfaces are represented by a hydrogen-bonded HCN linear chain. The latter choice allows the computation of reaction energies by means of DLPNO-CCSD(T) on B3LYP geometries, which would be prohibitive in a periodic approach. The trade-off is that these linear chains cannot (and are not meant to) represent the bulk electronic structure of the 2D material.

5.2 The Holy Grail of Chemical Accuracy

Chemical accuracy is the accuracy that is considered necessary to make quantitative chemical predictions [220]. It is commonly defined as $\pm 1\,\mathrm{kcal\,mol^{-1}}$, which is comparable to what is possible to measure using high precision thermochemistry and kinetic experiments. For comparison, a change in Gibbs energy of $1.4\,\mathrm{kcal\,mol^{-1}}$ at RT corresponds to a change of one order of magnitude in many thermodynamic properties, such as the equilibrium constant or the rate of a reaction.

Achieving chemical accuracy is an outstanding challenge. The difficulty primary lies in the scale of the energies involved: the electronic energy of a molecular system is huge. The energy differences that dictate chemistry (e.g., bonds, conformational changes, reaction barriers) are a tiny fraction of that total. As Charles A. Coulson famously wrote, computing bond strengths using ground state energies is the same as "to determine the weight of the captain

of a large ship by weighing the ship when he is and when he is not on board" [221]. The precision of such scale must be outstanding.

5.3 Acknowledging and Managing Model Errors

The sources of error in computational chemistry are numerous, and their compounding effect is virtually unpredictable. Yet, computations often returns a number with twelve decimal places. In fact, the result that calculations provide is as precise as it could be. With the critical clause: it is exact *within* that specific theoretical framework.[1] However, the computational program is ignorant on the chemical system that the model is trying to represent, and it is the responsibility of the latter to assess a model error. In the following, I will discuss the main sources of error in context of the research presented in this thesis.

5.3.1 Level of theory

The accuracy of a model is largely determined by the level of theory. Method and basis set primarily affect the electronic energy and the shape of the PES, from which are derived crucial factors such as equilibrium geometries, thermal corrections, intermolecular interactions, and solvation effects.

In DFT, the choice of the functional is crucial for the accuracy at which the electronic energy is computed. Essentially, the choice depends on the properties of interest. For example, GGA functionals like PBE, while computationally inexpensive, are generally heavily affected by self-interaction error. This is particularly evident for reaction barrier heights, which are systematically underestimated by $5 - 10\,\mathrm{kcal\,mol^{-1}}$ [200]. On the other hand, PBE performs relatively better for solid-state systems [222], although it often underestimate band gaps [223]. Hybrid functionals such as B3LYP provide more reliable reaction energies and barrier heights, with mean absolute errors around roughly $2 - 5\,\mathrm{kcal\,mol^{-1}}$ [224], although the performance varies widely across different types of system [200]. In general, empirical dispersion corrections (Section 4.1.5) improve the results and should always be included [200]. In all works presented in this thesis, the D3(BJ) empirical dispersion is employed, coupled with both PBE and B3LYP.

To mitigate inaccuracies, a computationally efficient approach is to perform single-point energy corrections (see Section 4.2.5). This correction should be carried out using at least a hybrid functional with a large basis set. In **Paper I** B3LYP-D3(BJ)/aug-cc-pVTZ single-point correction was carried out on B3LYP-D3(BJ)/cc-pVDZ geometries. Alternatively, the use of DLPNO-CCSD(T) combined with large basis sets often provides substantially improved results [225, 226]. In **Papers II–III**, reaction energies and barrier heights are

[1] Including numerical approximations.

obtained following DLPNO-CCSD(T)/aug-cc-pVTZ refinement on B3LYP-D3(BJ)/6-31+G(d,p) geometries.

5.3.2 Conformational sampling

In static calculations, the correct determination of the equilibrium geometry is crucial. As these methods lack a description of dynamics, identifying the global minimum is fundamental to correctly modeling the system. This importance stems from the exponential dependence on the Gibbs energy of the Boltzmann distribution (according to the aforementioned $1.4\,\mathrm{kcal\,mol^{-1}} \rightarrow$ one order of magnitude relation). Therefore, the lowest-energy conformer is the most representative species and is the one that typically dominates the chemical properties. Missing low-lying conformers leads to errors that can even exceed those that come from the level of theory [200].

Unfortunately, the exponential increase of the conformational space with the number of atoms [227] makes a thorough conformational sampling prohibitive even for medium-sized systems. This issue is aggravated for TSs. Not only does the search for first-order saddle points suffer from algorithmic instability [228], but it is also difficult to locate the specific TS that connects the correct minima. Furthermore, multiple paths can exist; thus, from the user's perspective, finding the minimum energy path is often a matter of "trial and error".

Automated sampling tools

Automated sampling algorithms represent a promising approach for a rigorous exploration of the PES. These tools typically perform unbiased sampling by generating multiple initial conformers via stochastic algorithms, which are then optimized using computationally efficient methods [200]. These methods are often semi-empirical, such as **GFN2-xTB** [229], which are extremely fast but rely on heavy parametrization. The strength of automated approaches lies in their ability to overcome human limitations and biases, as well as greatly reduce the workload. The goal is to sample a multitude of structures, filter out high-energy conformers, and provide initial geometries that can be refined at the DFT level.

While powerful, these automated approaches present a major limitation, which stems from using low-level methods as a "filter" to screen for good initial structures. As low-level methods can struggle to describe the underlying physics, they may approximate the PES too poorly, which might make the filter faulty. In particular, two closely related problems, as illustrated in 5.1, may arise:

- A minimum on the low-level PES may be too shifted compared to high-level PES, causing the structure to fall into a higher energy minimum after refinement.

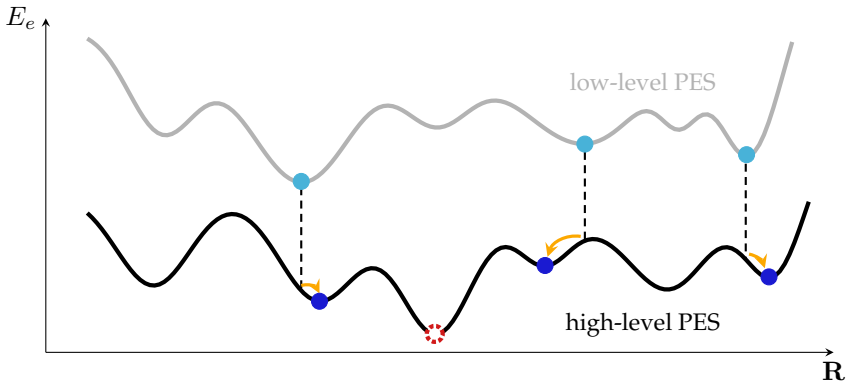

Figure 5.1: Representation of the "filter" problem in automated approached, where a low-level optimization (light blue circles) is followed by high-level refinement (orange arrows to blue circles). The approximate low-level PES may incorrectly discard low-energy conformers, including the global minimum (red dotted circle).

- The low-level method provides an unreliable ranking [200]: the global minimum of the high-level PES might correspond as a high-energy conformer for the low-level method, and therefore filtered out.

A potential solution for the latter problem is to increase the energy cutoff to obtain more conformers. However, the number of selected structures expands exponentially, which increases the workload for visual inspection and refinement. Automated approaches tend to struggle especially with TSs, as their search is non-trivial, and the ranking accuracy of low-level methods is often poorer than for minima. One strategy to improve reliability is to increase the level of theory of the low-level method, as described below. Nevertheless, the difficulty in finding specific mechanisms remains.

Three automated search tools are used in this thesis:

- In **Paper I**, the Conformer-Rotamer Ensemble Sampling Tool (CREST) [230] (with GFN2-xTB) is employed to generate conformers of large polymeric fragments. Although visual inspection was necessary, the tool proved robust for the considered molecules.
- In **Paper II**, the conformational sampling on molecular clusters is carried out with autodE [231]. The advantage of autodE is that it allows for a full customization of the routine. In particular, a sampling scheme with B3LYP-D3(BJ)/4-31G* as a low-level has been devised. While computationally heavier, this was crucial to minimize the filter problem.
- In parallel with autodE, in **Paper II**, the GOAT [210] and SOLVATOR [213] algorithms (using GFN2-xTB, implemented in ORCA 6.1 [213]) are also employed for kinetically relevant steps, to ensure a systematic

exploration of the conformational space.

5.3.3 Low-frequency vibrational modes

In this thesis, the entropy contribution from vibrational motion is computed using the harmonic oscillator approximation. However, this approximation fails for low-frequency vibrational modes, as the calculated entropy diverges to infinity as the frequency approaches zero, leading to unphysically large entropy corrections [218]. These errors are typically of 1–2 kcal mol^{-1} at RT. To mitigate this, the standard approach is to employ quasi-harmonic treatments, such as the rigid-rotor-harmonic-oscillator (RRHO) approximation [203]. This approach, adopted in **Papers II–III**, treats low-frequency modes (typically $<$100 cm^{-1}) as rotations [200, 218].

5.3.4 Solvation modeling

The treatment of solvation represents a critical aspect of static calculations, and must be approached with caution.

Implicit solvation

Implicit solvation is the standard approach to introduce solvation effects. Their major limitation stems from their lack in the description of specific solute-solvent interactions [215, 218]. PCM, employed in **Papers I–II**, is often associated to error of 4–5 kcal mol^{-1} for neutral molecules, and of $\sim$10–12 kcal mol^{-1} for charged species. Thus, a certain number of explicit solvent molecules should be included to improve the description of solvation effects [218].

Hybrid approaches

A hybrid approach combines explicit and implicit solvation. While this approach provides a description of specific interactions, it introduces a new layer of complexity. The challenge is to include enough molecules to describe the first solvation shell, but without rendering a thorough conformational sampling unfeasible. The problem of an exhaustive conformational search is aggravated by the typical flat PES, characterized by a multitude of minima very close in energy which correspond to slightly different positions of the solvent molecules around a solute. Finally, one should be careful not to introduce uneven descriptions of solvation effects, for example by using a different number of solvent molecules across a reaction profile, or by performing a biased conformational search [218].

In **Paper II**, in modeling liquid HCN, a robust conformational sampling scheme of explicit solvation is adopted (mentioned above, Section 5.3.2). It is found that all barrier heights and reaction energies converge within

$\pm$ 1 kcal mol^{-1} with respect to the number of solvent molecules when including up to five explicit HCN solvent molecules.

Furthermore, in **Paper II**, the solvation energy of HCN in itself is computed using a more refined approach: the cluster/continuum scheme proposed by Bryantsev et al. [232]. In this approach, the Gibbs energy of solvation of a species (ΔG_{solv}) is obtained by considering growing clusters of solvent molecules, and extrapolating to the bulk limit. Using DLPNO-CCSD(T) single-point corrected energies, the method provides a value of $\Delta G_{\text{solv}}(\text{HCN}) = -2.69$ kcal mol^{-1}, in good agreement with the experimental estimate of -3.52 kcal mol^{-1} .[2]

5.3.5 Error cancellation

Predicting the compounding effect of all sources of errors is virtually impossible. Fortunately, as computational chemistry typically focuses on relative energies, errors often partially cancel out.[3] This occurs because systematic errors (e.g., from level of theory or implicit solvation) tend to be similar in magnitude for similar systems. Therefore, relative energies between conformers and reaction energies can be fairly accurate [200]. However, this cancellation diminishes as the structural difference increases. For instance, barrier heights benefit less from this effect, as TSs may deviate significantly from minima in both geometry and electronic structure [200].

5.3.6 Kinetic modeling

The reaction rates derived evaluated from the Eyring equation (Eq. (4.27)) are approximate. Aside from the approximations of transition state theory, the exponential in the formula makes the rates extremely sensitive to errors in barrier estimates. This sensitivity propagates and compound over time, which can lead to erroneous predictions at long time scales [215].

One way to address the rate sensitivity issue is exemplified in **Paper II**, which adopts a **Monte Carlo** approach. In a Monte Carlo simulation, barrier heights ($\Delta G^{\ddagger}$) and reaction energies ($\Delta_r G$) are varied stochastically from normal distributions centered on the computed values. The simulation is then repeated thousands of times ($\sim$50,000 in **Paper II**), to obtain a distribution of concentration at each time step. This approach is based on the assumption that the errors on reaction parameters are random (normally distributed) and uncorrelated. Nevertheless, it provides a good estimate of the sensitivity of the results on the model error. A critical parameter is the standard deviation of such normal distributions, which reflects the expected model error. As to obtain a general estimate of this quantity is impractical, in **Paper II** a value of 1 kcal mol^{-1} is chosen to represent a best-case scenario.

[2] Obtained from the liquid density and the saturation vapor density; data from ref. [233].

[3] Not to be confused with fortuitous error compensation, which can lead to good results for the wrong reason [200].

CHAPTER 6

Thermodynamics of HCN-derived Molecules and Polymers

The chemical characterization of HCN-derived products presents enormous challenges. In this chapter, summary of **Paper I**, the following research question is addressed:

Q1. Which molecules and polymers can form spontaneously from HCN?

This question opens a series of related sub-questions, e.g., which species are plausible intermediates, and which do not react further? Furthermore: how does temperature affect the results?

Paper I elucidates the thermodynamic landscape of HCN-derived molecules and polymers in liquid water. The Gibbs energy (at 298 K) of seven molecules and twenty-four polymers proposed in the literature is computed. This set of polymers also includes one two-dimensional and one three-dimensional polymer, which I will not discuss here. All considered structures have a H:C:N ratio of 1:1:1, and the Gibbs energy are expressed per mole of HCN, which allows for an immediate comparison of relative stability across the whole map. These results, shown in Figure 6.1, represent a first step toward the exploration of the thermodynamics of HCN self-reaction. In what follows, I discuss the key takeaways of the study, which are strictly valid for the modeled conditions.

Key Takeaway 1: HCN is rich in chemical energy. Most examined molecules and polymers lie below HCN's Gibbs energy in liquid water at 298 K (Figure 6.1). This first observation is in line with the well-known high energy density of HCN. Thus, once initiated, the aqueous self-reaction of HCN is likely to give rise to plethora of different products. Among the considered products, the formation of only three polymers and only one species (the carbene isomer, **12**) is computed to be endergonic.

Key Takeaway 2: Some polymers can be ruled out; heterocycles are most favored. Both double (**5**) and single (**6**) ladder structures proposed by Völker [171] and Umemoto et al. [172], respectively, are computed to be two of the most energetic polymers. Thus, their formation from HCN is unlikely under the modeled conditions. Also the nitrogen-substituted polyacetylene (**7**), suggested by Mamajanov and Herzfeld as major product of a radical-initiated HCN polymerization [170], is predicted to be uphill from aqueous HCN. Nevertheless, formation of **7** may still occur following radical initiation producing the carbene isomer (**12**). Shifting the focus to the lowest-lying structures, it becomes evident the prevalence of highly conjugated five-member heterocycles. This result agrees with experimental evidence suggesting that the polymers are composed of highly unsaturated structures [148, 169].

Key Takeaway 3: Adenine is the lowest-lying species. Adenine (**10**) appears to be the thermodynamic sink of HCN self-reaction. The notorious low yield of adenine from HCN polymerization experiments ($\sim$0.1%-0.01% [137]) prompts three non-mutually exclusive interpretations: (i) there might exist other structures outside this sample which formation is more favorable, including beyond the $H_nC_nN_n$ chemical space; (ii) the combinatorial explosion of HCN self-reaction makes adenine statistically unlikely; finally, (iii) adenine formation might be kinetically hindered, a hypothesis that will be explored in the following chapter.

Key Takeaway 4: Most products are predicted to be uphill in DAMN polymerization. DAMN is computed to be one of the most stable structures, as only two molecules and eight polymers lie below it. The two molecular species are AICN (**9**) and adenine, in agreement with previous computational studies [185, 234, 235]. Several products have been suggested in the literature to form upon DAMN polymerization [136]. Albeit DAMN necessitates temperatures higher than 298 K to polymerize [154, 236, 237], the thermodynamic map can hint toward the feasibility of a pathway. For example, the suggested formation of the peptide-like polymer proposed by Matthews and co-workers [152], **20**, is computed to be thermodynamically disfavored. Conversely, the polyaminoimidazoles **31** and **32**, as suggested by Mamajanov and Herzfeld [236], are predicted to be likely products of DAMN polymerization.

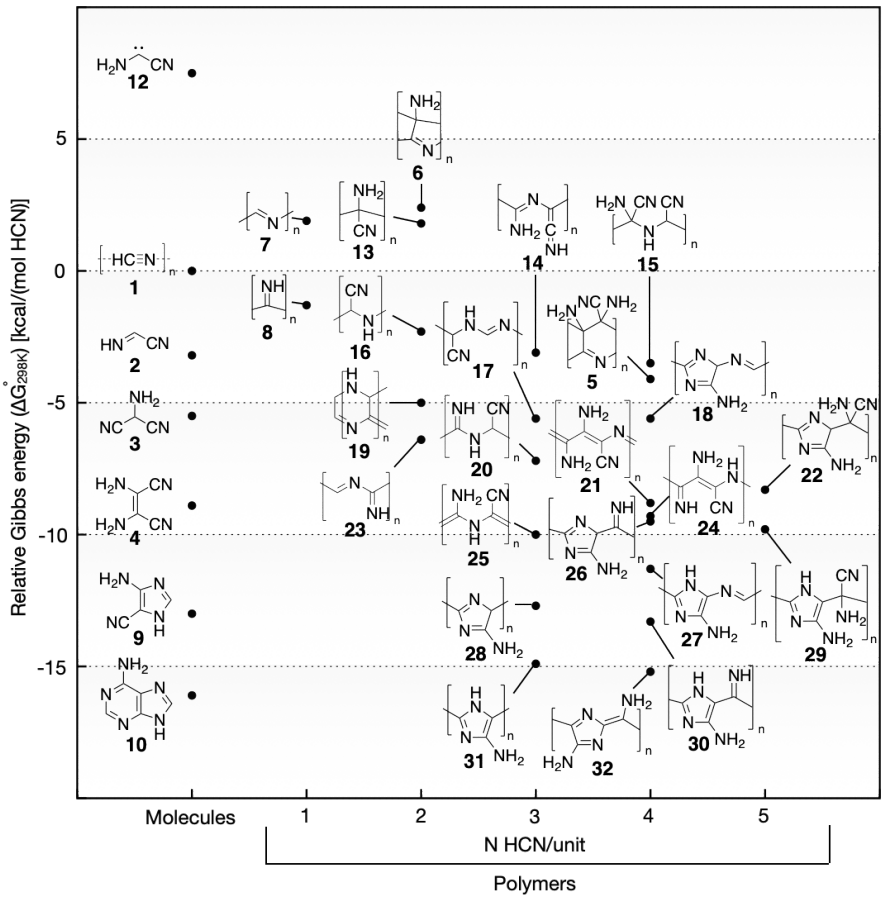

Figure 6.1: Standard Gibbs energies at 298 K in kcal (mol HCN)$^{-1}$ of the considered HCN-derived molecules and polymers. Figure reproduced from **Paper I**, licensed under CC BY 4.0.

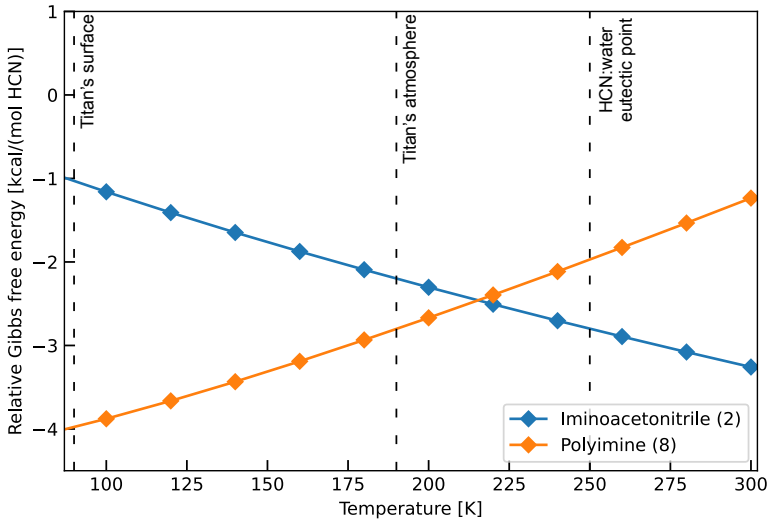

Figure 6.2: Predicted temperature dependence of Gibbs energies (in kcal (mol HCN)$^{-1}$) polyimine (**8**, in orange) and IAN (**2**, in blue). Polyimine formation is more favored at lower temperatures due to a reduced entropic penalty. Figure reproduced from **Paper I**, licensed under CC BY 4.0.

Key Takeaway 5: Polymerization is more thermodynamically favored in cold environments. Polymerization reactions are disfavored by entropy due to the loss of internal degrees of freedom. This energy penalty has (approximately) a linear dependence on the temperature, due to the $-T\Delta S$ term. This implies that entropic effects become gradually less relevant with decreasing temperatures, which likely changes the thermodynamic landscape. Figure 6.2 shows how the Gibbs energies (kcal (mol HCN)$^{-1}$) of IAN (**2**) and polyimine (**8**) vary with temperature. In particular, while IAN formation is predicted to be more favored at RT, a shift in the energetic order is computed for temperatures below ∼220 K. Although reaction rates become exponentially smaller as temperature decreases, eventually leading to the kinetic control regime, this simple example illustrates how temperature can affect prebiotic reaction chemistry.

CHAPTER 7

Abiotic Adenine Formation From HCN

"Messing with adenine is one of the main things life does."
– Hank Green

Adenine is arguably the most intriguing product of HCN self-reaction chemistry. This chapter is a summary of **Paper II**, which addresses the following research question:

Q2. Through which reaction mechanism(s) does adenine form from HCN?

The previously proposed routes to adenine from base-catalyzed HCN self-reaction are summarized in Figure 7.1. The arrows can represent multiple mechanisms.[1] Common to all routes is formation of the trimer AMN (**3**, gray path). Thereafter, three major routes can be distinguished, which are referred to as the **AMN path** (red), the **DAMN path** (blue and teal), and the **Voet and Schwartz (VS) path** (purple). Both the AMN and the DAMN path converge to AICN (**9**). The VS path, proposed by A. B. Voet and A. W. Schwartz in 1983 [183], branches from DAMN and leads to 8-cyano-adenine (**11**), which is speculated to convert to adenine upon hydrolysis (dashed arrow), via 8-carboxamide-adenine.

Figure 7.1 also denotes the novel reaction steps suggested in **Paper II**, marked as yellow glowing arrows. Two crucial links are proposed. Diverging from the DAMN path is an alternative route to adenine (in teal), via intermediate **37**, and that does not involve AICN as an intermediate. This newly proposed branch of the DAMN path will be referred to as "revised", whereas the previously known one will be denoted as "conventional". The second

[1] Most HCN addition reactions involve imine-enamine tautomerization. Owing to the fast and favored conversion to enamine forms in the modeled conditions, imine forms are omitted from the discussion. The only exception is AISN (**4'**), which actively partakes in a reaction.

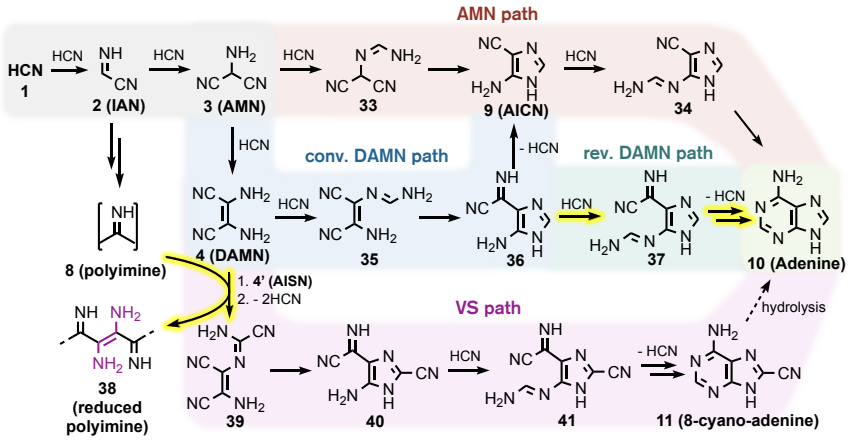

Figure 7.1: Summary of the base-catalyzed chemical reaction network from HCN to adenine in a HCN-rich environment. Arrows can represent multiple steps. Arrows highlighted in yellow represent the newly proposed steps from **Paper II**. Figure reproduced from **Paper II**.

newly suggested steps concern the onset of the VS path. Voet and Schwartz speculated that condensation of DAMN and its tautomer, AISN, yield product **39** [183]. However, they did not specify any mechanistic detail. In **Paper II**, a kinetically viable route from DAMN to **39** has been proposed.

All reactions are modeled in liquid HCN at 278 K. Such a temperature is chosen to allow comparison with previous computational studies by Sandström and Rahm [139, 159], as well as to coincide with HCN polymerization experiments reported in the literature [170]. A kinetic modeling using Monte Carlo sampling is also employed to study how the considered routes compete, as well as to predict and rationalize the experimentally observed adenine yields. These simulations span up to 30 thousand years (ky), offering insight on the long-term chemical evolution of the reaction network. Importantly, hydrolysis has not been considered in the kinetic model, as such reaction is not compatible with the modeled conditions. This means that the VS path terminates at 8-cyano-adenine (**11**). However, as it is also interesting to analyze the effect of the VS path on adenine yield, a hydrolysis step can be assumed. In this case, that I will clearly denoted as "assuming fast hydrolysis", the conversion of **11** to adenine is assumed to be rapid and quantitative. In the following, I summarize the key takeaways of **Paper II**.

Key Takeaway 1: DAMN formation is predicted to be fast and favored. The kinetics and thermodynamics of the HCN self-reaction pathway up to DAMN is shown in Figures 7.2-7.3. Also displayed is the onset of the AMN path (red line), compared to DAMN formation (blue, left). The initiation (24.0 kcal mol^{-1}) is predicted to be fairly slow at the considered temperature,

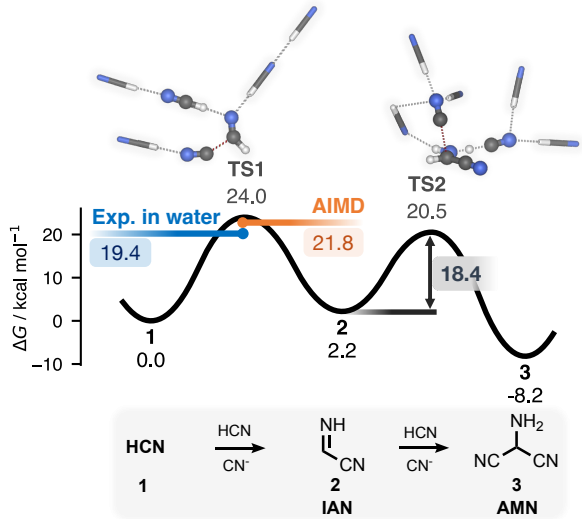

Figure 7.2: Reaction Gibbs energy profile (in kcal mol^{-1} relative to HCN) of AMN formation via IAN. The barrier of the first step is compared kinetic experiments in diluted aqueous solutions [149] (blue) and with previous computational studies employing AIMD (orange) [139]. The gray box displays the barrier height of **TS2**. The TS structures are shown on top of their respective labels. Explicit HCN solvent molecules are represented as sticks. Figure reproduced from **Paper II**.

and represents the rate-determining step for HCN tetramerization. In fact, once IAN forms, DAMN is predicted to form extremely rapidly: an activation barrier of 20.5 kcal mol^{-1} is overcome in the order of minutes at 278 K, assuming first-order kinetics. Furthermore, the conversion to the tetramer (**4**) from HCN is computed to be favored by *ca.* 20 kcal mol^{-1}. These values clearly translate into a fast accumulation of DAMN in the reaction medium, as shown in Figure 7.4. In the kinetic simulations, DAMN becomes the most abundant species in a time scale of days. These predictions are supported by several experimental and computational results [149, 164], and point towards DAMN being a key intermediate in the HCN self-reaction chemistry.

Key Takeaway 2: The VS path is suggested to involve a redox step. Figure 7.5 shows the beginning of the VS path, up to product **39**, as proposed in **Paper II**. While the first step is the condensation of DAMN (**4**) and AISN (**4'**), the fast and favorable formation of the former must be taken into account to avoid discussing artificially low barrier heights. Thus, the initiation of the VS path possesses an activation energy of approximately 28 kcal mol^{-1}. However, the actual bottleneck of the VS path is represented by **TS17** (29 kcal mol^{-1}), which is also the most noteworthy step. In this reaction, intermediate **43**

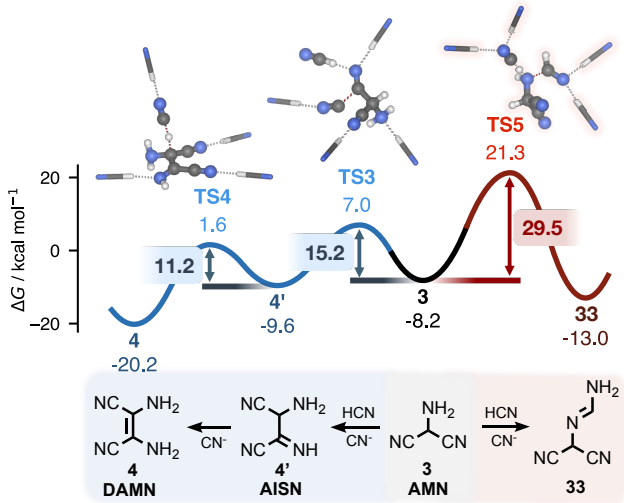

Figure 7.3: Gibbs energy profile (in kcal mol^{-1} relative to HCN) of the bifurcation into the AMN path and the DAMN paths, from AMN (**3**). Colored boxes display the barrier heights. The TS structures are shown on top of their respective labels. Explicit HCN solvent molecules are represented as sticks. Figure adapted from **Paper II**.

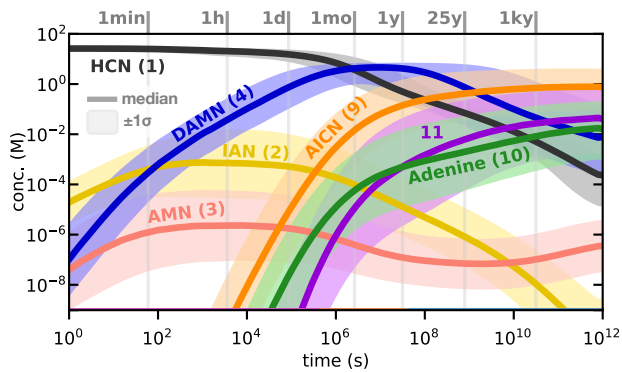

Figure 7.4: Time evolution of concentrations for key species (log-log scale) from the kinetic modeling based on Monte Carlo sampling (∼50,000 simulations). Solid colored lines indicate the median trajectories, while shaded bands represent the 68.3% of the simulations. Figure adapted from **Paper II**.

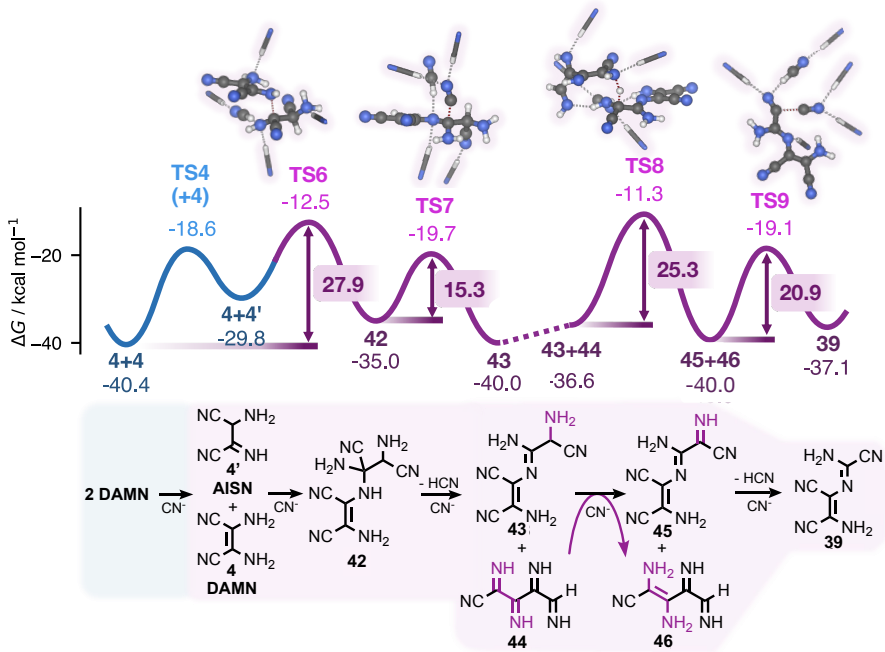

Figure 7.5: Gibbs energy profile (in kcal mol^{-1} relative to HCN) of the beginning of the VS path, until product **39**. Colored boxes display the barrier heights. The TS structures are shown on top of their respective labels. Explicit HCN solvent molecules are represented as sticks. Figure adapted from **Paper II**.

undergoes to a redox reaction, which involves a polyimine fragment (**44**) as an oxidizing agent. In particular, an amino group of **43** is converted into an imino group by mean of an hydride transfer, similarly to the reaction mediated by the Flavin Adenine Dinucleotide (FAD) coenzyme [238].[2] Polyimine might not be the sole oxidant: other HCN-derived products containing suitable functional groups may mediate such a reaction. While this speculation concurs with a previous suggestion that redox processes occur during HCN polymerization [239], this proposed mechanism and the results related to the VS path should be considered tentative.

Key Takeaway 3: All considered routes are in close kinetic competition. As seen from the Gibbs energy profiles (shown in full in **Paper II**, Figures 1-5), the four considered routes to adenine are characterized by barriers of similar magnitude, between 26-29 kcal mol^{-1}.[3] These sizable barriers support the kinetic hindrance as a plausible explanation for the low observed adenine yields. Furthermore, they also underline the importance of accurate chemical

[2] In fact, this was the inspiration for considering this step.
[3] Within the expected model error.

modeling, as any small error can drastically alter the conclusions. For example, the computed barrier for **TS5** with only a single explicit solvent HCN molecule results to be 3 kcal mol^{-1} smaller than the converged value with respect to explicit solvation (see Section 5.3.4). This difference would dramatically change the results from the kinetic simulations, as the kinetic bottleneck of the AMN path would be markedly smaller.

Key Takeaway 4: The main route to adenine is strongly time-dependent. The results demonstrate the kinetic competition between the computed pathways. Plotted in Figure 7.6a are the contributions to AICN rate of formation from the AMN path (from **33**, in red) and the conventional DAMN path (from **36**, in blue); similarly, Figure 7.6b shows the contributions to adenine rate of formation from the three pathways: from AICN, from the revised DAMN path, and from the VS path, thus assuming fast hydrolysis. Taken together, these the plots delineate the mass flow. Notably, no single route to adenine establishes as governing; instead, all four routes occupy a certain time window in which they contribute the most to adenine formation. Four time frames can be distinguished:

- Less than 1 minute: the AMN path appears to be the primary contributor via intermediates **33** and **9**, although the concentration of adenine is predicted not to be detectable.
- Less than ∼1 month: as DAMN rapidly accumulates, the predominant route shifts to the revised DAMN path (via **37**).
- Between ∼months and ∼years: DAMN concentration reaches a near steady-state, and the contribution from the VS path becomes significant.
- Beyond ∼25 years: the conventional DAMN path establishes as the main route.

Key Takeaway 5: The predicted adenine yields are compatible with experimental data. Table 7.1 summarizes the median of the predicted adenine yield (mol % from the initial HCN) at various times. Despite the inherent approximations of the model, these order of magnitudes are in line with adenine yields reported in the literature. It should be noted, however, that this comparison should be taken with caution, as the modeled reaction conditions, such temperature and employed solvent, vary wildly across the numerous experiments. Nevertheless, some parallels are worth noting. For example, in their excellent work, Borquez et al. [174][4] extrapolated the yield of adenine from the unhydrolyzed supernatant at infinite time, and reported a range of ~ 0.04–0.16%. This value is in line with the predicted $\sim 0.3\%$ yield at 30 ky (central column in Table 7.1). Furthermore, the contribution from the VS path, assuming fast hydrolysis (right column), is comparable with the

[4]Including Stanley L. Miller!

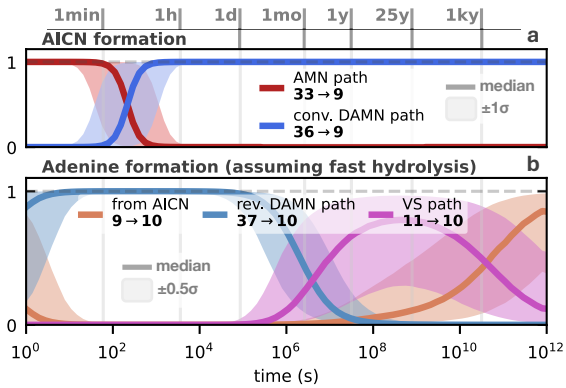

Figure 7.6: (a) Fractional contribution to AICN (**9**) rate formation from intermediates **33** (red, AMN pathway) and **36** (blue, conventional DAMN pathway). (b) Normalized contributions to the formation rate of adenine (**10**), assuming fast hydrolysis. Solid colored lines represent the median, while bands encompass 38.3% of the simulations. The pathways compared are: via AICN (**9**, orange), the revised DAMN path (**37**, blue), and the VS path (**11**, purple). Figure adapted from **Paper II**.

observed increase in adenine yield upon hydrolysis of the soluble fraction, where precursors as 8-cyano-adenine might be present. For example, in the one-of-a-kind experiment where diluted NH_4CN solutions were kept for 27 years at −78 °C, the adenine yield passed from 0.00016% to 0.029% before and after acid hydrolysis.

	Median adenine yield (mol %)	
Time	no hydrolysis	assuming fast hydrolysis
1 week	0.00009%	0.00009%
1 month	0.001%	0.001%
1 year	0.01%	0.02%
25 years	0.04%	0.2%
30 ky	0.3%	1.3%

Table 7.1: Median adenine yields from the kinetic modeling at different times (mol % of initial HCN at 278 K). Center: form the unhydrolyzed medium. Right: assuming fast hydrolysis. Table adapted from **Paper II**.

Chapter 8

HCN Crystal Surfaces May Spark Prebiotic Chemistry

I now return to cold astrochemical environments, where HCN may be found in the solid-state. One such setting is Titan. Titan's atmosphere is a complex environment with an extremely rich chemical inventory. Although the cryogenic temperatures hamper standard thermal chemistry, the atmosphere is teeming with reactions driven by UV light and cosmic particles. Importantly, Titan's atmosphere to host large quantities of HCN ice [95]. Understanding how the HCN crystals may affect the chemistry of Titan's atmosphere and, possibly, surface, is worth exploring. Therefore, owing to the intriguing though underexplored solid-state properties of HCN, this chapter addresses the following research question:

Q3. Can HCN crystal surfaces promote prebiotic chemistry in cold environments?

This chapter serves as a summary of **Papers III–IV**, which focus on the physicochemical properties of HCN crystal surfaces. As the context is cryogenic environments, the computational modeling concerns the low-temperature *Imm2* phase of HCN, shown in Figure 8.1a, which is likely to persist under these conditions. Quantitative conclusions therefore pertain to such temperatures ($T < 170\,\mathrm{K}$).

Crystal morphology and electric fields

The peculiar phenomena observed during HCN crystallization experiments – such as glow and electrical breakdowns – plausibly derive from the polarity of its solid-state structure, which creates an internal electric field that manifests at the polar termination. Thus, determining the extent to which polar surfaces are exposed is a crucial first step toward assessing the role of solid HCN in the chemistry of Titan.

In **Paper III**, surface energies (γ_{hkl}) are computed using slab models.[1] These energies are then used to predict the equilibrium crystal shape by means of Wulff constructions (see Section 4.3.4). Finally, electric fields and isomerization mechanisms are evaluated by means of a linear chain model (see Section 5.1.2).

Key Takeaway 1: Large surface energy anisotropy drives growth into needles. The HCN crystal surfaces show a stark contrast in their relative stability. In particular, surface energies of polar facets are predicted to be significantly higher than those of non-polar surfaces.[2]. The computed range is 66 – 1650 mJ m^{-2}, with (001), the most polar, being the most energetic. These energies drive HCN single crystals to grow as an elongated shape (Figure 8.1b-c) with an aspect ratio of about 1:26. In this needle-like structure, non-polar surfaces occupy the vast majority ($\approx 93\%$) of the surface area, whereas polar surfaces are exposed only near the tips. This finding not only is in line with the structures seen in experiments (Figure 3.2), but also offers a likely explanation to the "cobweb" framework: as the two terminations are oppositely charged, electrostatic attraction likely pulls the polar tips together, causing them to link. Therefore, these polar facets likely remain hidden, but might be transiently exposed following crystal fracture. This also suggests an explanation for the electrical breakdowns observed in experiments [5]: short exposure of polar surfaces might give rise to strong electrostatic forces between the disconnected tips.

Key Takeaway 2: Substantial electric fields are exerted at polar surfaces. When polar surfaces are exposed, the electrostatic potential buildup across the crystal remains uncompensated. Estimates from the linear chains suggest that these fields are enhanced by cooperativity, in line with previous computational studies [66, 68]. In particular, the electric field at both surfaces is predicted to be $\sim$20% higher than that of a single HCN molecule, and converges to a value of 1.10 V/Å at the H-end, and of 1.25 V/Å at the N-end (Figure 8.2). These magnitudes could potentially drive chemical transformations [241]. However, it should be noted that calculations on extended slab models show a dampening of these fields, especially at the H-end, likely due to charge screening effects stemming from metallicity, as I will discuss later in this chapter.

Furthermore, these fields are predicted to alter the chemical nature of the surface: the proton affinity (PA) of the terminal HCN at the N-end increases by $\sim$20% and reaches nearly the same value of that of ammonia; at the opposite side, the HCN gas-phase acidity becomes comparable of that of the strong

[1] Due to symmetry, (hkl) and $(hk\bar{l})$ are different and represent the two opposite facets. In **Paper III**, for simplicity, γ_{hkl} is computed as the average between γ_{hkl} and $\gamma_{hk\bar{l}}$ (see Section 4.3.4). In this chapter, I will omit this distinction.

[2] As the polar axis is the c-axis, a surface is polar if $l \neq 0$. Notably, γ_{hkl} scales linearly with the "degree of polarity" $l/\sqrt{h^2+k^2+l^2}$, the cosine of the angle between the c-axis and the surface normal, which measures the density of exposed N- and H-ends.

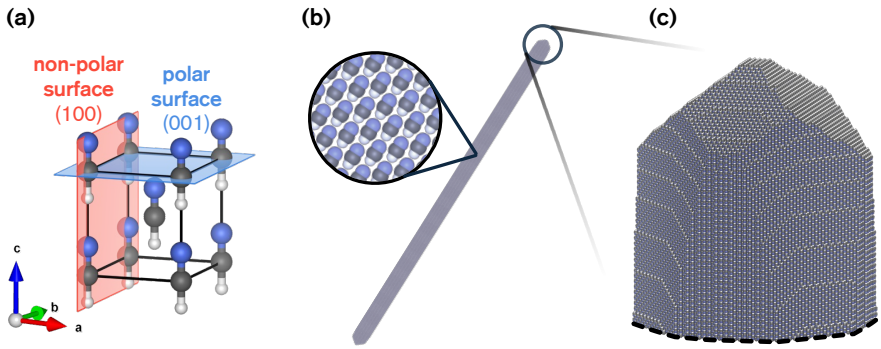

Figure 8.1: (a) Example of Miller planes of a nonpolar (100) (red) and a polar (001) (blue) surface of HCN *Imm*2 crystal. (b) Predicted equilibrium morphology of the HCN single crystal. The Wulff construction is composed of 10^7 atoms; (c) zoomed-in tip of the same crystal, at the H-termination. The majority of the tip surface area is composed of slanted, less polar facets, while highly polar facets are exposed at the very top. Figure adapted from **Paper III**, licensed under CC BY 4.0.

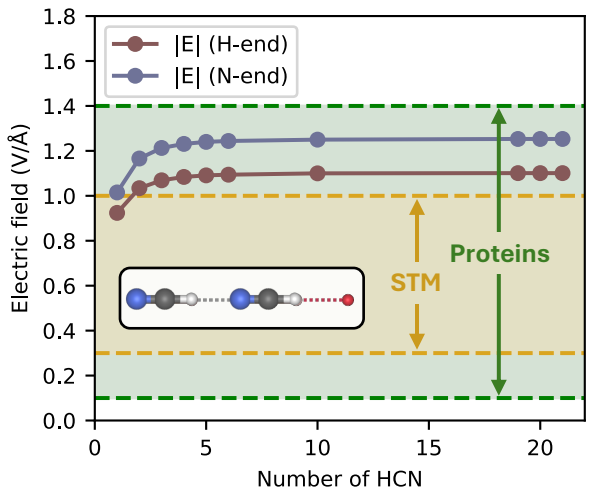

Figure 8.2: Computed electric field strengths from growing linear chains. The electrostatic potential is evaluated at a distance of 1.9 Å from the H-end (brown circles) and N-end (pale blue circles), reflecting a typical hydrogen bond distance. The inset depicts the measurement geometry relative to the H-end. These fields are compared to those inside proteins (green) and scanning tunneling microscopy (STM, yellow), which are known electrostatic catalysts [240]. Figure adapted from **Paper III**, licensed under CC BY 4.0.

acid HBr. This result provides the basis to rationalize the following predicted mechanisms.

Key Takeaway 3: Polar surfaces can drive the isomerization to HNC. A remarkable consequence of these electric fields is the potential to enable chemical reactions. As shown in Figure 8.3, when an ion is present at a terminal position, the catalyzing chain induces a proton transfer that effectively shifts the charge inward, giving a terminal HNC. This process, likely driven by electrostatics, is computed to be near-barrierless and energetically favorable with chains of four or more HCN units. This is in direct contrast to the non-catalyzed mechanisms, in which it is the reversed process (from HNC to HCN) that is predicted to be barrierless and favored.

The results for these two mechanisms offer a tentative explanation for part of the non-equilibrium HNC abundance in Titan's atmosphere [102]. As illustrated in Figure 8.3 and detailed in **Paper III**, the isomerization mechanism could become self-propagating, leading to continuous surface ablation, provided the existence of a process that liberates HNC. Given the computed binding energy of HNC on a HCN chain ($\approx$18 kcal mol^{-1}), thermal desorption may become kinetically viable only at the upper temperature limits observed in Titan's atmosphere (180-190 K). Alternatively, the mechanism could be triggered by UV photons or cosmic rays, which are known to penetrate the atmosphere, reaching lower attitudes [242].

Crucially, the plausibility of HNC formation depends on the actual exposure of these polar facets. However, modeling this accurately under realistic conditions is a non-trivial task. A major challenge concerns assessing the stability of these surfaces, which requires modeling phenomena as surface reconstruction or passivation. Another problem is determining the likelihood for crystalline HCN to grow in Titan's atmosphere. Whether single crystals or amorphous ices are predominant remains unresolved.

Electronic structure of polar surfaces

The polar nature of the HCN crystal is predicted to fundamentally alter its electronic structure. In this section, I summarize the key takeaways of **Paper IV**. For simplicity, this section focuses on the most polar facets (001) only, while noting that similar results are predicted for other polar surfaces as well.

Key Takeaway 4: Polar surfaces are predicted to exhibit localized metallic states. In the bulk, solid HCN is a typical organic crystal: an electronic insulator with a large band gap ($\approx$7 eV), a feature that is retained non-polar surfaces. In striking contrast, polar surfaces are predicted to exhibit a metallic character. Figure 8.4 shows the band structure of a (001) surface composed of 16 layers of HCN (relaxed 8-unit-cell-thick slab). It can be seen that some of the high-energy bands belonging to the H-end termination (from π_x^* and π_y^*

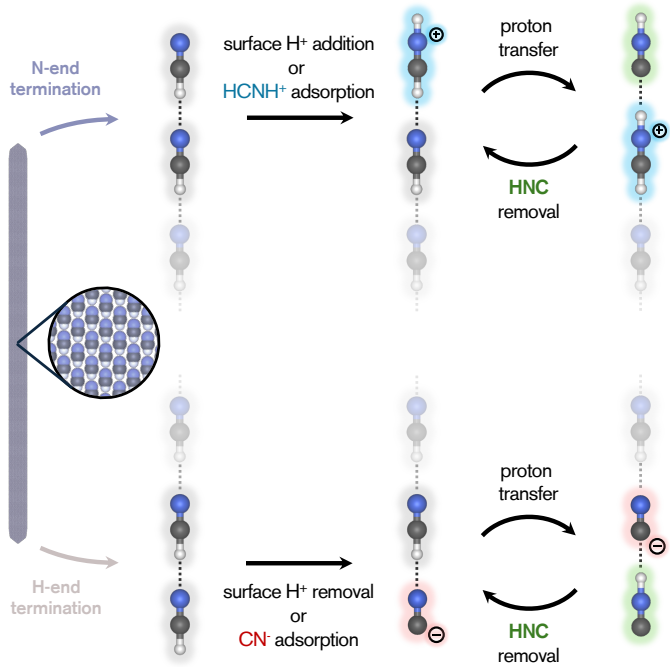

Figure 8.3: Schematic representation of the proposed surface-catalyzed isomerization mechanisms. Top: Protonation of the N-end or adsorption of $HCNH^+$ (light blue). Bottom: Deprotonation of the H-end or adsorption of CN^- (red). Both mechanisms are near-barrierless and favorable, and generate a terminal HNC (green) following proton transfer. The process can iterate upon HNC removal. Figure adapted from **Paper III**, licensed under CC BY 4.0.

MOs, see Figure 3.1) become partially occupied around the X- and Y-points. At the same time, the valence bands of the N-end termination (from π_x and π_y MOs) become partially unoccupied around the Y- and Γ-points. In essence, calculations predict that the N-end facet becomes oxidized (p-doped), whereas the H-end facet becomes reduced (n-doped). These metallic states are localized on their respective surfaces, as seen from the insets in Figure 8.4. The system therefore undergoes an electronic reconstruction to avoid a "polar catastrophe" [243, 244]. However, a question remains: how does this reconstruction occur?

To rationalize the emergence of metallicity, it is beneficial to take a step back and analyze the simplest case: a single layer of HCN in a rectangular lattice. This system an electronic insulator, with a band gap of 6 eV. Figure 8.5 shows the three uppermost valence bands of such a simple system, projected onto the atomic orbitals of all atoms. Overlaid are the interacting MOs of HCN, shown from above: σ (blue), π_x (red), and π_y (green, see Figure 3.1b). The shape and dispersion of the bands can be understood based on in-phase and

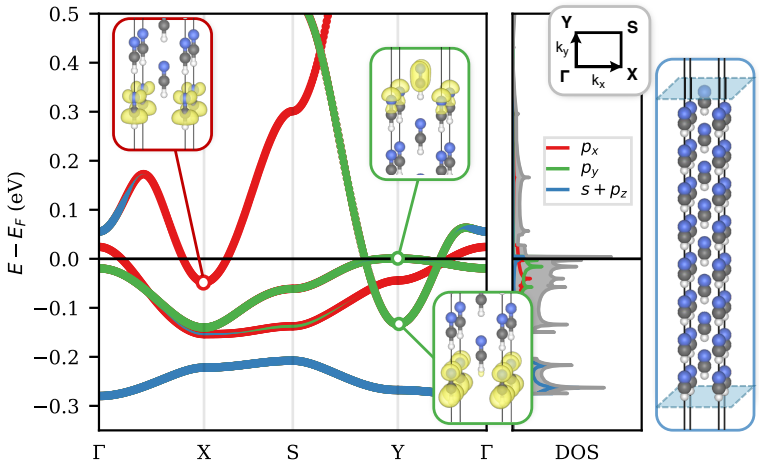

Figure 8.4: Electronic band structure of an optimized 8-layer HCN slab (unit cell shown on the right). The bands are color-coded by orbital character, indicating contributions from p_x (red), p_y (green), and $s + p_z$ (blue) atomic orbitals. Overlaid insets depict the projected density of states (PDOS) at the marked point. The k-point path is shown in top right panel. Figure adapted from **Paper IV**.

out-of-phase interactions. For example, as the a-axis is shorter than the b-axis, the π_x–π_x interactions are stronger than π_y–π_y ones, and the dispersions of the π_x band results the largest. The smallest dispersion is exhibited by the σ band, due to the small side-to-side overlap.

The inclusion of additional layers introduces head-to-tail interactions. As they are primary electrostatic in nature, their main effect is to rigidly shift the energy levels on opposite sides of the crystal. In particular, the occupied states at the (negative) N-end are destabilized and pushed up in energy, whereas the virtual states at the (positive) H-end are stabilized and pulled down. Already with three layers of HCN, both these bands cross the Fermi level, creating metallic states (shown in **Paper IV**).

The formation of metallic states can therefore be understood on the basis of a synergy between head-to-tail and side-by-side interactions in the lattice. While the former power the polarization catastrophe, the latter allow for a redistribution of the charges. Furthermore, as visible from the insets in Figure 8.4, small σ–π and π–π mixing occurs between the MOs of the outermost and the second outermost layer. These interactions create an additive effect to the "pushing" and "pulling" by the aligned dipoles, as they introduce a small perturbation in the states. For example, strong σ–π_y mixing at the N-end in the Y-point further destabilizes the upper state, allowing it to cross the Fermi level.

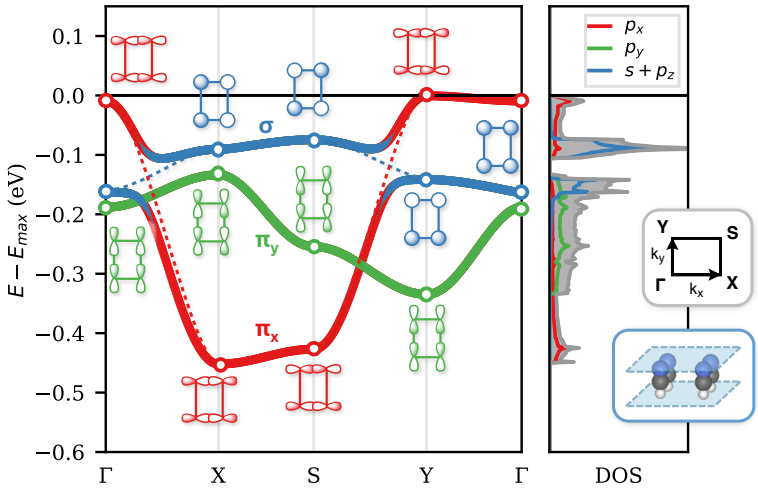

Figure 8.5: Electronic band structure and density of states (DOS) of side-interacting HCN molecules in a rectangular lattice (unit cell shown on the bottom right). The bands are color-coded by orbital character, indicating contributions from p_x (red), p_y (green), and $s + p_z$ (blue) atomic orbitals. The k-point path is shown in the panel on the right. Figure adapted from **Paper IV**.

Given these tentative yet exciting results, it is worth to explore the potential role that these surfaces might play in realistic astrochemical environments. One example is studying adsorption by water. Not only is water more ubiquitous than HCN in the Universe [245], but the interaction between HCN and water is of prebiotic relevance, for instance for the synthesis of amino acids.

Key Takeaway 5: Polar HCN surfaces can induce metallicity in adsorbed molecules. Preliminary electronic structure calculations from **Paper IV** suggest that the electronic structure of water molecules is altered profoundly when adsorbed to the (001) surface. Figure 8.6 shows the electronic bands projected onto the oxygen atoms (red), carbon and nitrogen $p_x + p_y$ orbitals (green) and carbon and nitrogen $s + p_z$ atomic orbitals (blue). Crossing the Fermi level are states that are mainly localized on the water molecules. An electronic reconstruction similar to the naked system occurs: the band arising from the lowest unoccupied MO (LUMO) of water (top two insets) is pulled down in energy, whereas the band originating from its highest occupied MO (HOMO) – the lone pair of the oxygen – is pushed right on the Fermi level. Notably, this latter band is essentially flat on the Fermi level, which contributes to a large peak in the density of states (DOS). This result indicates that the metallic states effectively shift from the HCN facets to the chemisorbed water layers. Hence, water molecules at the N-end are predicted to be reduced, while those at the H-end are predicted to be oxidized.

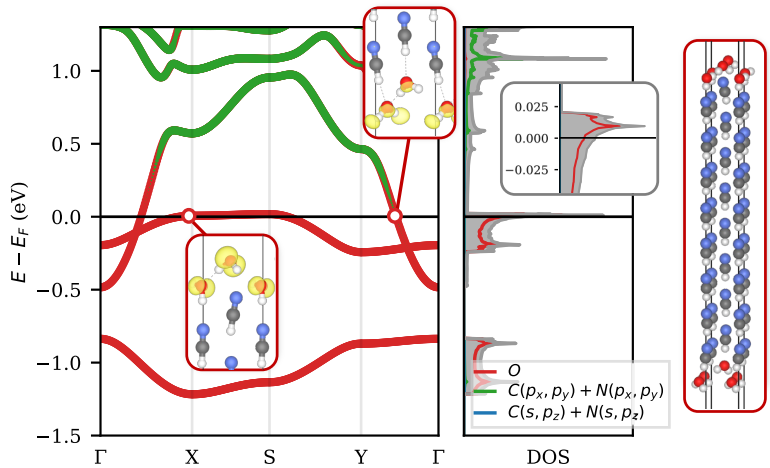

Figure 8.6: Electronic band structure and density of states (DOS) of a relaxed 16-layer (8 unit cells) HCN slab passivated by a single layer of water molecules (unit cell shown on the right). The bands are color-coded by atomic orbital characters, indicating contributions from all orbitals of oxygen atoms (red), carbon and nitrogen $p_x + p_y$ (green), and carbon and nitrogen $s + p_z$ (blue). Overlaid insets depict the PDOS at the marked point. The k-point path is the same as in Figure 8.4. Figure adapted from **Paper IV**.

Preliminary hybrid-DFT (HSE06) calculations yield the same electronic structure. The major computational limitation for this system is the size of the cell. A thick slab is required to describe bulk behavior in the middle, while a large supercell is crucial to permit water molecules to chemisorb more realistically, and possibly alter the geometry of the outermost HCN layer. Furthermore, an extensive sampling of the water molecules on the surfaces is necessary to minimize spurious findings.[3] Nevertheless, provided that the surface does not structurally rearrange rapidly, these localized metallic states are likely to interact with the environment and induce highly enhanced reactivity on chemisorbed molecules, such as water molecules.

[3]These investigations are currently being made, and they appear to be in line the results presented here.

CHAPTER 9

Conclusion and Outlook

In this thesis, I explore how hydrogen cyanide (HCN), a deadly poison to most life on Earth, could be the very seed of its origin. Through the four papers presented, I investigate the unique reactivity of HCN in the context of prebiotic chemistry. While truly captivating, this journey was also humbling, owing to the sheer complexity of HCN chemistry and the unavoidable compromises needed to model it. In this chapter, I summarize the key results discussed in this thesis, highlight the open (and newly-opened) questions, and suggest directions to guide future work.

HCN self-reaction is known to cause a plenitude of different molecules and polymers. The thermodynamic landscape investigated in **Paper I** reveals the elevated energy of this simple nitrile, and sheds light on some of the possible structural composition of the HCN-derived products. While most considered structures are predicted to be thermodynamically favored to form from aqueous HCN polymerization, the results indicate that some polymers proposed in the literature are unlikely to form. Conversely, highly conjugated polymers containing imidazole rings are computed to be among the lowest-lying species, in line with experimental observation of highly unsaturated structures [148, 169].

The thermodynamic sink of the investigated structures is predicted to be adenine. This nucleobase is the subject of **Paper II**, which addresses a long-standing question in prebiotic chemistry: what is the mechanism of adenine formation from HCN? Several proposed pathways are studied using state-of-the-art quantum chemistry and modeled in a liquid HCN reference environment. Kinetic simulations coupled with uncertainty quantification suggest that the governing mechanism for adenine formation varies with time. These results offer new insight into several experimental observations, such as the rapid formation of DAMN, as well as the low adenine yields and their increase upon hydrolysis.

The second part of the thesis is devoted to the solid-state properties of HCN, which are as compelling as they are underexplored. **Paper III** begin

to fill this knowledge gap by investigating the equilibrium morphology and surface chemistry of the HCN crystal. The work explains why solid HCN crystallizes into elongated structures, whose terminations are predicted to be composed by highly energetic polar surfaces. These facets can exert strong electric fields, which are suggested to facilitate chemical transformations on chemisorbed molecules.

The exceptional properties of these facets are further explored in **Paper IV**, which presents an analysis of their electronic structure. In striking contrast to the bulk phase, the polar surfaces are predicted to be metallic. The emergence of these localized metallic states is attributed to a synergistic effect of electrostatic and molecular orbital interactions, which drive an electronic reconstruction to avoid a polar catastrophe. This induces p-type and n-type doping at opposite sides of the crystal. Furthermore, water molecules chemisorbed to the surface appear to inherit this metallic character. The predicted absence of an electronic band gap on exposed polar facets of HCN crystals suggests the presence of unusual reactivity. Such potential enhanced reactivity is especially relevant for cryogenic environments, such as Saturn's moon Titan, where solid HCN is likely to persist.

The work presented in this thesis represents an important step toward better understanding the role of HCN reactivity in prebiotic chemistry and astrochemistry. It demonstrates that computational chemistry, when coupled with rigorous chemical modeling, can provide in-depth insights into the complex physics and chemistry of HCN. Furthermore, this thesis establishes both a robust methodology and a foundational reference upon which future research can be built.

The results presented in **Paper II** provide an opportunity to study and compare on equal footing the formation of other prebiotically relevant molecules. However, unlike adenine, many of these products feature oxygen-containing functional groups. Therefore, their study requires an investigation into the role of water. Experimental observations indicate that oxygen is incorporated into HCN-derived products during aqueous polymerization [168, 173], which suggests that water acts as an active reactant. A possible direction for future research is to kinetically compare the pathways computed in **Paper II** with the hydrolysis rates of HCN and its early self-reaction products. This comparison might reveal whether water participates at the onset of HCN self-reaction. Furthermore, such mechanistic insight may lay the ground for the exploration of other pathways, such to amino acids, and to the nucleobase guanine. Guanine is structurally similar to adenine but possesses a carbonyl group. Several studies reported the detection of guanine following hydrolysis of the HCN-derived products, with yields that are one order of magnitude lower than those of adenine [137]. These studies could help delineate the mechanisms leading to the diverse range of HCN-derived molecules.

Finally, the results on solid-state HCN (**Papers III–IV**) introduce several compelling research questions, particularly regarding the predicted metallic states on polar surfaces. A critical aspect that remains to be addressed is

the stability of these facets: the potential for these surfaces to drive chemical transformations depends on their lifetime. To this end, *ab initio* molecular dynamics (AIMD) represents a valid approach to predict their stability against surface reconstruction. However, given the complexity in modeling this phenomenon, verifying the existence of these high-energy facets is ultimately an experimental question.

Assuming these surfaces persist long enough to interact with the environment, the next challenge may be to investigate how their metallicity affects chemical reactivity. A first step could be to study the self-reactivity of HCN in this environment. For instance, investigating whether polar surfaces reduce the barrier of the rate-determining initiation step may provide insight into the reaction viability at cryogenic temperatures. This catalytic potential is arguably best explored via AIMD simulations. Investigating these phenomena could significantly improve our understanding of chemical evolution in cold environments, such as on Titan.

In conclusion, I hope that the contributions made by this thesis inspire researchers to explore the fascinating world of hydrogen cyanide and take us one step closer to understanding of how chemistry brought life to our Universe.

Bibliography

(1) Cleaves, H. J. *Evolution: Education and Outreach* **2012**, *5*, 342–360 (cit. on pp. 1, 7).

(2) Derrick, P. *International Journal of Microbiology Research and Reviews* **2021**, *10*, 1–1 (cit. on p. 2).

(3) Oró, J. *Biochemical and Biophysical Research Communications* **1960**, *2*, 407–412 (cit. on pp. 2, 12, 13).

(4) Smyth, C. P.; McAlpine, K. B. *Journal of the American Chemical Society* **1934**, *56*, 1697–1700 (cit. on p. 2).

(5) Mozhaev, P. S.; Kiryukhin, D. P.; Kichigina, G. A.; Barkalov, I. M. *Mendeleev Communications* **1994**, *4*, 17–19 (cit. on pp. 2, 10, 60).

(6) NASA About Life Detection (cit. on p. 3).

(7) Rosing, M. T. *Science* **1999**, *283*, 674–676 (cit. on p. 4).

(8) Van Zuilen, M. A.; Lepland, A.; Arrhenius, G. *Nature* **2002**, *418*, 627–630 (cit. on p. 4).

(9) Bell, E. A.; Boehnke, P.; Harrison, T. M.; Mao, W. L. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 14518–14521 (cit. on p. 4).

(10) Dodd, M. S.; Papineau, D.; Grenne, T.; Slack, J. F.; Rittner, M.; Pirajno, F.; O'Neil, J.; Little, C. T. *Nature* **2017**, *543*, 60–64 (cit. on p. 4).

(11) Maas, R.; Kinny, P. D.; Williams, I. S.; Froude, D. O.; Compston, W. *Geochimica et Cosmochimica Acta* **1992**, *56*, 1281–1300 (cit. on p. 4).

(12) Wilde, S. A.; Valley, J. W.; Peck, W. H.; Graham, C. M. *Nature* **2001**, *409*, 175–178 (cit. on p. 4).

(13) Orgel, L. E. *Trends in Biochemical Sciences* **1998**, *23*, 491–495 (cit. on pp. 4, 7).

(14) Tian, F.; Toon, O. B.; Pavlov, A. A.; De Sterck, H. *Science* **2005**, *308*, 1014 (cit. on p. 4).

(15) Sossi, P. A.; Burnham, A. D.; Badro, J.; Lanzirotti, A.; Newville, M.; O'Neill, H. S. C. *Science Advances* **2020**, *6*, DOI: 10.1126/SCIADV.ABD1387 (cit. on p. 4).

(16) Mojzsis, S. J.; Harrison, T. M.; Pidgeon, R. T. *Nature* **2001**, *409*, 178–181 (cit. on p. 4).

(17) Cnossen, I.; Sanz-Forcada, J.; Favata, F.; Witasse, O.; Zegers, T.; Arnold, N. F. *Journal of Geophysical Research: Planets* **2007**, *112*, 2008 (cit. on p. 4).

(18) Canuto, V. M.; Levine, J. S.; Augustsson, T. R.; Imhoff, C. L. *Nature* **1982**, *296*, 816–820 (cit. on p. 4).

(19) Weiss, M. C.; Sousa, F. L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W. F. *Nature Microbiology* **2016**, *1*, 16116– (cit. on p. 4).

(20) Moody, E. R. et al. *Nature Ecology & Evolution* **2024**, *8*, 1654–1666 (cit. on p. 4).

(21) Oparin, A. I., *The Origin of Life (Proiskhozhdenie zhizny)*; Moskovskii rabochii: Moscow, 1924 (cit. on p. 5).

(22) Haldane, J. B. S. *Rationalist Annual* **1929**, *148*, 3–10 (cit. on p. 5).

(23) Miller, S. L. *Science* **1953**, *117*, 528–529 (cit. on p. 5).

(24) Do Nascimento Vieira, A.; Kleinermanns, K.; Martin, W. F.; Preiner, M. *FEBS Letters* **2020**, *594*, 2717–2733 (cit. on pp. 5, 6).

(25) Whitaker, D.; Powner, M. W. *Nature Reviews Chemistry* **2024**, *8*, 817–832 (cit. on p. 5).

(26) Huber, C.; Wächtershäuser, G. *Science* **1998**, *281*, 670–671 (cit. on p. 5).

(27) Lambert, J. F. *Origins of Life and Evolution of Biospheres* **2008**, *38*, 211–242 (cit. on p. 5).

(28) Hazen, R. M.; Sverjensky, D. A. *Cold Spring Harbor perspectives in biology* **2010**, *2*, DOI: 10.1101/CSHPERSPECT.A002162 (cit. on p. 5).

(29) Westall, F.; Brack, A.; Fairén, A. G.; Schulte, M. D. *Frontiers in Astronomy and Space Sciences* **2023**, *9*, 1095701 (cit. on p. 5).

(30) Orgel, L. E. *Critical Reviews in Biochemistry and Molecular Biology* **2004**, *39*, 99–123 (cit. on p. 5).

(31) Follmann, H.; Brownson, C. *Naturwissenschaften* **2009**, *96*, 1265–1292 (cit. on p. 5).

(32) Green, N. J.; Xu, J.; Sutherland, J. D. *Journal of the American Chemical Society* **2021**, *143*, 7219–7236 (cit. on p. 5).

(33) Gilbert, W. *Nature* **1986**, *319*, 618 (cit. on p. 6).

(34) Joyce, G. F. *Nature* **2002**, *418*, 214–221 (cit. on p. 6).

(35) Lazcano, A. *Origins of Life and Evolution of Biospheres* **2010**, *40*, 161–167 (cit. on p. 6).

(36) Bernhardt, H. S. *Biology Direct* **2012**, *7*, 23 (cit. on p. 6).

(37) Benner, S. A.; Kim, H. J.; Yang, Z. *Cold Spring Harbor Perspectives in Biology* **2012**, *4*, a003541 (cit. on p. 6).

(38) Martin, W.; Baross, J.; Kelley, D.; Russell, M. J. *Nature Reviews Microbiology* **2008**, *6*, 805–814 (cit. on p. 6).

(39) Sojo, V.; Herschy, B.; Whicher, A.; Camprubí, E.; Lane, N. *Astrobiology* **2016**, *16*, 181–197 (cit. on p. 6).

(40) Branscomb, E.; Russell, M. J. *BioEssays* **2018**, *40*, 1700179 (cit. on p. 6).

(41) Branscomb, E.; Russell, M. J. *BioEssays* **2018**, *40*, 1700182 (cit. on p. 6).

(42) Bada, J. L.; Lazcano, A. *Science* **2002**, *296*, 1982–1983 (cit. on p. 6).

(43) Kawamura, K. *International Journal of Astrobiology* **2004**, *3*, 301–309 (cit. on p. 6).

(44) Sutherland, J. D. *Nature Reviews Chemistry* **2017**, *1*, 0012– (cit. on p. 6).

(45) Hurowitz, J. A. et al. *Nature 2025 645:8080* **2025**, *645*, 332–340 (cit. on p. 6).

(46) De Mol, M. L. *Life* **2023**, *13*, DOI: 10.3390/LIFE13030675 (cit. on p. 6).

(47) Sagan, C.; Thompson, W. R.; Khare, B. N. *Accounts of Chemical Research* **1992**, *25*, 286–292 (cit. on p. 6).

(48) Fulchignoni, M. et al. *Nature* **2005**, *438*, 785–791 (cit. on pp. 6, 11).

(49) Vuitton, V.; Yelle, R. V.; Klippenstein, S. J.; Hörst, S. M.; Lavvas, P. *Icarus* **2019**, *324*, 120–197 (cit. on pp. 6, 12).

(50) Nixon, C. A. *ACS Earth and Space Chemistry* **2024**, *8*, 406–456 (cit. on pp. 6, 11).

(51) Schulze-Makuch, D.; Grinspoon, D. H. *https://home.liebertpub.com/ast* **2005**, *5*, 560–567 (cit. on p. 6).

(52) Raulin, F.; Owen, T. *Space Science Reviews* **2002**, *104*, 377–394 (cit. on p. 6).

(53) Kaiser, R. I.; Mebel, A. M. *Chemical Society Reviews* **2012**, *41*, 5490–5501 (cit. on p. 6).

(54) Sandström, H.; Rahm, M. *Science Advances* **2020**, *6*, eaax0272 (cit. on p. 6).

(55) Chyba, C.; Sagan, C. *Nature* **1992**, *355*, 125–132 (cit. on pp. 7, 12).

(56) Altwegg, K. et al. *Science Advances* **2016**, *2*, DOI: 10.1126/SCIADV.1600285 (cit. on p. 7).

(57) Kitadai, N.; Maruyama, S. *Geoscience Frontiers* **2018**, *9*, 1117–1153 (cit. on pp. 7, 17).

(58) Oba, Y.; Takano, Y.; Furukawa, Y.; Koga, T.; Glavin, D. P.; Dworkin, J. P.; Naraoka, H. *Nature Communications* **2022**, *13*, 1–10 (cit. on p. 7).

(59) Furukawa, Y.; Sunami, S.; Takano, Y.; Koga, T.; Hirakawa, Y.; Oba, Y.; Naraoka, H.; Saigusa, D.; Yoshikawa, T.; Tanaka, S.; Glavin, D. P.; Dworkin, J. P.; Connolly, H. C.; Lauretta, D. S. *Nature Geoscience* **2025**, *19*, 19–24 (cit. on p. 7).

(60) Mojarro, A.; Aponte, J. C.; Dworkin, J. P.; Elsila, J. E.; Glavin, D. P.; Connolly, H. C.; Lauretta, D. S. *Proceedings of the National Academy of Sciences of the United States of America* **2025**, *122*, e2512461122 (cit. on p. 7).

(61) Yadav, M.; Kumar, R.; Krishnamurthy, R. *Chem Rev* **2020**, *120*, 4766–4805 (cit. on p. 7).

(62) Bhattacharya, B. N.; Gordy, W. *Physical Review* **1960**, *119*, 144 (cit. on p. 9).

(63) Mettee, H. D. *Journal of Physical Chemistry* **1973**, *77*, 1762–1764 (cit. on p. 9).

(64) Rocher-Casterline, B. E.; Chng, L. C.; Mollner, A. K.; Reisler, H. *Journal of Chemical Physics* **2011**, *134*, 211101 (cit. on p. 9).

(65) Ruoff, R. S.; Emilsson, T.; Klots, T. D.; Chuang, C.; Gutowsky, H. S. *The Journal of Chemical Physics* **1988**, *89*, 138–148 (cit. on p. 9).

(66) Adrian-Scotto, M.; Vasilescu, D. *Journal of Molecular Structure: THEOCHEM* **2007**, *803*, 45–60 (cit. on pp. 9, 60).

(67) Góra, R. W.; Zaleśny, R.; Zawada, A.; Bartkowiak, W.; Skwara, B.; Papadopoulos, M. G.; Silva, D. L. *Journal of Physical Chemistry A* **2011**, *115*, 4691–4700 (cit. on p. 9).

(68) Brandão, I.; Rivelino, R.; Fonseca, T. L.; Castro, M. A. *Chemical Physics Letters* **2013**, *580*, 9–13 (cit. on pp. 9, 60).

(69) De Oliveira, P. M. M. C.; Silva, J. A.; Longo, R. L. *Journal of Molecular Modeling* **2017**, *23*, 1–10 (cit. on p. 9).

(70) Freitas, D. P.; Pansini, F. N.; Varandas, A. J. *Chemical Physics Letters* **2023**, *828*, 140734 (cit. on p. 9).

(71) Coates, G. E.; Coates, J. E. *Journal of the Chemical Society (Resumed)* **1944**, 77–81 (cit. on p. 9).

(72) Malmberg, C. G.; Maryott, A. A. *Journal of Research of the National Bureau of Standards* **1956**, *56* (cit. on p. 9).

(73) Costa Cabral, B. J. *Journal of Molecular Liquids* **2018**, *272*, 778–786 (cit. on p. 10).

(74) Dulmage, W. J.; Lipscomb, W. N *Acta Crystallographica* **1951**, *4*, 334 (cit. on p. 10).

(75) Pézolet, M.; Savoie, R. *Canadian Journal of Chemistry* **1011**, *47*, 3041–3048 (cit. on p. 10).

(76) Panas, I. *Chemical Physics Letters* **1992**, *194*, 239–246 (cit. on p. 10).

(77) Giauque, W. F.; Ruehrwein, R. A. *Journal of the American Chemical Society* **1939**, *61*, 2626–2633 (cit. on p. 10).

(78) Peng, J.; Zhang, S.; Refson, K.; Dove, M. T. *Journal of Physics: Condensed Matter* **2021**, *34*, DOI: 10.1088/1361-648X/ac3e1f (cit. on p. 10).

(79) Makhnev, V. Y.; Kyuberis, A. A.; Zobov, N. F.; Lodi, L.; Tennyson, J.; Polyansky, O. L. *Journal of Physical Chemistry A* **2018**, *122*, 1326–1343 (cit. on p. 10).

(80) Baiano, C.; Lupi, J.; Barone, V.; Tasinato, N. *Journal of Chemical Theory and Computation* **2022**, *18*, 3111–3121 (cit. on p. 10).

(81) Snyder, L. E.; Buhl, D.; Snyder, L. E.; Buhl, D. *ApJL* **1971**, *163*, L47 (cit. on p. 10).

(82) Rice, T. S.; Bergin, E. A.; Jørgensen, J. K.; Wampfler, S. F. *The Astrophysical Journal* **2018**, *866*, 156 (cit. on p. 10).

(83) Pizzarello, S. *The Astrophysical Journal Letters* **2012**, *754*, DOI: 10.1088/2041-8205/754/2/L27 (cit. on p. 10).

(84) Dello Russo, N.; Kawakita, H.; Vervack, R. J.; Weaver, H. A. *Icarus* **2016**, *278*, 301–332 (cit. on p. 11).

(85) Rodgers, S. D.; Charnley, S. B. *The Astrophysical Journal* **1998**, *501*, L227–L230 (cit. on p. 11).

(86) Huebner, W. F.; Snyder, L. E.; Buhl, D. *Icarus* **1974**, *23*, 580–584 (cit. on p. 11).

(87) Tokunaga, A. T.; Beck, S. C.; Geballe, T. R.; Lacy, J. H.; Serabyn, E. *Icarus* **1981**, *48*, 283–289 (cit. on p. 11).

(88) Lellouch, E.; Romani, P. N.; Rosenqvist, J. *Icarus* **1994**, *108*, 112–136 (cit. on p. 11).

(89) Lellouch, E. et al. *Icarus* **2017**, *286*, 289–307 (cit. on p. 11).

(90) Hanel, R. et al. *Science* **1981**, *212*, 192–200 (cit. on p. 11).

(91) Tanguy, L.; Bézard, B.; Marten, A.; Gautier, D.; Gérard, E.; Paubert, G.; Lecacheux, A. *Icarus* **1990**, *85*, 43–57 (cit. on p. 11).

(92) Peter, J. S.; Nordheim, T. A.; Hand, K. P. *Nature Astronomy* **2023**, *8*, 164–173 (cit. on p. 11).

(93) Mumma, M. J.; Charnley, S. B. *Annual Review of Astronomy and Astrophysics* **2011**, *49*, 471–524 (cit. on p. 11).

(94) Burgdorf, M.; Cruikshank, D. P.; Dalle Ore, C. M.; Sekiguchi, T.; Nakamura, R.; Orton, G.; Quirico, E.; Schmitt, B. *Astrophysical Journal Letters* **2010**, *718*, L53–L57 (cit. on p. 11).

(95) De Kok, R. J.; Teanby, N. A.; Maltagliati, L.; Irwin, P. G.; Vinatier, S. *Nature* **2014**, *514*, 65–67 (cit. on pp. 11, 12, 59).

(96) Gerakines, P. A.; Yarnall, Y. Y.; Hudson, R. L. *Monthly Notices of the Royal Astronomical Society* **2021**, *509*, 3515–3522 (cit. on p. 11).

(97) Bergner, J. B.; Rajappan, M.; Öberg, K. I. *The Astrophysical Journal* **2022**, *933*, 206 (cit. on p. 11).

(98) Hirota, T.; Yamamoto, S.; Mikami, H.; Ohishi, M. *The Astrophysical Journal* **1998**, *503*, 717–728 (cit. on p. 11).

(99) Graninger, D.; Öberg, K. I.; Qi, C.; Kastner, J. *Astrophysical Journal Letters* **2015**, *807*, L15 (cit. on p. 11).

(100) Irvine, W. M.; Dickens, J. E.; Lovell, A. J.; Schloerb, F. P.; Senay, M.; Bergin, E. A.; Jewitt, D.; Matthews, H. E. *Earth, Moon and Planets* **1997**, *78*, 29–35 (cit. on p. 11).

(101) Rodgers, S. D.; Butner, H. M.; Charnley, S. B.; Ehrenfreund, P. *Advances in Space Research* **2003**, *31*, 2577–2582 (cit. on p. 11).

(102) Moreno, R.; Lellouch, E.; Lara, L. M.; Courtin, R.; Bockelée-Morvan, D.; Hartogh, P.; Rengel, M.; Biver, N.; Banaszkiewicz, M.; González, A. *Astronomy & Astrophysics* **2011**, *536*, DOI: 10.1051/0004-6361/201118189 (cit. on pp. 11, 12, 62).

(103) Hirota, T.; Yamamoto, S.; Kawaguchi, K.; Sakamoto, A.; Ukita, N. *The Astrophysical Journal* **1999**, *520*, 895–900 (cit. on p. 11).

(104) Agúndez, M.; Biver, N.; Santos-Sanz, P.; Bockelée-Morvan, D.; Moreno, R. *Astronomy & Astrophysics* **2014**, *564*, L2 (cit. on p. 11).

(105) Biver, N.; Bockelée-Morvan, D.; Colom, P.; Crovisier, J.; Davies, J. K.; Dent, W. R.; Despois, D.; Gérard, E.; Lellouch, E.; Rauer, H.; Moreno, R.; Paubert, G. *Science* **1997**, *275*, 1915–1918 (cit. on p. 11).

(106) Irvine, W. M. et al. *Nature* **1996**, *383*, 418–420 (cit. on p. 11).

(107) Irvine, W. M.; Bergman, P.; Lowe, T. B.; Matthews, H.; McGonagle, D.; Nummelin, A.; Owen, T. *Origins of Life and Evolution of the Biosphere* **2003**, *33*, 609–619 (cit. on p. 11).

(108) Rodgers, S. D.; Charnley, S. B. *Monthly Notices of the Royal Astronomical Society* **2001**, *323*, 84–92 (cit. on p. 11).

(109) Nixon, C. A. et al. *Planetary and Space Science* **2018**, *155*, 50–72 (cit. on p. 11).

(110) Sagan, C.; Khare, B. N. *Nature* **1979**, *277*, 102–107 (cit. on p. 11).

(111) Vuitton, V.; Bonnet, J. Y.; Frisari, M.; Thissen, R.; Quirico, E.; Dutuit, O.; Schmitt, B.; Le Roy, L.; Fray, N.; Cottin, H.; Sciamma-O'Brien, E.; Carrasco, N.; Szopa, C. *Faraday Discussions* **2010**, *147*, 495–508 (cit. on pp. 12, 15).

(112) Cable, M. L.; Hörst, S. M.; Hodyss, R.; Beauchamp, P. M.; Smith, M. A.; Willis, P. A. *Chemical Reviews* **2012**, *112*, 1882–1909 (cit. on p. 12).

(113) Lavvas, P. P.; Coustenis, A.; Vardavas, I. M. *Planetary and Space Science* **2008**, *56*, 67–99 (cit. on p. 12).

(114) Teanby, N. A.; Irwin, P. G.; de Kok, R.; Vinatier, S.; Bézard, B.; Nixon, C. A.; Flasar, F. M.; Calcutt, S. B.; Bowles, N. E.; Fletcher, L.; Howett, C.; Taylor, F. W. *Icarus* **2007**, *186*, 364–384 (cit. on p. 12).

(115) Lavvas, P.; Griffith, C. A.; Yelle, R. V. *Icarus* **2011**, *215*, 732–750 (cit. on p. 12).

(116) Willacy, K.; Allen, M.; Yung, Y. *The Astrophysical Journal* **2016**, *829*, 79 (cit. on p. 12).

(117) Ennis, C.; Cable, M. L.; Hodyss, R.; Maynard-Casely, H. E. *ACS Earth and Space Chemistry* **2020**, *4*, 1195–1200 (cit. on p. 12).

(118) Izquierdo-Ruiz, F.; Cable, M. L.; Hodyss, R.; Vu, T. H.; Sandström, H.; Lobato, A.; Rahm, M. *Proceedings of the National Academy of Sciences of the United States of America* **2025**, *122*, e2507522122 (cit. on p. 12).

(119) Loison, J. C.; Hébrard, E.; Dobrijevic, M.; Hickson, K. M.; Caralp, F.; Hue, V.; Gronoff, G.; Venot, O.; Bénilan, Y. *Icarus* **2015**, *247*, 218–247 (cit. on p. 12).

(120) Petrie, S. *Icarus* **2001**, *151*, 196–203 (cit. on p. 12).

(121) Strecker, A. *Justus Liebigs Annalen der Chemie* **1854**, *91*, 349–351 (cit. on p. 12).

(122) Miller, S. L. *Annals of the New York Academy of Sciences* **1957**, *69*, 260–275 (cit. on p. 12).

(123) Pearce, B. K. D.; Molaverdikhani, K.; Pudritz, R. E.; Henning, T.; Cerrillo, K. E. *The Astrophysical Journal* **2022**, *932*, 9 (cit. on p. 12).

(124) Navarro-González, R.; Molina, M. J.; Molina, L. T. *Geophysical Research Letters* **1998**, *25*, 3123–3126 (cit. on p. 12).

(125) Todd, Z. R.; Öberg, K. I. *Astrobiology* **2020**, *20*, 1109–1120 (cit. on p. 12).

(126) Itcovitz, J. P.; Rae, A. S. P.; Citron, R. I.; Stewart, S. T.; Sinclair, C. A.; Rimmer, P. B.; Shorttle, O. *The Planetary Science Journal* **2022**, *3*, 115 (cit. on p. 12).

(127) McDonald, C. H.; Bonsor, A.; Rae, A. S.; Rimmer, P. B.; Anslow, R. J.; Todd, Z. R. *Icarus* **2025**, *441*, 116704 (cit. on p. 12).

(128) Tian, F.; Kasting, J. F.; Zahnle, K. *Earth and Planetary Science Letters* **2011**, *308*, 417–423 (cit. on p. 12).

(129) Miyakawa, S.; Cleaves H., J.; Miller, S. L. *Origins of Life and Evolution of the Biosphere* **2002**, *32*, 195–208 (cit. on pp. 12, 13).

(130) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *Science* **1966**, *153*, 72–73 (cit. on pp. 12, 13).

(131) Miyakawa, S.; Cleaves, H. J.; Miller, S. L. *Origins of Life and Evolution of the Biosphere* **2002**, *32*, 209–218 (cit. on pp. 12, 13, 17).

(132) Ferus, M.; Pietrucci, F.; Saitta, A. M.; Knížek, A.; Kubelík, P.; Ivanek, O.; Shestivska, V.; Civiš, S. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 4306–4311 (cit. on p. 12).

(133) Saladino, R.; Crestini, C.; Costanzo, G.; Dimauro, E. *Topics in Current Chemistry* **2005**, *259*, 29–68 (cit. on p. 12).

(134) Signorile, M.; Pantaleone, S.; Balucani, N.; Bonino, F.; Martra, G.; Ugliengo, P. *Molecules 2020, Vol. 25, Page 2274* **2020**, *25*, 2274 (cit. on p. 12).

(135) Proust, J. L. *Ann. Chim. Phys* **1806**, *60* (cit. on p. 13).

(136) Ruiz-Bermejo, M.; de la Fuente, J. L.; Pérez-Fernández, C.; Mateo-Martí, E. *Processes* **2021**, *9*, 597 (cit. on pp. 13, 15, 48).

(137) Ruiz-Bermejo, M.; Zorzano, M.-P.; Osuna-Esteban, S. *Life* **2013**, *3*, 421–448 (cit. on pp. 13, 16, 17, 48, 68).

(138) He, C.; Lin, G.; Upton, K. T.; Imanaka, H.; Smith, M. A. *The Journal of Physical Chemistry A* **2012**, *116*, 4751–4759 (cit. on pp. 13, 15).

(139) Sandström, H.; Rahm, M. *ACS Earth and Space Chemistry* **2021**, *5*, 2152–2159 (cit. on pp. 13, 52, 53).

(140) Budil, D. E.; Roebber, J. L.; Liebman, S. A.; Matthews, C. N. *Astrobiology* **2003**, *3*, 323–329 (cit. on p. 13).

(141) Mizutani, H.; Mikuni, H.; Takahasi, M.; Noda, H. *Origins of life* **1975**, *6*, 513–525 (cit. on pp. 13, 15).

(142) Mozhaev, P. S.; Kichigina, G. A.; Kiryukhin, D. P. *High Energy Chemistry* **1994**, *29* (cit. on p. 13).

(143) Niketić, V.; Draganić, Z.; Nešković, S.; Draganić, I. *Journal of Molecular Evolution* **1982**, *18*, 130–136 (cit. on pp. 13, 15).

(144) Yuasa, S.; Ishigami, M. *Geochemical Journal* **1977**, *11*, 247–252 (cit. on p. 13).

(145) Matthews, C. N.; Moser, R. E. *Proceedings of the National Academy of Sciences* **1966**, *56*, 1087–1094 (cit. on pp. 13, 17).

(146) Yamada, Y.; Kumashiro, I.; Takenishi, T. *Journal of Organic Chemistry* **1968**, *33*, 642–647 (cit. on p. 13).

(147) Santalucia, R.; Pazzi, M.; Bonino, F.; Signorile, M.; Scarano, D.; Ugliengo, P.; Spoto, G.; Mino, L. *Phys. Chem. Chem. Phys* **2022**, *24*, 7224 (cit. on pp. 13, 17).

(148) Bonnet, J.-Y.; Thissen, R.; Frisari, M.; Vuitton, V.; Quirico, Orthous-Daunay, F.-R.; Dutuit, O.; Le Roy, L.; Fray, N.; Cottin, H.; Hörst, S. M.; Yelle, R. V. *International Journal of Mass Spectrometry* **2013**, *354-355*, DOI: 10.1016/j.ijms.2013.06.015 (cit. on pp. 13–15, 48, 67).

(149) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *J. Mol. Biol* **1967**, *80*, 223–253 (cit. on pp. 13, 14, 17, 53).

(150) Orgel, L. E. *Origins of Life and Evolution of the Biosphere* **2004**, *34*, 361–369 (cit. on pp. 13, 17).

(151) Levy, M.; Miller, S. L.; Brinton, K.; Bada, J. L. *Icarus* **2000**, *145*, 609–613 (cit. on p. 13).

(152) Moser, R. E.; Claggett, A. R.; Matthews, C. N. *Tetrahedron Letters* **1968**, *9*, 1599–1603 (cit. on pp. 13, 48).

(153) Moser, R. E.; Claggett, A. R.; Matthews, C. N. *Tetrahedron Letters* **1968**, *9*, 1605–1608 (cit. on p. 13).

(154) Ruiz-Bermejo, M.; García-Armada, P.; Mateo-Martí, E.; de la Fuente, J. L. *European Polymer Journal* **2022**, *162*, 110897 (cit. on pp. 13, 48).

(155) Hortelano, C.; Ruiz-Bermejo, M.; de la Fuente, J. L. *Polymers 2023, Vol. 15, Page 486* **2023**, *15*, 486 (cit. on p. 13).

(156) Cataldo, F.; Patanè, G.; Compagnini, G. *Journal of Macromolecular Science Part a-Pure and Applied Chemistry* **2009**, *46*, 1039–1048 (cit. on p. 13).

(157) Hudson, J. S.; Eberle, J. F.; Vachhani, R. H.; Rogers, L. C.; Wade, J. H.; Krishnamurthy, R.; Springsteen, G. *Angewandte Chemie International Edition* **2012**, *51*, 5134–5137 (cit. on pp. 13, 17).

(158) Moffat, J. B. *Journal of the Chemical Society, Chemical Communications* **1975**, 888–890 (cit. on p. 13).

(159) Sandström, H.; Rahm, M. *The Journal of Physical Chemistry A* **2023**, *127*, 4503–4510 (cit. on pp. 14, 18, 52).

(160) Oró, J.; Kimball, A. P. *Archives of Biochemistry and Biophysics* **1962**, *96*, 293–313 (cit. on p. 14).

(161) Nandi, S.; Bhattacharyya, D.; Anoop, A. *Chemistry – A European Journal* **2018**, *24*, 4885–4894 (cit. on p. 14).

(162) De la Fuente, J. L.; Ruiz-Bermejo, M.; Nna-Mvondo, D.; Minard, R. D. *Polymer Degradation and Stability* **2014**, *110*, 241–251 (cit. on pp. 14, 15).

(163) Ferris, J. P.; Hagan, W. J. *Tetrahedron* **1984**, *40*, 1093–1120 (cit. on p. 14).

(164) Ferris, J. P.; Edelson, E. H. *Journal of Organic Chemistry* **1978**, *43*, 3989–3995 (cit. on pp. 14, 53).

(165) Ferris, J. P.; Matthews, C. N. *Science* **1979**, *203*, 1135–1137 (cit. on p. 14).

(166) Marin-Yaseli, M. R.; Cid, C.; Yagüe, A. I.; Ruiz-Bermejo, M. *Chemistry & Biodiversity* **2017**, *14*, e1600241 (cit. on p. 15).

(167) Mas, I.; de la Fuente, J. L.; Ruiz-Bermejo, M. *European Polymer Journal* **2020**, *132*, 109719 (cit. on p. 15).

(168) Villafañe-Barajas, S. A.; Ruiz-Bermejo, M.; Rayo-Pizarroso, P.; Colín-García, M. *Processes* **2020**, *8*, 968 (cit. on pp. 15, 68).

(169) Ruiz-Bermejo, M.; de la Fuente José, L.; Rogero, C.; Menor-Salván, C.; Osuna-Esteban, S.; Martín-Gago, J. A. *Chemistry & Biodiversity* **2012**, *9*, 25–40 (cit. on pp. 15, 48, 67).

(170) Mamajanov, I.; Herzfeld, J. *The Journal of Chemical Physics* **2009**, *130*, 134503 (cit. on pp. 15, 48, 52).

(171) Völker, T. *Angewandte Chemie* **1960**, *72*, 379–384 (cit. on pp. 15, 48).

(172) Umemoto, K.; Takahasi, M.; Yokota, K. *Origins of Life and Evolution of the Biosphere* **1987**, *17*, 283–293 (cit. on pp. 15, 48).

(173) Ruiz-Bermejo, M.; de la Fuente, J. L.; Carretero-González, J.; García-Fernández, L.; Aguilar, M. R. *Chemistry – A European Journal* **2019**, *25*, 11437–11455 (cit. on pp. 15, 68).

(174) Borquez, E.; Cleaves, H. J.; Lazcano, A.; Miller, S. L. *Origins of Life and Evolution of Biosphere* **2005**, *35*, 79–90 (cit. on pp. 17, 56).

(175) Yamada, Y.; Nagashima, N.; Iwashita, Y.; Nakamura, A.; Kumashiro, I. *Tetrahedron Letters* **1968**, *9*, 4529–4532 (cit. on p. 17).

(176) Ferris, J. P.; Orgel, L. E. *Journal of the American Chemical Society* **1965**, *87*, 4976–4977 (cit. on p. 17).

(177) Shuman, R. F.; Shearin, W. E.; Tull, R. J. *Journal of Organic Chemistry* **1979**, *44*, 4532–4536 (cit. on p. 17).

(178) Ferris, J. P.; Orgel, L. E. *Journal of the American Chemical Society* **1966**, *88*, 1074 (cit. on p. 17).

(179) Ferris, J. P.; Orgel, L. E. *Journal of the American Chemical Society* **1966**, *88*, 3829–3831 (cit. on p. 17).

(180) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *Journal of Molecular Biology* **1968**, *38*, 121–128 (cit. on p. 17).

(181) Barks, H. L.; Buckley, R.; Grieves, G. A.; Di Mauro, E.; Hud, N. V.; Orlando, T. M. *ChemBioChem* **2010**, *11*, 1240–1243 (cit. on p. 17).

(182) Boulanger, E.; Anoop, A.; Nachtigallova, D.; Thiel, W.; Barbatti, M. *Angewandte Chemie* **2013**, *125*, 8000–8003 (cit. on pp. 17, 18).

(183) Voet, A. B.; Schwartz, A. W. *Bioorganic Chemistry* **1983**, *12*, 8–17 (cit. on pp. 17, 51, 52).

(184) Wang, J.; Gu, J.; Nguyen, M. T.; Springsteen, G.; Leszczynski, J. *Journal of Physical Chemistry B* **2013**, *117*, 14039–14045 (cit. on p. 18).

(185) Roy, D.; Najafian, K.; von Rague Schleyer, P. *Proceedings of the National Academy of Sciences* **2007**, *104*, 17272–17277 (cit. on pp. 18, 48).

(186) Armas-Vázquez, M. Z.; González-Espinoza, C. E.; Segura, A.; Heredia, A.; Miranda-Rosete, A. *Astrobiology* **2023**, *23*, DOI: 10.1089/AST.2022.0050 (cit. on p. 18).

(187) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, I., Ed., Mineola, New York, 1996 (cit. on p. 19).

(188) Born, M.; Oppenheimer, R. *Annalen der Physik* **1927**, *389*, 457–484 (cit. on p. 20).

(189) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chemical Physics Letters* **1989**, *157*, 479–483 (cit. on p. 24).

(190) Stanton, J. F. *Chemical Physics Letters* **1997**, *281*, 130–134 (cit. on pp. 24, 32).

(191) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864 (cit. on p. 24).

(192) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133 (cit. on p. 25).

(193) Mardirossian, N.; Head-Gordon, M. *Molecular Physics* **2017**, *115*, 2315–2372 (cit. on p. 25).

(194) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. *Science* **2017**, *355*, DOI: 10.1126/SCIENCE.AAH5975 (cit. on p. 26).

(195) Perdew, J. P.; Schmidt, K. *AIP Conference Proceedings* **2001**, *577*, 1–20 (cit. on p. 26).

(196) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Journal of Physical Chemistry A* **2007**, *111*, 10439–10452 (cit. on p. 26).

(197) Gerrits, N.; Smeets, E. W.; Vuckovic, S.; Powell, A. D.; Doblhoff-Dier, K.; Kroes, G. J. *Journal of Physical Chemistry Letters* **2020**, *11*, 10552–10560 (cit. on pp. 26, 41).

(198) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Physical Review Letters* **1996**, *77*, 3865 (cit. on p. 26).

(199) Csonka, G. I.; Perdew, J. P.; Ruzsinszky, A.; Philipsen, P. H.; Lebègue, S.; Paier, J.; Vydrov, O. A.; Ángyán, J. G. *Physical Review B* **2009**, *79*, 155107 (cit. on p. 26).

(200) Bursch, M.; Mewes, J. M.; Hansen, A.; Grimme, S. *Angewandte Chemie International Edition* **2022**, *61*, e202205735 (cit. on pp. 26, 27, 36, 39, 42–46).

(201) Tirado-Rives, J.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2008**, *4*, 297–306 (cit. on p. 26).

(202) Pribram-Jones, A.; Gross, D. A.; Burke, K. *Annual Review of Physical Chemistry* **2015**, *66*, 283–304 (cit. on p. 27).

(203) Grimme, S. *Chemistry – A European Journal* **2012**, *18*, 9955–9964 (cit. on pp. 27, 45).

(204) Grimme, S.; Ehrlich, S.; Goerigk, L. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465 (cit. on p. 27).

(205) Hoffmann, R., *Solids and surfaces : a chemist's view of bonding in extended structures*; VCH Publishers: 1988, p 142 (cit. on p. 29).

(206) Hehre, W. J.; Ditchfield, K.; Pople, J. A. *The Journal of Chemical Physics* **1972**, *56*, 2257–2261 (cit. on p. 30).

(207) Dunning, T. H.; Dunning, T. H. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023 (cit. on p. 30).

(208) Blöchl, P. E. *Physical Review B* **1994**, *50*, 17953 (cit. on p. 31).

(209) Riplinger, C.; Neese, F. *The Journal of Chemical Physics* **2013**, *138*, DOI: 10.1063/1.4773581 (cit. on p. 32).

(210) De Souza, B. *Angewandte Chemie International Edition* **2025**, *64*, e202500393 (cit. on pp. 33, 44).

(211) M.J. Frisch; et al. Gaussian 16, Revision B.01, 2016 (cit. on p. 33).

(212) Kresse, G.; Furthmüller, J. *Physical Review B* **1996**, *54*, 11169 (cit. on p. 33).

(213) Neese, F. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2025**, *15*, e70019 (cit. on pp. 34, 44).

(214) Mennucci, B. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 386–404 (cit. on p. 37).

(215) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. *ACS Catalysis* **2019**, *9*, 6803–6813 (cit. on pp. 37, 39, 45, 46).

(216) Wulff, G. *Zeitschrift für Kristallographie - Crystalline Materials* **1901**, *34*, 449–530 (cit. on p. 38).

(217) Rahm, J. M.; Erhart, P. *The Journal of Open Source Software* **2020**, *5*, DOI: 10.21105/joss.01944 (cit. on p. 38).

(218) Ryu, H.; Park, J.; Kim, H. K.; Park, J. Y.; Kim, S. T.; Baik, M. H. *Organometallics* **2018**, *37*, 3228–3239 (cit. on pp. 39, 45).

(219) Evarestov, R. A.; Bredow, T.; Jug, K. *Physics of the Solid State 2001 43:9* **2001**, *43*, 1774–1782 (cit. on p. 41).

(220) Pople, J. A. *Reviews of Modern Physics* **1999**, *71*, 1267 (cit. on p. 41).

(221) Coulson, C. A.; McWeeny, R., *Coulson's valence*; OUP: 1979, p 435 (cit. on p. 42).

(222) Kurth, S.; Perdew, J. P.; Blaha, P. *International Journal of Quantum Chemistry* **1999**, *75*, DOI: 10.1002/(SICI)1097-461X(1999)75:4/5 (cit. on p. 42).

(223) Borlido, P.; Aull, T.; Huran, A. W.; Tran, F.; Marques, M. A.; Botti, S. *Journal of Chemical Theory and Computation* **2019**, *15*, 5069–5079 (cit. on p. 42).

(224) Lynch, B. J.; Truhlar, D. G. *Journal of Physical Chemistry A* **2001**, *105*, 2936–2941 (cit. on p. 42).

(225) Paulechka, E.; Kazakov, A. *Journal of Physical Chemistry A* **2017**, *121*, 4379–4387 (cit. on p. 42).

(226) Sandler, I.; Chen, J.; Taylor, M.; Sharma, S.; Ho, J. *Journal of Physical Chemistry A* **2021**, *125*, 1553–1563 (cit. on p. 42).

(227) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368–1372 (cit. on p. 43).

(228) Schlegel, H. B. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 790–809 (cit. on p. 43).

(229) Bannwarth, C.; Ehlert, S.; Grimme, S. *Journal of Chemical Theory and Computation* **2019**, *15*, 1652–1671 (cit. on p. 43).

(230) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. *Journal of Chemical Physics* **2024**, *160*, 114110 (cit. on p. 44).

(231) Young, T. A.; Silcock, J. J.; Sterling, A. J.; Duarte, F. *Angewandte Chemie International Edition* **2021**, *60*, 4266–4274 (cit. on p. 44).

(232) Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A. *Journal of Physical Chemistry B* **2008**, *112*, 9709–9719 (cit. on p. 46).

(233) Börnstein, L., *Vapor Pressure and Antoine Constants for Nitrogen Containing Organic Compounds*; Springer-Verlag: 2001; Vol. 20C (cit. on p. 46).

(234) R. Glaser; B. Hodgen; D. Farrelly; E. McKee *Astrobiology* **2007**, *7*, 455–470 (cit. on p. 48).

(235) Jung, S. H.; Choe, J. C. *Astrobiology* **2013**, *13*, 465–475 (cit. on p. 48).

(236) Mamajanov, I.; Herzfeld, J. *The Journal of Chemical Physics* **2009**, *130*, 134504 (cit. on p. 48).

(237) Mas, I.; Hortelano, C.; Ruiz-Bermejo, M.; de la Fuente, J. L. *European Polymer Journal* **2021**, *143*, 110185 (cit. on p. 48).

(238) Gaweska, H.; Fitzpatrick, P. F. *Biomolecular Concepts* **2011**, *2*, 365–377 (cit. on p. 55).

(239) Ferris, J. P.; Ryan, T. J. *J. Org. Chem.* **1973**, *38*, 3302–3307 (cit. on p. 55).

(240) Shaik, S.; Ramanan, R.; Danovich, D.; Mandal, D. *Chemical Society Reviews* **2018**, *47*, 5125–5145 (cit. on p. 61).

(241) Stuyver, T.; Danovich, D.; Joy, J.; Shaik, S. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10*, e1438 (cit. on p. 60).

(242) Molina-Cuberos, G. J.; López-Moreno, J. J.; Rodrigo, R.; Lara, L. M.; O'Brien, K. *Planetary and Space Science* **1999**, *47*, 1347–1354 (cit. on p. 62).

(243) Okamoto, S.; Millis, A. J. *Nature* **2004**, *428*, 630–633 (cit. on p. 63).

(244) Nakagawa, N.; Hwang, H. Y.; Muller, D. A. *Nature Materials* **2006**, *5*, 204–209 (cit. on p. 63).

(245) Herbst, E.; Van Dishoeck, E. F. *Annual Review of Astronomy and Astrophysics* **2009**, *47*, 427–480 (cit. on p. 65).

Appended Papers