THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Nitriles in Prebiotic Chemistry and Astrobiology

HILDA SANDSTRÖM

Department of Chemistry and Chemical Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2022

Nitriles in Prebiotic Chemistry and Astrobiology
HILDA SANDSTRÖM
ISBN 978-91-7905-652-0

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5118
ISSN 0346-718X

Department of Chemistry and Chemical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: Some of the nitriles studied in this thesis in front of Earth and Titan.

Chalmers Digitaltryck
Gothenburg, Sweden 2022

Nitriles in Prebiotic Chemistry and Astrobiology

HILDA SANDSTRÖM

Department of Chemistry and Chemical Engineering

Chalmers University of Technology

ABSTRACT

Life appeared on Earth within a billion years of the planet's formation. How? - no one knows. Theories regarding the origin of life involve reactions of molecules predicted to have existed on early Earth in what is called prebiotic chemistry. In this thesis, I use computational methods to investigate hypotheses in prebiotic chemistry and astrobiology. With computational chemistry, it is possible to predict the thermodynamics and kinetics of chemical processes. The long-term goal of this line of research is furthering our understanding of the origin of life.

The first part of this thesis is devoted to hydrogen cyanide (HCN) chemistry. HCN is believed to have been present on early Earth. Molecular building blocks of DNA, RNA, and proteins have been detected among HCN reaction products. However, because of HCN's reactivity, the molecule forms numerous other compounds as well. One such set of proposed reaction products are HCN-derived polymers - a diverse group of structures which have been proposed to form in many ways. Here I present a thermodynamic landscape of HCN-derived molecules and polymers. Using the thermodynamic map, some hypothesized reaction pathways are proven to be unfeasible. Polyaminoimidazole is estimated to be one of the most stable polymers, while the nucleobase adenine is computed as the most stable of all studied structures. We also investigate the first steps in the formation of two proposed HCN reaction products: diaminomaleonitrile and polyimine. Our results reveal that all studied competing reactions have similar activation barriers. These results open for the possibility of a diverse beginning to HCN oligomerization under kinetic control. The estimated timescale of HCN oligomerization suggests that reactions in low-temperature environments as cold as 200 K could occur within thousands of years. I discuss the implications of the predicted reaction rates for HCN chemistry in astrochemical environments like comets and Saturn's moon Titan.

The second part of the thesis investigates an astrobiological hypothesis: the possibility for cryogenically operable membranes in the seas of Titan. It is concluded that a previously suggested polarity-inverted membrane made from acrylonitrile, a so-called azotosome, cannot spontaneously self-assemble and is therefore unlikely to exist on Titan.

KEYWORDS: astrobiology, prebiotic chemistry, hydrogen cyanide, polymer chemistry, molecular dynamics, umbrella sampling, metadynamics, density functional theory

LIST OF PAPERS:

This thesis is based on the following papers, referred to in the text by Roman numerals

I. **A Thermodynamic Landscape of Hydrogen Cyanide-Derived Molecules and Polymers.** Hilda Sandström, Fernando Izquierdo-Ruiz, Rana Dogan, Siddhant Sharma and Martin Rahm. *Manuscript.*

II. **The Beginning of HCN Polymerization: Iminoacetonitrile Formation and Its Implications in Astrochemical Environments.** Hilda Sandström and Martin Rahm. (2021) ACS Earth and Space Chemistry, **5**, 2152-2159.

III. **Simulating the Origin of Hydrogen Cyanide Polymerization.** Hilda Sandström and Martin Rahm. *Manuscript.*

IV. **Can Polarity-Inverted Membranes Self-Assemble on Titan?** Hilda Sandström and Martin Rahm. (2020) Science Advances, **6**, eaax0272.

CONTRIBUTION REPORT

Description of the author's contribution to the appended papers.

I. Performed quantum chemical calculations together with F.I-R, R.D and S.S. Analyzed the data. Wrote the manuscript together with F.I-R and M.R. Designed the study together with M.R.

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

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

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

LIST OF ABBREVIATIONS

ATP	Adenosine triphosphate
B3LYP	Becke, 3-parameter, Lee-Yang-Parr
BO	Born-Oppenheimer
DFT	Density functional theory
DNA	Deoxyribonucleic acid
Ga	Billion years
KS	Kohn-Sham
MAD	Mean absolute deviation
PBE	Perdew–Burke-Ernzerhof
PCM	Polarizable continuum model
RNA	Ribonucleic acid
VASP	Vienna ab initio simulation package

TABLE OF CONTENTS

1
Introduction

The Universe encompasses a vastly diverse set of environments, characterized by different pressures, temperatures, and chemical compositions. Among these environments are stars, planets, moons, comets, asteroids, and interstellar gas clouds, to name a few. Their chemical processes can provide information on the nature and evolution of the Universe and are of fundamental scientific interest. The term *astrochemistry* refers to the study of any chemical reaction, molecule or material in the Universe. However, the term is often used to signify only the study of chemistry outside Earth. In the past century, technological advances have made it possible to study the chemistry of remote locations; radio telescopes can detect molecules by their rotation emission spectra; space aircrafts can probe the chemistry of planets, comets, moons, and asteroids using common equipment also found in chemistry laboratories.

One of the large unanswered questions in science, which overlaps with astrochemistry and many other research fields, is how common life is in the Universe. Research on the nature and ubiquity of life in the Universe is called *astrobiology*. Astrobiologists are concerned with finding environments that have chemical features called biosignatures, indicating the presence of biological processes and potentially habitable environments. Another part of astrobiology research, called *prebiotic chemistry,* is concerned with finding the chemical processes which could have led to the origin of life. Much of prebiotic chemistry research focuses on finding plausible formation routes for building blocks of

proteins, nucleic acids, and membranes, all of which are integral parts of the cell and its machinery (1, 2). Over the past century, progress has been made in finding chemical systems which can generate biomolecular building blocks from simple molecules, in reactions thought to be relevant to the early Earth. Demonstrations of such processes are taken as first steps towards understanding how life can develop from abiotic chemical reactions. Yet many questions remain unanswered and hypotheses untested in prebiotic chemistry and astrobiology

The work presented in this thesis aims to contribute to the symbiosis between experimental and computational astrobiology which has gradually developed over the past decades. The idea is to test hypotheses in astrobiology and prebiotic chemistry by computationally evaluating different aspects of chemical stability and reactivity. The motivation for pursuing these questions with computational chemistry is that astrobiology hypotheses can be challenging to test experimentally, either because of the exotic chemistry, cost of the experiment, or the timescales associated with the underlying chemistry. To this end, computational chemistry has developed as an important tool at the astrobiologist's disposal. Computational chemistry can be used for prediction of abundances of molecules, to simulate molecular spectra, and predict chemical stability. The chemical data gathered through observations and space missions can be integrated into theoretical models (3, 4). There is continuous feedback between experimental observations and computational chemistry predictions which helps to improve the understanding of the chemical processes in the Universe.

1.1 Chemistry on Titan

One of the environments I will focus on is Saturn's moon Titan. Titan is a unique place in our solar system because it has a rich atmospheric chemistry filled with many organic molecules. The atmosphere consists of methane and nitrogen (5). Photochemistry in the atmosphere generates a plenitude of organic species of varying complexity (6, 7). Figure 1.1 illustrates some of Titan's atmospheric processes and characteristics. Several of the organic molecules are envisioned to combine to form aerosols which, in turn, make out Titan's thick haze (8). However, the chemical processes leading to the formation of larger organic particles on Titan are largely unknown (9). The temperature in the atmosphere ranges between ~70-~190 K (10). The warmest temperature being found at an ~300 km altitude. Meanwhile, the surface temperature, has been measured to 90-94 K (11).

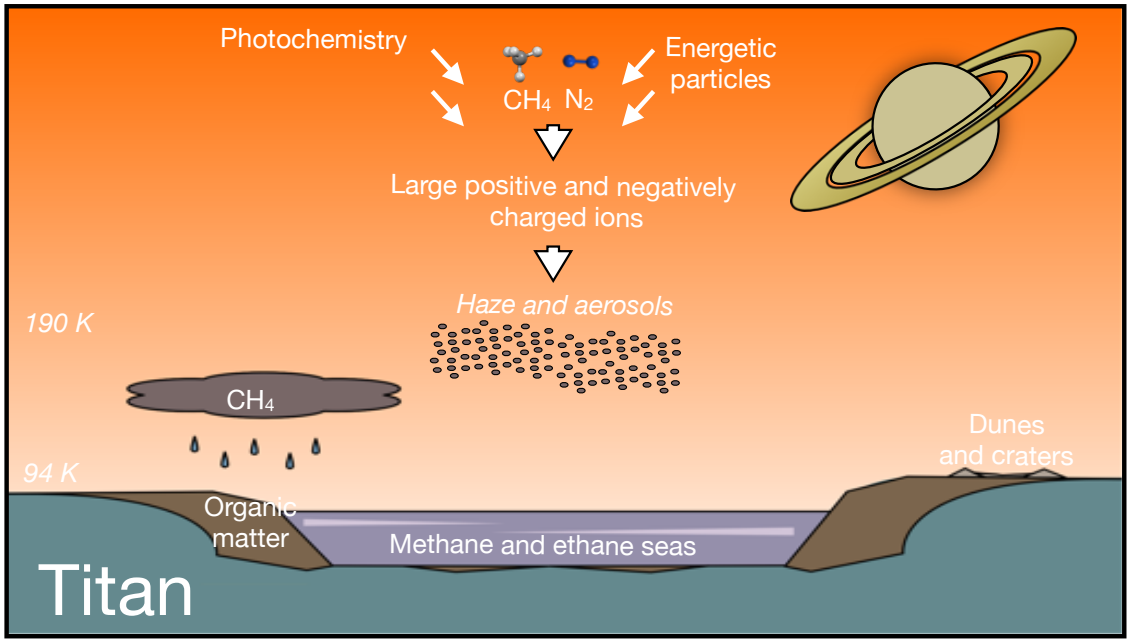

Figure 1.1. The hydrocarbon world of Titan. The photochemistry occurring in the outer layers of the atmosphere is connected to the formation of more complex molecules such as large positive and negatively charged ions, as well as aerosol particles at lower altitudes. Titan's surface is covered with organic materials. There are hydrocarbon seas made from methane and ethane. The craters on the surface are footprints of ancient collisions with high-energy impactors. 94 K and 190 K are representative maximum temperatures on Titan's surface and in the moon's atmosphere (~300 km altitude), respectively.

The surface of Titan is covered in most-part by solid organic dust. The surface is also in part speckled, mainly in the polar regions, by liquid methane and ethane lakes (12). Observations of these lakes makes Titan the only other body in the solar system which is known to have liquids on its surface. The methane in the atmosphere forms clouds which take part in a methane cycle like the hydrological cycle on Earth (13). However, the liquids in the seas of Titan can also be considered as chemically opposite of Earth's water seas. Methane and ethane are hydrophobic molecules and non-polar solvents. Therefore, anything that dissolves well in water will be poorly soluble in the seas of Titan.

The similarities of Titan and the early Earth have generated many suggestions regarding the potential for life on the cold moon. Titan is one of the main subjects of interest for current astrobiology research outside Earth (14). There have even been suggestions that lifeforms could harvest energy through methanogenic metabolic cycles in the moon's hydrocarbon world (15, 16). '

The low temperature and lack of radiation due to the thick atmosphere means that the available energy in the form of heat and light on Titan's surface is scarce (9). Other possible forms of energy available at the surface of Titan could be that

caused by high-impact collisions. Titan's surface has many craters which are footprints of high-energy impacts predicted to have exposed liquid water to the organic material at the surface (17). Titan's dunes and craters will be the targets of the Dragonfly mission data collection (14).

1.2 Theoretical chemistry as a tool for prebiotic chemistry and astrobiology

Computational chemistry can simulate processes in worlds drastically different from that of Earth, such as the frigid world of Titan. When predicting the outcome of chemical processes two concepts are central: kinetics and thermodynamics. Computational chemistry allows for the calculation of a free energy landscape encompassing reactants, products, and transition states of a reaction. The thermodynamics of a reaction is determined by the relative energy between reactants and products. Temperature can alter the thermodynamics of a reaction via the temperature dependence of the reaction enthalpy and entropy. For example, at low temperatures the entropic contribution to the free energy is small, while weak enthalpic interactions can be of a similar size to the available thermal energy. The second aspect of chemical stability, kinetics, is governed by the relative energy of the reactants and transition state, the activation barrier. Barring tunnelling effects, a high activation barrier will only allow a reaction to occur at high temperatures. Figure 1.2 demonstrates the stark effect of temperature on a first-order reaction half-life. A reaction with a barrier of 80 kJ/mol that occurs on an order of minutes at ambient Earth conditions has an associated timescale of a century at 200 K.

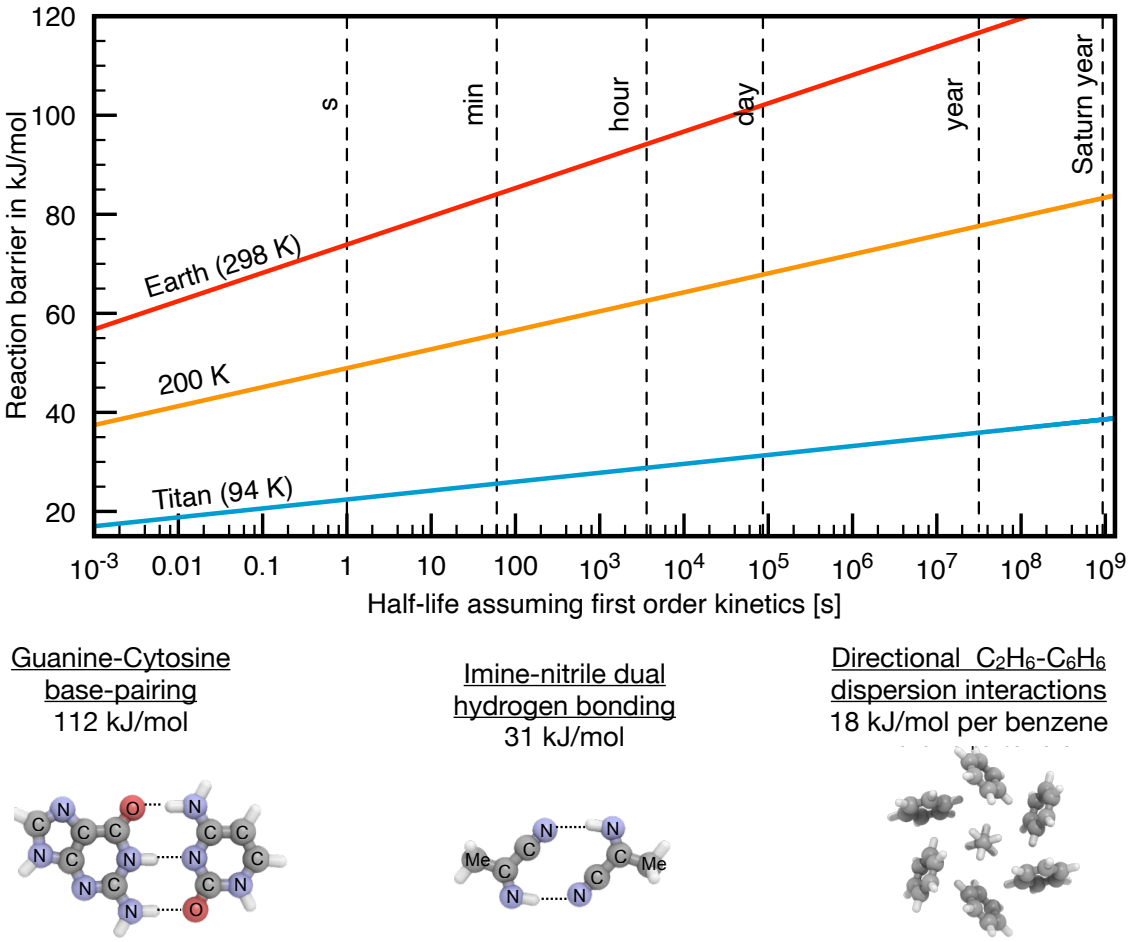

Figure 1.2. Top: The reaction barrier of a chemical process governs the reaction rate. Depending on temperature, the reaction can occur in ms to century timescales. (half-life computed via transition-state theory). **Bottom:** Examples of intermolecular interactions of different strengths. The guanine-cytosine base pair have an interaction strong enough to help support the double helix structure of DNA at ambient conditions on Earth. On Saturn's cold moon Titan, the interaction between the nucleobases is too strong to allow for a dynamic chemistry. The figure is adapted with permission from ref. (18).

1.3 Aim/Research questions

The long-term goal of the work presented in this thesis is to develop hypotheses within astrobiology and origin of life research. Ideally the work will be used to guide experimental efforts which explore prebiotic chemistry in different environments and investigate the potential of life beyond Earth. To this end, this thesis aims to answer a set of questions. The first part of the thesis is centred on the role of hydrogen cyanide (HCN) and HCN-derived molecules and polymers in prebiotic chemistry (Paper I-III). The overall research questions regarding HCN chemistry addressed in this thesis can be summarized as follows. What are plausible structures for the HCN-derived polymers? Which formation routes to HCN-derived polymers are thermodynamically feasible? What are the effects of temperature on the structure of the HCN reaction product? What is the kinetics and thermodynamics for the beginning of base-catalyzed HCN polymerization? Lastly, I answer the question: What is the likelihood of membrane formation on Saturn's moon Titan (Paper IV)?

The thesis is structured as follows. Chapter 2 introduces the reader to origin of life research and the predicted conditions of early Earth. An introduction to HCN chemistry in a prebiotic chemistry context is provided in Chapter 3. A brief introduction to the computational methods that were used for this thesis are presented in Chapter 4. Chapter 5 presents HCN-derived polymers, and a thermodynamic evaluation of various propositions for their structure and formation mechanism. Chapter 6 focuses on HCN reactivity through the study of base-catalyzed HCN reactions. In Chapter 7, the focus is shifted towards Titan, and I discuss the potential for cell membrane formation in such a vastly different world.

2
Origin of Life Research

Origin of life as a research domain in chemistry dates to the 19th century. When Charles Darwin published *The Origin of Species*, ideas of an evolutionary origin of life began to spread. The debate regarding the nature of life's origin sprung among physicists, chemists, and biologists alike. Could biological systems have arisen through a set of evolutionary abiotic chemical processes? Simultaneous and parallel developments in geology, astronomy, chemistry, and molecular biology over the last 200 years, have continued to shape origin of life hypotheses. The discoveries and characterizations of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as other cellular machinery, were central in propelling origin of life research forward. This chapter gives a brief introduction to the history and working hypotheses in prebiotic chemistry and astrobiology and sets the contextual stage upon which we explore nitrile chemistry (Papers I-IV).

2.1 The defining features of life

The current working definition of life by NASA (19) reads as follows "Life is a self-sustaining chemical system capable of Darwinian evolution." However, it widely recognized that a definition of life is difficult to formulate. Functional and molecular attributes of living systems as we know them on Earth have shaped origin of life research so far.

Figure 2.1 shows three fundamental key features of life as we know it: the cell membrane which surrounds the cell, storage of genetic information in DNA, and metabolism, here represented by the energy storing adenosine triphosphate (ATP) molecule. The cell membrane (Fig. 2.1A) protects, concentrates, and compartmentalizes cellular machinery. In the cell, DNA stores genetic information (Fig. 2.1B). DNA replication allows for the propagation of information. Replication and other types of cellular processes require energy. Through the cell's metabolism, stored chemical energy is converted into ATP molecules (Fig 2.1C) for this purpose. Yet, there is currently no agreement on which property developed first (2).

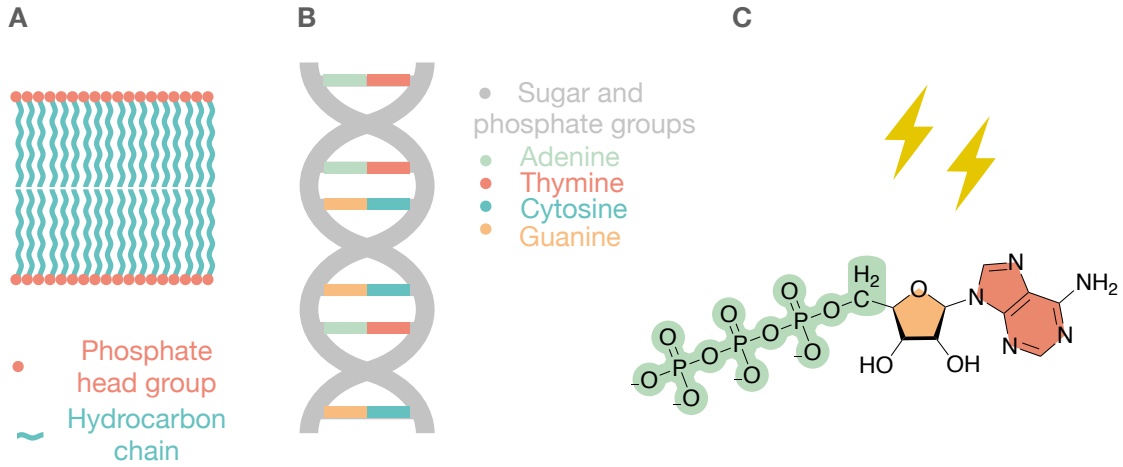

Figure 2.1. Three important aspects of living systems: A. The bilayer membrane of cells B. A model of deoxyribonucleic acid (DNA) C. The energy storing adenosine triphosphate (ATP) molecule, here representing the process of metabolism.

Throughout the history of prebiotic chemistry there have been those who advocate that either one of the three systems shown in Figure 2.1 was the first to develop. The prebiotic origin of the cell membrane was one of the first to be studied (20). Today, there is still ongoing research on the first protocells (see e.g., (21)). Nevertheless, much of the ongoing origin of life debate focuses on information storing and catalytic polymers, versus metabolic cycles (1). In the cell, the polymers in charge of catalysis and genetic storage are proteins and DNA, respectively. However, a co-origin and co-evolution of proteins and DNA is generally considered unlikely (22). The RNA hypothesis provides an alternate scenario, as RNA can perform both information storage and catalysis (see review (22)). The diverse function of RNA has led many to suggest that it was the first information-storing polymer. This theory is referred to as the RNA world hypothesis. Alternatively, a metabolism-first approach instead envisions life beginning with autocatalytic metabolic cycles (i.e., complex replicating chemical systems) (1).

Whichever property came first, much of prebiotic chemistry research focuses on finding ways to make building blocks of DNA, RNA, proteins, metabolic cycles, or membranes, from small organic molecules that are readily available either on early Earth, or elsewhere in the Universe. Among the important biomolecular building blocks are amino acids (which make up proteins), nucleobases and sugars (needed in RNA and DNA), lipids (which make up membranes), or metabolites used in metabolic cycles. Any attempt to uncover the origin of terrestrial life must consider prebiotic synthesis in the context of the early Earth.

2.2 Where on *Earth* did life originate?

Life on Earth has existed as far back as the geological record stretches. The earliest evidence of life on Earth is from ~3.4-3.7 billion years (Ga) ago, and is in the form

of preserved microfossils as well as stromatolites thought to have formed as a result of photosynthetic microorganisms (23–25). Any direct evidence of life in the geological record from earlier periods has been erased during rock metamorphosis (26, 27). Nevertheless, the earliest evidence provides an upper estimate for the time it took for life to establish on Earth. The period which precedes the earliest records of life coincides with the late heavy bombardment, a period that is theorized to have occurred between ~3.5 Ga and ~4.0-4.1 Ga (28) characterized by a high frequency of collisions with asteroids.

Little is known of Earth's chemical conditions during the emergence of life. Uncertainties in factors such as atmospheric composition and temperature affect the outcomes of chemical models of early Earth. There exists geological evidence of liquid water being present on early Earth, indicating a surface temperature above 273 K (29). Yet, the young sun was less bright than it is currently (30). Studies predict that the Earth's surface temperature could have been close to 220 K (31), unless sufficient greenhouse gases were present in the atmosphere (29, 30, 32). Predictions also suggest that the radiation at the Earth's surface was more intense 3.5 Ga ago (33). The composition of the atmosphere could have been mainly controlled by outgassing from the early Earth mantle. In that scenario, the atmosphere would have consisted mainly out of carbon dioxide, nitrogen gas and water vapour (31). Both carbon dioxide and nitrogen gas are very stable and therefore not very reactive. Minor contributions to the atmosphere from reduced species, such as carbon monoxide, methane, hydrogen gas and ammonia, could have been present due to degassing of impact material (31, 34). It was earlier believed that the early Earth atmosphere must have been highly reduced since such a composition favours the formation of prebiotic compounds (35). However, as the composition of the early Earth atmosphere is unknown, the produced quantities of prebiotic molecules remain uncertain (see e.g., (31, 36))

2.2.1 Endogenous production or exogenous delivery

Prebiotic molecules on Earth could have originated through endogenous production or exogenous delivery (37). Endogenous production involves the development of life on Earth through native chemical processes. Clues to an endogenous origin of life could be found by examining the chemical processes occurring on early Earth. For example, formation of organic molecules in the early Earth atmosphere could have been driven by photochemistry, lightning, and shock-impacts (37–39). One of the most famous examples of atmospheric prebiotic chemistry is the Miller-Urey experiment in 1953 (40, 41). Miller and Urey were able to demonstrate the production of amino acids from a simulated early Earth atmosphere (made of ammonia, methane, hydrogen, and water). An electrical spark was used to simulate the effect of atmospheric lightning (35, 40, 41). Later experiments have shown how the formation of organic molecules in such spark-discharge experiments varies with the atmospheric composition (31). A second possible endogenous origin of prebiotic molecules is the early Earth's ocean. However, theoretical predictions suggest that the concentration of organic

molecules in the ocean would have been low (see e.g., ref (42)). This has led to considerations of concentrated organic solutions formed in shallow ponds via high-energy impacts (38) or via eutectic freezing (43). A third major hypothesis for an endogenous origin of life is in hydrothermal vents located on the seafloor. Hydrothermal vents could drive chemical reactions through a temperature and redox potential between the cold ocean and warm seafloor (44). On the other hand, theories of an origin of life through exogenous delivery refers to the distribution of organic and prebiotic material from impact events. For example, organic bearing comets and asteroids can enter the atmosphere (as meteors), and sometime collide with the Earth's surface (as meteorites). One famous example is the Murchison meteorite in which 86 amino acids had been identified up until 2017 (45–47), along with nucleobases (48). Theoretical models have described how the probability that the meteorites' organic material remains intact (does not undergo reaction) depends on the atmosphere and size of the impactor (37).

2.3 Summary

Current prebiotic chemistry and astrobiology research is conducted considering the known characteristics of life on Earth. Astrobiology research involves finding possibilities for metabolic cycles, cell membranes and formation building blocks for proteins, DNA and RNA beyond Earth. In this thesis, I focus on the prebiotic and potential role in astrobiology of two readily available nitriles in the solar system: acrylonitrile and HCN. As will be discussed in the next chapter, HCN has an important role in prebiotic chemistry and astrochemistry hypotheses.

3
HCN, a Molecule at the Center of Prebiotic Chemistry

The search for prebiotic molecules, which may have been important in the origin of life, focuses on certain molecular properties. The prebiotic molecule should (1) be able to form biologically relevant building blocks, by itself or with other prebiotic molecules, (2) be predicted to have existed on early Earth, and ideally (3) be abundant in the Universe. The third criterion would suggest that the molecule might commonly partake in prebiotic reactions beyond Earth. HCN is a small nitrile which fits with all these three criteria. In this chapter, I introduce different aspects which all implicate HCN, and HCN-derived polymers, as main suspects in origin of life research. This chapter serves to set the stage for the astrochemical and prebiotic context within which we investigated HCN chemistry in Papers I-III.

3.1 The HCN molecule

The HCN (Figure 3.1a) molecule is striking in many ways. HCN is a weak acid (pKa of 9.2) that once went under the name prussic acid. Under standard conditions, pure HCN is a volatile liquid with a boiling point close to room temperature (299 K)(49) which freezes already at 260 K (50). Comparing the properties of HCN with water – also a polar protic solvent – highlights the nature of the HCN liquid. The HCN liquid has a dielectric constant of 144.8 at 278 K (51), and the sixth largest reported for a liquid in the CRC handbook (after a group of amides) (52). The dielectric constant of HCN is ~70 % larger than that of water at the same temperature (85.8)(53). The viscosity of HCN is 0.2 cP at ambient conditions (49) – a fifth of that of water. Theoretical simulations have shown that the HCN molecule is highly coordinated by other neighbouring molecules in the liquid (54). The typical strength of a hydrogen bond between HCN molecules is 5-6 kcal/mol, like that of water (55, 56).

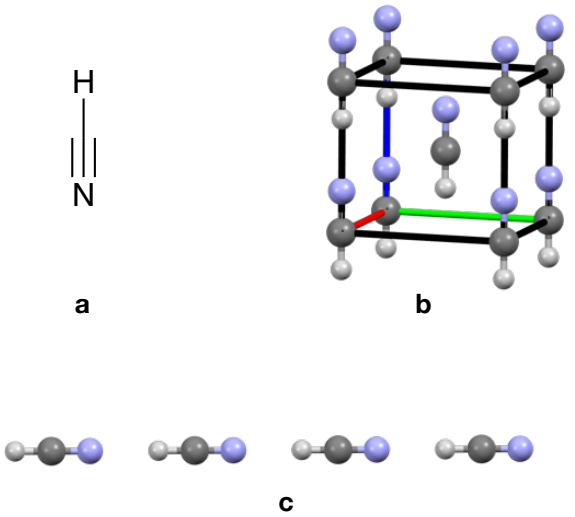

Figure 3.1. a. The Lewis structure of HCN; b. The conventional unit cell of crystalline HCN (57); c. example of a linear cluster of HCN molecules. Figure reproduced from ref. (58).

When HCN solidifies it forms a crystal with a polar space group, seen in Figure 3.1b (57). The polar crystal is pyroelectric. Ampoules storing frozen HCN at 77 K have been observed to glow in the dark (59). In the gas phase, HCN forms small hydrogen bonded linear clusters (shown in Figure 3.1c) (60). Cooperative effects are important in these hydrogen bonded systems. The hydrogen bond strength of HCN increases by 15 % when comparing the HCN hydrogen bonded dimer and pentamer (55). Similarly, theoretical studies show that in a hydrogen bonded chain of HCN molecules, the dipole moment (exp. value 2.98 D (61)) increases by a factor of 1.4 (62).

Anhydrous HCN can be synthesized by different methods, such as by reacting an alkali salt (e.g., potassium or sodium cyanide) with stearic acid, followed by fractional condensation (63). Still, HCN is a hazardous chemical –it can inhibit oxygen uptake by binding to the protein cytochrome C oxidase (64) – and requires careful handling. Because HCN is both volatile and toxic, HCN chemistry is not something easily studied in the laboratory.[1]

[1]Text adapted from a previously published licentiate thesis (58).

3.2 HCN in prebiotic chemistry

HCN has a central role in prebiotic chemistry, experimentally established through the Urey-Miller experiments in the 1950s (35, 40). Their seminal experiments demonstrated the formation of amino acids upon applying an electrical discharge to gas mixtures (H_2O, NH_3, H_2 and CH_4) meant to imitate the early Earth atmosphere. HCN was a key intermediate in the reaction process. Through considerations of both organic and inorganic materials that likely were available on early Earth, Sutherland and colleagues have connected HCN chemistry to the formation of amino acids, sugars, nucleobases, and lipid building blocks (65). Moreover, the cyanide anion has been proposed to have had a role in the early development of metabolism as a reducing agent (66). Moreover, Oró was first to demonstrate that base-catalyzed reactions of aqueous HCN form adenine, along with a plethora of other compounds (67, 68). Today, all natural nucleobases of DNA and RNA have been identified as products in HCN reaction mixtures (see e.g., ref. (69)). In experiments, the formation of adenine has been observed in aqueous solutions of HCN at kept at 195 K (42). Low-temperature synthesis of adenine suggests that its formation could be feasible in various cold astrochemical environments. Indeed, HCN is implicated in prebiotic chemistry hypotheses beyond Earth. For example, the diverse nature of HCN reaction products has inspired proposals of a CHN-based biochemistry in oxygen-free worlds, such as that of Titan (70).

HCN chemistry is often studied through polymerization experiments. Along with the formation of larger products, referred to as HCN polymers (often precipitated solid residues), a plethora of compounds form in such experiments. Amino acids, pteridines (which include building blocks of cofactors), sugars, and nucleobases have all been derived from HCN polymerization experiments (69, 71). Detections of amino acids and nucleobases are often reported after hydrolysis of either or both the insoluble and soluble fractions of HCN reaction products (see ref. (69) for a review). I note that hydrolysis reportedly increases the observed yield of nucleobases in such experiments (72). In conclusion, HCN chemistry has been linked to many of the defining molecular features of life (Fig 3.2).

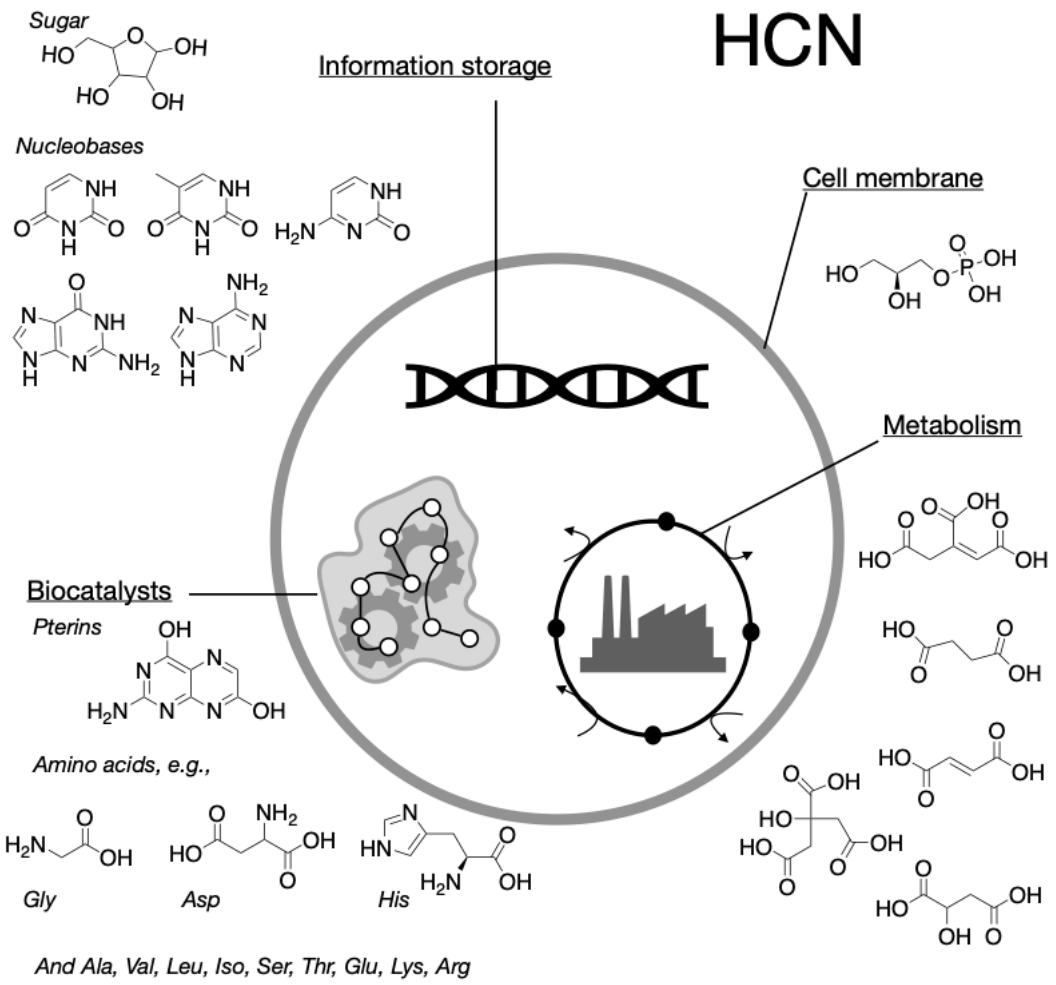

Figure 3.2. Prebiotic HCN chemistry can form building blocks for the central pillars of life. Sugars, nucleobases, and phosphate make out the molecular structure of DNA and RNA. Proteins and cofactors are used in catalysis. Cell membranes are made up of a hydrophobic and hydrophilic part. The head group glycerol-1-phosphate has been formed in HCN based prebiotic chemistry. Carboxylic acids, detected among HCN reaction products, are also metabolites in the citric acid cycle.

3.2.1 HCN on the Early Earth

Atmospheric processes on early Earth are predicted to have formed HCN through photochemistry, shockwave or discharge chemistry caused by coronal mass ejections or high-energy impact events (73–75). HCN produced in the atmosphere would reach the Earth's surface and ocean. The resulting HCN steady state concentration in the ocean has been estimated to 10^{-6} M at 273 K and pH 8 (42). This estimate was based on the assumption by Sagan and Chyba, that the HCN production rate made out a tenth of the total atmospheric production of organics on early Earth (computed as 10^{11} kg/year)(37). The assumed HCN production rate matches recent estimates on the order of 10^{10} kg produced HCN/year on early Earth

(36). The resulting abundance of HCN in the early ocean is not sufficiently high to favour HCN polymerization (<0.01 M). Various suggestions have been proposed for concentrating HCN enough to allow for polymerization on early Earth. Eutectic freezing is one mechanism whereby HCN could become concentrated (74.5 mole% at the eutectic point 250 K) (43). A study simulating the effect of the late heavy bombardment predicted that high-energy impact events could temporarily raise the HCN concentration to a sufficient degree which would allow for polymerization (with an >80 % probability in shallow pools, and > 24 % probability in deeper pools) (38). Since HCN is more volatile than water it cannot concentrate through evaporation.

3.3 HCN as an abundant molecule in the Universe

HCN is one of the more common small organic molecules in the solar system and beyond. Figure 3.3 shows three radically different astrochemical environments where HCN has been detected: the interstellar medium (76), Saturn's moon Titan (77) and comets (78, 79). HCN has also been detected on other planets (80–83) and on Pluto (82), as well as in protoplanetary disks, whose formation precedes that of planets (84). This section briefly introduces these astrochemical environments and their HCN chemistry.

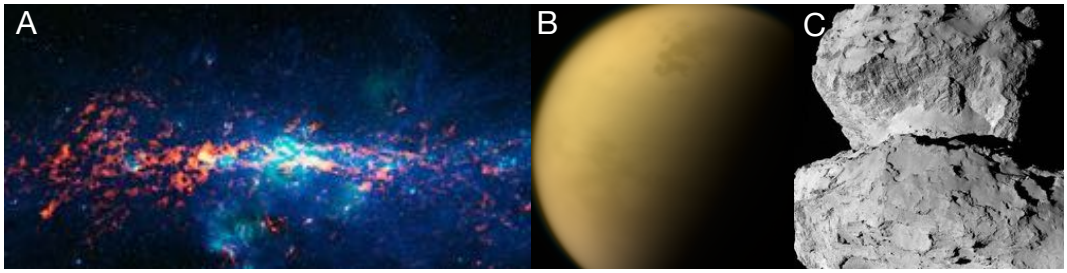

Figure 3.3. Three astrochemical environments where HCN is found. A. The molecular cloud Sagittarus B2 (Image: European Space Observatory(ESO) and NASA) (85) B. Saturn's moon Titan. (Image: NASA/Jet Propulsion Laboratory-Caltech, Space Science Institute) (86) C. The comet 67P/Churyumov-Gerasimenko photographed by the navigational camera onboard Rosetta (Image: European Space Agency (ESA)) (87). Figure reproduced from ref. (58).

3.3.1 HCN and the interstellar medium

The interstellar medium makes out the region between stars in galaxies. Low temperature and low density are two chemical characteristics of the interstellar medium In these regions, the matter is either gaseous or dust like (88). The dust is made up of small solid ice grains. HCN can be found in regions of the interstellar medium called molecular clouds (Figure 3.3A) (89). Molecular clouds are the birthplace of stars, planets, and other celestial bodies such as asteroids and comets. The temperature in dense molecular clouds is below 50 K (88).

Interstellar HCN can form from different high-energy carbon, nitrogen, and hydrogen-containing species (90, 91). In dark (a type of dense) molecular clouds, HCN typically exists at abundance of 10^{-8} relative to molecular hydrogen (90), although the abundance varies. For context, the most common heteroatomic molecules, gaseous carbon monoxide and water ice, both exist at a 10^{-4} abundance (92). The abundance of the high-energy isomer hydrogen isocyanide (HNC) is similar to that of HCN in the interstellar medium (93). It is the high barrier of isomerization and low temperatures which hinder HNC to HCN conversion in the interstellar medium (94).

HCN containing ice analogues have been found to produce glycine and other amino acids upon irradiation (see ref (95) for a review, or ref (96)). Hence, HCN chemistry in ice grains is interesting for prebiotic chemistry. Still, solid HCN has not been detected in the interstellar medium (but has been detected in clouds on Titan (97) and tentatively identified on Neptune's moon Triton (98)).

3.3.2 HCN on Titan

HCN in Titan's atmosphere forms as a photochemical product of nitrogen and methane (99). In the atmosphere of Titan, the volume mixing ratio of HCN is lower at the surface than at 250-300 km altitude (10^{-8}-10^{-6}) (100). Observations suggest that HCN-based clouds exist in the atmosphere and contain HCN ice (97, 101–103). HCN has been predicted to form co-crystals with benzene and acetylene on Titan (104). HCN is also predicted to incorporate in large complex organic materials, referred to as tholins. Sagan and Khare coined the name tholin to refer to organic materials formed when cosmically abundant gases are subjected to light or spark discharge (105).[2] Incorporation of HCN into tholins can be observed during experiments simulating tholin formation (106). However, characterization studies suggest that HCN polymers considerably differ from tholins, and therefore should at best be considered to make up parts of tholins (107, 108). HCN polymers have been predicted to absorb the scarce sunlight available on the surface (109).

3.3.3 HCN in cometary chemistry

Comets are important carriers of organic material in the solar system. Cometary chemistry can be studied both with remote observations and spacecraft missions. Observational studies have found that the surface of comets is covered with organic material (110, 111). The temperature on the surface of comets varies with their distance to the sun but have been measured to ~190-350 K at distances < 4 AU (see summary in ref. (112)).[3] Materials on the cometary surface can sublimate upon heating, forming an envelope made of dust and gaseous molecules around the comet, called the cometary coma.

HCN on comets can either exist as a gas in the coma or as a solid in the cometary ice due to the varying temperature and low-pressure. HCN has been detected in

[2] The other name they considered was the more evoking *star-tar*.
[3] 1 AU corresponds to the average distance between the Earth and the sun.

the coma of comets in an abundance of 0.08-0.25 % relative to water (113). When the comet approaches the sun the HCN abundance increases (114). According to observations, the production rate of HCN in the coma follows a power law, $Q_{HCN} \propto R_H^{-x}$, where R_H is the distance to the sun and x=2-4.7 (79, 115, 116).

Evidence of aqueous alteration of minerals on comets, as well as theoretical models, suggest that the subsurface temperatures of some comets could allow water below the surface to melt (112, 117). However, it is uncertain whether sufficient pressures could be reached to keep the liquid water stable (112). I note that if liquid water does form in the subsurface of comets, aqueous HCN solutions (which melt at lower temperatures than water) might form in the subsurface of comets as well.

Suggestions that HCN polymers exist on cometary surfaces are based on model predictions that HCN and HNC in the cometary coma come from a parent molecule on the comet nucleus (78). HCN polymerization has also been proposed to explain cometary outbursts (118). While HCN polymers have been proposed to form on comets, the most complex molecules that have been identified on comets are not very large, partly owing to the instrumentational setup. For example, the mass spectrometer used in the Rosetta mission to study the comet 67P only measured up to a mass of 100 Da. The largest molecules identified on 67P was ethylene glycol and ethanethiol with a mass of 62 Da (119), as well as toluene which was identified at peaks at 91 and 92 Da.

3.4 HCN-derived polymers

Because of the ubiquity and reactivity of HCN, products derived from it (including polymers) are suspected to contribute to complex organic matter in a range of environments. HCN polymerizes in a variety of conditions into a intricate reaction mixture (see e.g., ref (120)). The heterogenous polymer products have been intensely studied experimentally, and at rarer times, computationally for the past century. Below I provide a brief outline of the results of such efforts. For a complete review, I refer the reader to ref. (121).

The polymerization of HCN can proceed both in aqueous solution, and in pure liquid or gaseous HCN (see review) Attempts have been made to study radiation-induced solid-state polymerization of HCN (59). However, no polymerization was observed prior to melting the irradiated solid.

Factors such as pH, temperature and the presence of catalysts can affect the polymerization process. Aqueous HCN polymerization is known to proceed most efficiently at pH 9.2, the pK_a of HCN (122). Such an aqueous HCN solution has an equal concentration of HCN and cyanide anions. Hence, the base-catalyzed polymerization is sometimes referred to only as cyanide polymerization. In experiments performed in pure liquid HCN the concentration of cyanide anions has been ~ 0.05 – 0.2 M (123, 124). Although in one case, a two-component HCN:NH_3 system was used with an overall 10:1 molar ratio (125). Amines, ammonia,

or alkali hydroxides like NaOH are typically used as bases, while hydrochloric acid is used to lower the pH (126, 127).

There are various other ways to catalyze or initiate HCN polymerization. Mamajanov and Herzfeld have demonstrated catalysis using free radicals (126). Mozhaev et al.(59) and Draganić and colleagues (128–130) have performed radiation induced HCN polymerization. Schwartz and Goverde found that cyanohydrins and glyconitrile accelerate tetramer formation from cyanide solutions (131). Ammonia also causes a faster reaction kinetics (120). Moreover, Ferris et al. have reported that the formation of the HCN tetramer diaminomaleonitrile is catalyzed by bromine, iodine, copper ions, cyanogen and cyanamide, and polymer formation is catalyzed by presence of diaminomaleonitrile (122).

Polymerization in aqueous solution competes with hydrolysis. At concentrations below 0.01 M, hydrolysis is the governing process (122). In cold temperatures, the formation of eutectic phases could spontaneously concentrate HCN, favouring polymerization over hydrolysis (50).

The reactive nature of HCN brings about the question: At how low temperatures can HCN polymers form and over what timescale? Answering this question would provide strong hints as to which different astrochemical environments HCN polymers may be present in. So far, the range of temperatures where HCN polymer formation has been observed is between 195-363 K (72, 126, 127). Experiments performed close to 250 K report tetramer formation, albeit not polymer formation (120). In a one-of-a-kind low temperature experiment at 195 K, a 0.15 M HCN 0.1 M ammoniacal solution was left to react over a period of 27 years (42). Studies of the kinetics of HCN polymerization have also shown that the mechanism appears to change with temperature (132).[4]

3.4.1 HCN-derived molecules: alternative polymerization monomers

A polymerization process proceeds through the successive reaction of molecules referred to as monomers. In the case of the polymer products formed in HCN reaction mixtures, the identity of the monomer that polymerizes is still up for debate. Sometimes, *HCN-derived polymers* is used as a term instead of *HCN polymers*. HCN-derived polymers include polymers of HCN oligomers, or even of the hydrolysis product formamide (133). Figure 3.4 shows a set of HCN-derived molecules that have been proposed as the monomer in the formation of HCN-derived polymers. A presentation of these compounds will be helpful for the following discussion. From this point on, HCN and HCN-derived molecules and polymers will be enumerated.

[4] Text adapted from a previously published licentiate thesis (58).

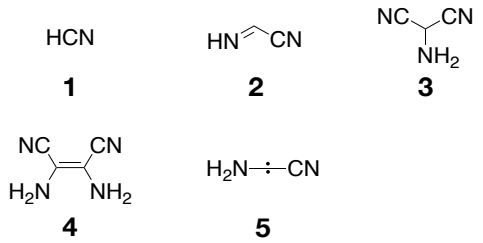

Figure 3.4. Molecules proposed as monomers or intermediates in HCN polymerization: HCN (**1**), iminoacetonitrile (**2**), aminomalononitrile (**3**) and diaminomaleonitrile (**4**) and aminocyanocarbene (**5**).

The HCN oligomers in Figure 3.4 are the dimer iminoacetonitrile (**2**), trimer aminomalononitrile (**3**), and tetramer diaminomaleonitrile (**4**). A second dimer aminocyanocarbene(**5**) is often also discussed because of its reactive nature (59, 124, 134, 135), and because radicals have been detected among the polymer products (136, 137). Early in the studies of HCN chemistry, Oró proposed a mechanism to form adenine (**6**) which included **2** and **3** as intermediates (68). Moreover, these molecules have been proposed to have key roles in the formation of HCN-derived polymers (see ref. (121) for a review). For example, the dimer formation is the proposed rate limiting step in the formation of **4** (138). However, the evidence for the formation of **2** and **3** in HCN reactions is only indirect (122, 138, 139). Out of these molecules only the tetramer has been detected in HCN reaction mixtures besides HCN (138). Moreover, Ferris and colleagues showed that **4** can photochemically form **6** upon UV-irradiation through two photoexcitation events (140, 141). Diaminomaleonitrile is also widely used to form N-heterocycles via Schiff base intermediates (R1N=CR2R3) in organic synthesis (142, 143). Ferris investigated HCN oligomerization and stated that **4** is the main product of base-catalyzed HCN oligomerization (144). We here note, however, that separate polymerization experiments starting from the tetramer do not produce a noteworthy polymer yield close to room temperature, like most HCN polymerization experiments do (126, 127).

Polymerization of the molecules in Figure 3.4 have been conducted over a wide range of temperatures. The dimer **2** is known to polymerize in dichloromethane solution above 233 K (145). The polymerization of **3** has been studied at above room temperature in aqueous solution and in the neat liquid (146–148). The tetramer **4** polymerizes in neat melt above 449 K (149) and neat solid above 398 K (127, 149, 150). Efficient aqueous polymerization of **4** has been observed above 353 K at pH 9.2, and to a smaller extent at room temperature (yield ~1 wt%)(126).

3.4.2 Characterization of HCN-derived polymers

Below I outline some of the experimental characterization studies that have been done on HCN-derived polymers. One of the main challenges with characterizing and resolving the structure of HCN polymers is their low solubility in most solvents. Nevertheless, numerous characterization studies have been performed. Some studies separate the soluble and non-soluble part of the HCN reaction

product before analysis. In a polymerization experiment performed in aqueous solution, the insoluble product yield (based on the carbon input) was 15-46 wt% at 311 K, and 1-12 wt% at 277 K, depending on the reaction conditions (151). Bonnet et al. reportedly were able to nearly completely solvate their reaction product in methanol (125). In addition, He et al. were able to dissolve their polymer solid in DMSO-d_6 (123). Meanwhile, Mizutani et al. studied their polymer product in fractions separated by solubility in DMSO, methanol and ethyl ether (152).

Reaction products formed in a mixture of liquid HCN (10:1 HCN:NH_3 ratio) have been studied with high-resolution mass spectrometry (153). The mass peaks lie between 50-600 m/z and 50-1000 m/z for negative and positive ions, respectively (125). Ruiz-Bermejo and colleagues have pointed out the challenge in recording the mass spectra of the polymer products - the large product masses suggest that the products likely have many charged sites (121). Analysis with liquid chromatography found mass peaks of between 100 and 2000 Da (152) Using electrophoresis, polymer products with masses up to 250 kDa have been detected in the insoluble product fraction of aqueous solutions of NaCN or NH_4CN (at 277-311 K) (151). The same study found masses up to 140 kDa in the soluble fraction. Meanwhile, Dragovic and colleagues measured that 7 % of their polymer product formed during radiation-induced aqueous polymerization had masses in the range of 10 and 22 kDa (128–130).

The HCN-derived polymers have a composition that varies with reaction conditions. The N/C and H/C ratios can vary between 0.7 and 1.4 (121). The N/C ratio decreases slightly at higher temperatures (see ref. (132) for details). In studies of HCN-derived polymers formed from neat solid- or melt diaminomaleonitrile, N/C and H/C ratios of about 0.9 were found (154). The loss of NH_3 and HCN during polymerization was observed with thermal gravimetric mass spectrometry (TGMS) analysis and could explain the unequal molar ratios. The polymer oxygen content has been measured to be between 10 and 25 mole percent when polymerization experiments are conducted in aqueous solution, (126, 132, 155, 156). The incorporated oxygen originates from the aqueous solution and is present whether the reactions are carried out in contact with air or an inert atmosphere (see ref. (121) for a review).

3.4.3 Proposed polymer structures

Several studies have focused on characterizing the structure of HCN polymerization products using a variety of techniques including, but not limited to, vibrational spectroscopy, UV-vis spectroscopy, and nuclear magnetic resonance (NMR) spectroscopy. The characterization reveals that HCN polymers include a variety of functional groups, such as amines (132, 132, 157), amides (132, 158), hydroxyl groups (132), imines (123, 126) carbodiimides, triazines, nitriles, carbonyls (155) to name a few (Figure 3.4A). Different researchers have proposed different polymer structures based on the results of characterization studies. Three such polymer examples are shown in Figure 3.4B. For example, He et al. found using multidimensional NMR spectroscopy that new bonds were formed between carbons and proposed a polyimine structure (structure **7** in Figure 3.4B) (123). Rahm et al. (109) later showed computationally that polyimine has a tunable band gap which would allow the polymer to absorb visible light in the narrow window of wavelengths that reaches the surface of Titan. Mamajanov and Herzfeld (124) found that new bonds were formed between carbon and nitrogen atoms, and proposed structure **8** in Figure 3.4, in agreement with the proposed structure Mozhaev et al. (59). Structure **9** in Figure 3.4 was proposed by Völker as a polymer of the HCN dimer iminoacetonitrile (156, 157). Figure 3.4c shows examples of different the three-dimensional structures of the HCN-derived polymers. These, and several other hypotheses exist for the structure of HCN-derived polymers, and how they form. Given the potential relevance of these structures to prebiotic chemistry it is important to determine which are most likely to form.[5]

[5] Text adapted from a previously published licentiate thesis (58).

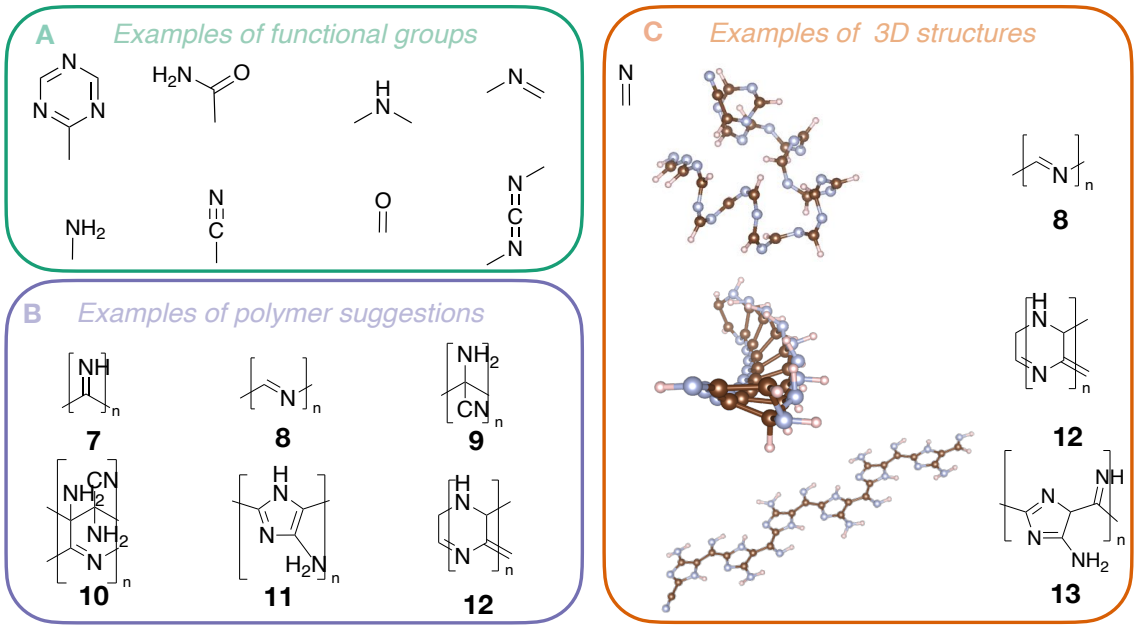

Figure 3.5. A. Examples of functional groups identified through experimental characterization of HCN polymers with methods such as NMR spectroscopy and vibrational spectroscopy. B. Examples of different proposed repeating units proposed for the HCN polymers based on the identified functional groups (here only examples without oxygen). C. Three dimensional structures of small fragments of different HCN polymers. Different repeating units result in both coiled and stretched conformations.

3.5 Summary

This chapter outlined the main motivation for studying the role of HCN in prebiotic chemistry. However, in HCN reaction mixtures biomolecular building blocks make up only a small fraction of a plethora of products. Still, many questions regarding the nature of these materials remain. In Chapter 5 and 6, we investigate HCN chemistry, including proposed products and reaction mechanisms. First, I will briefly introduce the methods we applied in our investigations in Chapter 4.

4
Computational Methods

The key concept behind the work presented in this thesis is *energy*. Energy governs the speed and direction of chemical reactions. A material's energy results from the interactions and kinetic energy of its electrons and atomic nuclei. In theory, the energy of a system can be obtained by solving the Schrödinger equation. However, the exact solution is often not possible to compute. Instead, much of computational chemistry involves finding approximate solutions to the Schrödinger equation for different systems under study. The approximate methods often balance a reduction in computational costs with an acceptable accuracy in the computed energy. This chapter introduces the computational methods which were used to calculate energies in this thesis.

In this chapter, I will use a partition of the total free energy, G, into an electronic contribution, E_{el}, and a term which corresponds to the thermal energy, $E_{thermal}$, as

$$G = E_{el} + E_{thermal}. \quad (4.1)$$

The electronic term refers to the part of the energy that arises due to interactions between electrons and atomic nuclei as well as the kinetic energy of the electrons. The thermal term has both enthalpic and entropic contributions.

4.1 Electronic energy calculations with density functional theory

In this thesis I have used the quantum mechanical method density functional theory (DFT) to compute the electronic energy (159). All DFT calculations were done using Kohn-Sham (KS) DFT (briefly described below). From now on, I refer to KS-DFT only as DFT.

DFT (among other methods) makes use of the Born-Oppenheimer (BO) approximation when computing the electronic energy. This approximation considers the electrons as moving around a frozen configuration of atomic nuclei (160). The BO approximation works well in most cases since atomic nuclei are much heavier and slower than electrons. Within the BO approximation, the problem of finding the energy of the nuclei and electrons can be treated separately. For example, the energy of the electrons, E_e, can be calculated as a solution to the time-independent Schrödinger equation

$$\mathrm{H}\Psi = E_e\Psi, \tag{4.2}$$

where Ψ is the electrons' wave function and H is a Hamiltonian that only considers the kinetic energy and interactions of the electrons. The energy and wave function in Eq. (4.2) can be found by applying the variational principle which states that the true ground state wave function corresponds to the one with the lowest energy. Nuclear-nuclear repulsions between atomic nuclei can be added to the electrons' energy to obtain the total electronic energy. As a result of the BO approximation, the electron energy can be computed separately in DFT as well, but not via the Schrödinger equation (Eq. 4.2). Below I briefly explain how.

The foundation of DFT is made up of two key publications. In the first, Hohenberg and Kohn proved that it is possible to compute the energy from the electron density of a system using the variational principle (161). In DFT, the energy is *a functional* of the spatially dependent electron density (i.e., it is a function of a function). The main challenge of the functional formulation in DFT is to describe the kinetic energy and electron-electron interactions. In a successful attempt to overcome this obstacle, Kohn and Sham showed that it is possible to model the electrons as non-interacting particles moving in an effective potential (162). The kinetic energy is then computed for a set of non-interacting particles. The functional formulation of the energy, $\mathrm{E_e}[\rho]$, in terms of the electron density, ρ, can be expressed as,

$$\mathrm{E_e}[\rho] = \mathrm{T}[\rho] + \int \mathrm{d}^3\mathrm{r}\rho(\mathbf{r})\mathrm{V_{ext}}(\mathbf{r}) + \frac{1}{2}\int \mathrm{d}^3\mathrm{r}\rho(\mathbf{r})\,\mathrm{V_H}(\mathbf{r}) + \mathrm{E_{xc}}[\rho], \tag{4.3}$$

where **r** is a positional vector. The first term in Eq. (4.3) corresponds to the kinetic energy, T[ρ], calculated for non-interacting electrons (using an orbital formulation). The effective potential functional corresponds to the remaining three terms: contribution to the energy from interactions between nuclei and electrons, represented by the external potential, V_{ext}; electrostatic electron-electron interactions represented by the Hartree potential, V_H; and the exchange-correlation functional, $E_{xc}[\rho]$. The form of the exchange-correlation functional in Eq. (4.3) is unknown. The term accounts for the correlated movement of electrons beyond the mean-field approximation. In theory, DFT is a relatively computationally cheap method because it can simplify the evaluation of the electron-electron interactions, in particular the exchange and correlation terms, compared to other methods (163). However, there are many variations of DFT which vary in computational cost.[6]

There exist several rungs of DFT which differ in the way they model the exchange-correlation term. Two of the most common types are generalized gradient approximation (GGA) DFT and hybrid DFT (164). GGA functionals evaluate the exchange-correlation as a function of the density and its gradient (165). Meanwhile, hybrid DFT functionals include an empirical amount of Hartree-Fock exchange in the computation of the exchange-correlation (163). For this reason, hybrid functionals are relatively computationally expensive to use ($O(N^4)$ scaling compared to $O(N^3)$) (164). In general, hybrid DFT functionals are more accurate than GGA functionals.

The required accuracy of a computational method depends on the research question. To be able to draw quantitative conclusions regarding reaction rates and chemical equilibria, the accuracy needs to be ~1 kcal/mol, which matches that typically obtainable in standard thermochemistry experiments (166). The accuracy of DFT can range from one to several kcal/mol depending on the level of theory and application (for a few different benchmarks, see refs.(167–171)). In summary, the choice of DFT method affects the accuracy and computational cost of the energy calculation, so too does the description of the electron wavefunctions which are used to construct the electron density.

The total electron density can be represented as the sum of fictitious KS-one-electron densities. The one-electron densities in turn are computed as the square of the one-electron wave functions, ψ_j. Taken together, the total electron density, ρ, can be expressed as

$$\rho = \sum_j |\psi_j|^2. \tag{4.4}$$

The linear combination of atomic orbitals (LCAO) method is used to represent the one-electron wave function of an electron in a molecule as a sum of atomic orbitals ϕ_i with coefficients, c_{ji}

[6] Text adapted from a previously published licentiate thesis (58).

$$\psi_j = \sum_i c_{ji}\phi_i \quad (4.5)$$

The atomic orbitals are typically also modelled using a linear combination of gaussian functions or slater functions. The appearance of this linear combination is called the basis set. As mentioned above, the choice of basis set affects the accuracy of the energy computation. A larger basis set provides higher accuracy but comes with a higher computational cost. According to the variational principle the optimal coefficients, c_{ji}, are those which minimize the energy. The states of the core electrons of atoms are typically assumed to be frozen to reduce computational cost without meaningful loss in accuracy. An alternative to using atom centered basis functions to construct the basis set is to use plane waves instead. The plane waves are especially well suited to describe electrons in calculations using periodic boundary conditions.

The electron density of a periodic system, such as a polymer or a crystal, can be described by the electronic density in just one unit cell. The wave functions of electrons in periodic materials can be represented as a sum of plane waves (in accordance with the Bloch theorem), according to

$$\psi_j = \sum_k c_{jk} e^{-i\mathbf{k}\cdot\mathbf{r}} \quad (4.6)$$

where c_{jk} is a coefficient and $\mathbf{r}$ is a positional vector in the crystal lattice. The plane wave periodicity is determined by the wave vector $\mathbf{k}$. Periodic calculations typically use pseudopotentials (172) or projected augmented waves (173) to reduce the number of plane waves. In periodic DFT, the size of the basis set and the accuracy are controlled by an energy cutoff for the plane-wave kinetic energy.

Long-range interactions within a crystal are typically computed in the reciprocal space (the Fourier transform of the crystal lattice). When doing so the energy is evaluated through an integration in reciprocal space. In software, integrals are replaced by sums which are evaluated on a *k*-point grid. The quality of such calculations needs to be converged with respect to k-point density to ensure sufficient accuracy.

4.2 Geometry optimizations

The energy of a material depends on the positions of its atomic nuclei. We typically search for the structure of the material at zero Kelvin (with no zero-point energy) because it corresponds to a well-defined point on the multidimensional potential energy surface. Thermal effects corresponding to a particular temperature are added after the geometry optimization, as described below. Reactant and product states correspond to minima on the multidimensional energy surface (Fig. 4.1). A transition state of a chemical reaction is a first-order saddle point (i.e., a minimum in all but one direction in which it is a maximum). A minimum or saddle point can be found through geometry optimizations. The optimization algorithm typically involves computation of the positional derivatives of the energy (forces). The endpoint of the optimization often corresponds to a local minimum and depends on the input guess structure. The global minimum could be found by performing many optimizations with different input structures, a work-intensive task depending on the size of the system under study. After optimization and calculation of thermal effects, the relative free energy of two structures can measured by their energy difference (See Fig 4.1).

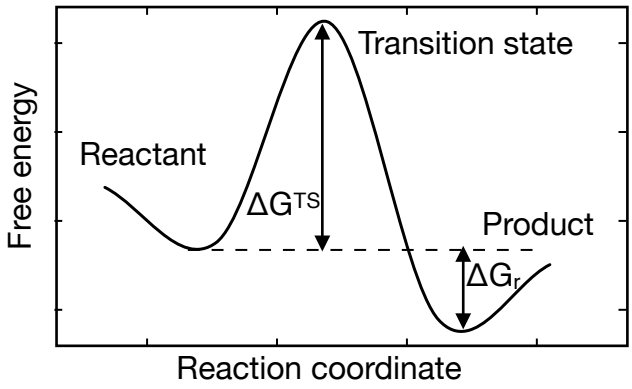

Figure 4.1 Schematic of an energy profile of a reaction, visualized in one dimension. The reactant and product correspond to minima on the potential energy surface. In a general number of n-dimensions, the transition state is a first-order saddle-point (a minimum in all coordinates but one for which it is a maximum). ΔG_r and ΔG^{TS} correspond to the reaction energy and reaction barrier, respectively. Figure adapted from ref. (58).

4.3 Calculating thermal effects in molecules and solids

The part of the total energy made up of thermal effects is calculated with the help of statistical thermodynamics. Solids, molecules, and atoms have the possibility of storing energy in their degrees of freedom. The degrees of freedom can either be translational, rotational, or vibrational in type. The total thermal energy term, $E_{thermal}$, can be written as a sum of all types of contributions

$$E_{thermal} = E_{vibration} + E_{rotation} + E_{translation}, \quad (4.7)$$

where E_i is the thermal energy contribution for degree of freedom i. An electronic contribution needs to be added to Eq. (4.7) for systems with degenerate electronic ground states. There is both an enthalpic, H_i, and entropic contribution, S_i, to the total thermal energy contribution, according to

$$E_i = H_i - TS_i, \quad (4.8)$$

where T is the temperature. The enthalpic and entropic contributions from a degree of freedom can be calculated from the corresponding partition function. For example, constructing the partition function for one vibrational degree of freedom first requires computing the force constant associated with that vibrational mode. Approximations such as the rigid rotor and harmonic approximation are used to treat rotations and vibrations, respectively.

The approach to compute thermal effects varies depending on what state or system is modelled. A molecule has all types of degrees of freedom, an atom only has translational ones, while a solid only has vibrational ones. The vibrational degrees of freedom in a solid are also called phonons. Computational calculation of phonons was done in Paper I and IV using the finite displacement method (174). This method relies on multiple energy evaluations in which the system is displaced in different ways depending on its symmetry group.

4.4 Modelling solvent effects

The reactions studied in this thesis are considered in a solvent environment, whether it be HCN, water or methane. A solvent influences the potential energy surface of a solute through both enthalpic and entropic effects. These effects depend on the specific interaction between the solute and solvent. As such, solvent effects influence the free energy profile of a reaction.

In this thesis, I used the implicit solvent model called the polarizable continuum model (PCM) to represent the solvent around static molecular models (175). In PCM, the solvent is represented by a dielectric medium surrounding a solute cavity. The shape of the solute cavity depends on the size of the solvent molecules (modelled as spheres). Specific solvents can be defined by the size of their solvent spheres and by their static and dynamic dielectric constants. Implicit models can also consider different parameters but those will not be discussed here (see e.g., ref. (176)). Implicit solvent models consider electrostatic and dispersion effects as well as the energy loss associated with making the solute cavity (177). HCN is not a commonly used solvent and does not exist among the available solvents in the software Gaussian 16 rev. B.01 (178) used in this thesis. I made an approximate PCM of HCN by adjusting the static dielectric constant of the water model to that of HCN (144.8 at 278 K (51)). All solvated species were computed in a 1 M concentration. Meanwhile, concentration of HCN in the neat liquid at 278 K is 26.2 M (49).

Implicit models cannot model direct coordination of the solvent with the solute, via e.g., hydrogen bonds. Solvation energies computed with implicit models like PCM have mean unsigned errors of 1-6 kcal/mol and 7-10 kcal/mol for neutral and charged species, respectively (176). However, thanks to error cancellation, *relative energies* of solvated molecules can be expected to be more accurate.

Explicit consideration of solvent molecules provides an alternative to implicit models. In an explicit model, solvent molecules are added around the solute, thereby modelling the solvent-solute interactions more accurately. It is possible to combine explicit and implicit solvation by using a few explicit solvent molecules and an implicit solvent model. However, a solvent is dynamic and consists of many solvent molecules which can arrange around the solute in an entire ensemble of solvent configurations. Without some type of configurational sampling there is no guarantee of the accuracy of the profile. Hence, configurational sampling is key when studying reactions in solutions (179).

4.5 From simulations to system properties

In this thesis I used atomistic simulations to explicitly study solute-solvent interactions. *Atomistic simulations* can generate a statistical ensemble of configurations by assuming the *ergodic hypothesis*. According to the ergodic hypothesis, the statistical ensemble overlaps with the simulation ensemble if enough time passes during the simulation. Therefore, the partition function (and thermodynamic properties) of a solvated system is theoretically possible to compute from a very long simulation. However, computational costs associated with simulations limit the timescale to below microseconds or nanoseconds, depending on the available resources.

Molecular dynamics is one type of simulation technique which models the time-evolution of a system. In molecular dynamics, energies and forces are computed with some method, classical or quantum chemical (e.g., DFT). I used ab initio BO molecular dynamics based on DFT in Papers II-IV. In BO molecular dynamics, the energy and forces are evaluated at each time step. The movement of the atoms between each step, follows Newtonian dynamics, with positions being updated based on atomic velocities and the forces acting on the atoms. The size of the time step is limited by the fastest motions in the system (typically covalent bonds to hydrogen atoms) and is often on the order of femtoseconds (180).

I simulated all systems studied with molecular dynamics in this thesis in a canonical NVT ensemble. A thermostat coupled to the motion of the atoms regulates the temperature of the system. For example, in Paper II and III the steered simulations were run using the canonical sampling through velocity rescaling thermostat (181). This thermostat regulates the temperature by velocity rescaling and ensures sampling of a canonical ensemble by introducing noise to the equations of motion.

The free energy change computed in an NVT ensemble is a Helmholtz free energy. However, in chemistry we typically are interested in Gibbs free energies, computed at constant pressure. Nevertheless, the difference between the Helmholtz reaction free energies and the Gibbs reaction free energies for condensed phase systems is small.[7] In this thesis, I refer to all free energies (also Helmholtz) as Gibbs free energies, or simply free energies.

4.6 Enhanced sampling methods

Proper sampling in any type of simulation is difficult to obtain if the system is stuck in a local minimum. Constructing reaction free energy profiles requires sampling regions of the multidimensional free energy surface which correspond to reactants, products, and transition states. Transitions between near-by minima, such as reactants and products of a reaction, are rare-events. Enhanced sampling methods make it possible to sample high-energy regions. In enhanced sampling methods, the multidimensional free energy surface is projected onto a small set of (often one or two) variables called collective variables. A choice of collective variables can be one that describes a reaction coordinate, bond (or coordination) topology or configurational parameters such as bond distances and angles. The free energy profile is dependent on a good choice of collective variables (see ref.(182)). Path collective variables, called s and z, were designed as a general set of collective variables (183, 184) . The s and z variables are constructed based on a characterization of two or more reference points. For example, one set of reference points could be the reactant and product of a reaction. Simply put, the s variable describes the path between the reference states, while the z variable describes the distance from the path between them (see ref. (184) for a more detailed description). The position of the reference states in s, z – space is determined through a distance function and a parameter called λ. The distance function can be described in many ways. For example, the topological path collective variables used in this thesis have a distance function that is based on coordination numbers between groups of atoms (185). In the work presented in this thesis, I have made use of path collective variables and two enhanced sampling methods: metadynamics and umbrella sampling.

[7] Differences between the Helmholtz and the Gibbs free energy of reaction are due to changes in the PV term, which for reactions in condensed phase systems is small compared to the change in internal energy.

4.6.1 Metadynamics

Metadynamics is a technique which biases the system to explore new areas of the potential energy surface (for a review see (183)). In a metadynamics simulation, a bias is deposited around the region of the system in collective variable space. The deposition occurs with a certain set frequency. The added bias results in a flattening of the energy surface with time, which encourages a broader sampling. The time-evolution of the deposited bias is illustrated in Figure 4.2. When the system is freely diffusing across the potential energy surface, the simulation is converged. At that point, the deposited bias has the inverse shape of the free energy landscape. Yet, converging the free energy surface this way is time-consuming and computationally costly. In Paper II, I used metadynamics for reaction exploration only.

Metadynamics is suitable for exploration of many competing reactions or the configurational space of molecules. During reaction exploration, the objective is to explore the collective variable space. It is possible to observe transitions between different reactant and product states within the same simulation – provided the simulation is long enough. The transition state can be identified through a statistical method known as committor analysis (186). When a simulation is started from the transition state, there is a 50:50 chance of ending up in either the product or reactant basin.

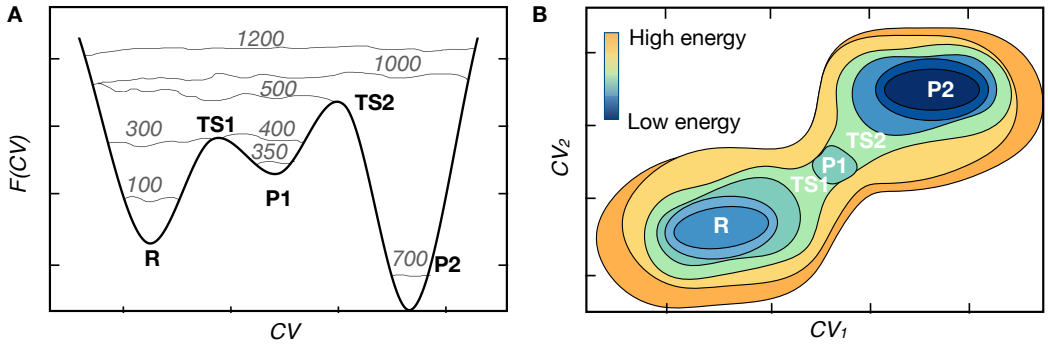

Figure 4.2 A. Schematic of snapshots of the progression of the energy (F(CV)) landscape in a metadynamics simulation in a one-dimensional potential as a function of a collective variable (CV). The free energy landscape is successively filled with deposited bias. Starting from the basin to the left, the system gradually explores states of higher energy until it ends up in another basin. The lines in the diagram represent snapshots of the bias at different times (illustrated by numbers above line) during the simulation. B. A multidimensional free energy landscape projected onto two collective variables. White regions represent areas without a free energy estimate. Often more than one collective variable is used in a metadynamics simulation. Figure adapted from ref. (58).

4.6.2 Umbrella sampling

In Paper II and III, I used umbrella sampling to obtain free energy profiles of reactions taking place in liquid HCN. Umbrella sampling is another enhanced molecular dynamics method (187). During umbrella sampling simulations, the system is constrained at different points along a reaction coordinate. As opposed to metadynamics, umbrella sampling requires a pre-existing knowledge of the reaction and reaction coordinate one wishes to study.

The free energy profile, ΔG, in umbrella sampling is calculated based on the relative probabilities of different states along the reaction coordinate. The free energy, ΔG, is calculated based on a probability distribution, $p(r)$, as

$$\Delta G(r) = -k_B T \, ln \, p(r) \qquad (4.3)$$

where k_B is Boltzmann's constant and T is the temperature. The probability distribution describes the system's position along the reaction coordinate and is obtained by gathering statistics from molecular dynamics simulations.

Some areas of the reaction path are more difficult to sample because they lie far from minima. Umbrella sampling uses a biased probability distribution to ensure that the whole reaction path is sampled. The added bias is in the form of a harmonic potential which constrains the position of the system. The potential is of the form

$$V(r) = k(r - r_0)^2, \qquad (4.4)$$

where k is a force constant, r is the reaction coordinate and r_0 is the center of the potential. Sampling the whole reaction path requires several parallel simulations constrained different r_0 values.

Statistical errors in the free energy profile obtained through umbrella sampling simulations arise due to insufficient sampling. Block-averaging is a method used to estimate the statistical uncertainty of the energy profile (188). The method is based on creating many different sets of system configurations from the biased simulations. Each subset of configurations is used to generate a free energy profile. The standard deviation of the free energy profiles obtained from the different subsets is used as a measure of the uncertainty in the energy.

Umbrella sampling of a reaction is computationally costly. The typical simulation time needed to obtain a free energy profile for a reaction using ab initio molecular dynamics can be 400-500 ps. In Section 6.3, I compare the effect of studying reactions in implicit solvent and simulations.

4.6.3 On using the PBE functional for atomistic simulations

In Paper II and III we made use of ab initio molecular dynamics simulations to study reactions in a dynamic solvent environment. All simulations referred to in Chapter 6 were carried out in a canonical ensemble at 278 K. The temperature is close to the center of the temperature span where HCN is liquid and matches that

used in a polymerization experiment performed in liquid HCN (124). The simulations were performed with the Perdew–Burke-Ernzerhof (PBE) (165) functional and Grimme's D3 dispersion corrections (189) (see Paper II and III) with BO molecular dynamics in CP2K v6.1 (190).

The well-known PBE functional was chosen because of its relative speed as a GGA functional compared to the more accurate hybrid DFT methods (see Section 4.1). Previous theoretical studies have shown that this level of theory correctly reproduces the experimental dipole moment of HCN (54). Yet, the PBE functional is known to underestimate barrier heights (168, 171). Therefore, we consider the known errors of the PBE functional, or compute error corrections, when discussing the implied reaction rates of computed activation energies. The mean absolute deviations are 3 and 4 kcal/mol for nucleophilic attacks and proton transfer respectively (170, 191). The composite error (assuming the two processes as independent) is 5 kcal/mol.

5
The Thermodynamic Stability of HCN-Derived Molecules and Polymers

Experimental studies of HCN reaction mixtures over the past century have resulted in a plenitude of proposed polymer structures and polymerization pathways. To this day, no single polymerization pathway or polymer product is agreed upon. This chapter explores the structure of HCN-derived molecules and polymers and is based on the work of Paper I. Our quest to elucidate the relative energies of HCN-based chemical structures is motivated by outstanding answers in a series of important questions: W*hich HCN-derived molecules and polymers form in astrochemical environments? Which molecules may constitute monomers in various hypothetical polymerization processes? How does temperature affect the spontaneity of these reactions?*

5.1 Evaluating the spontaneity of forming HCN-derived molecules and polymers

In a step towards untangling the many polymerization hypotheses found in literature, we applied quantum chemistry to map the thermodynamic landscape of HCN-derived molecules and polymers. According to the second law of thermodynamics, spontaneous chemical processes must be thermodynamically favoured (i.e., result in a negative change of the free energy). As such, a molecule or polymer formed from HCN should also be energetically downhill compared to HCN. Figure 5.1 shows our free energy map for HCN-derived molecules and polymers. From this point on I discuss Gibbs free energies unless otherwise stated. The relative position of two structures in the map indicates which one is more stable. If one omits consideration of reaction kinetics, then the relative free

energies can be translated into relative probabilities of HCN reacting to form different products, via the Boltzmann distribution.

We included 7 molecules and 24 different polymers in our study (a two-dimensional and a three-dimensional polymer were also included in Paper I but are not discussed here). The polymer set only consists of structures with an H,C,N 1:1:1 molar ratio. For such materials it is possible to evaluate the energy of an HCN unit in the structure which we then use to compare thermodynamic stability. Our selection of materials is not exhaustive, as the structural space of HCN-derived molecules and polymers is vast. Instead, we include molecules that are implicated in various prebiotic HCN reactions and polymerizations, as well as proposed polymers with a wide variety of structural motifs. Refer to Paper I for details on the studied set of structures and pathways.

We focus on the thermodynamic stability of HCN-based materials in aqueous solution at 298 K. Although HCN-derived polymers may form in a wide variety of environments, an aqueous solution and (Earth) ambient conditions are common in experiments. In what follows, I focus on the key takeaways and only discuss a subset of the studied materials using the nomenclature shown in Figure 5.1 (refer to Paper I for further details).

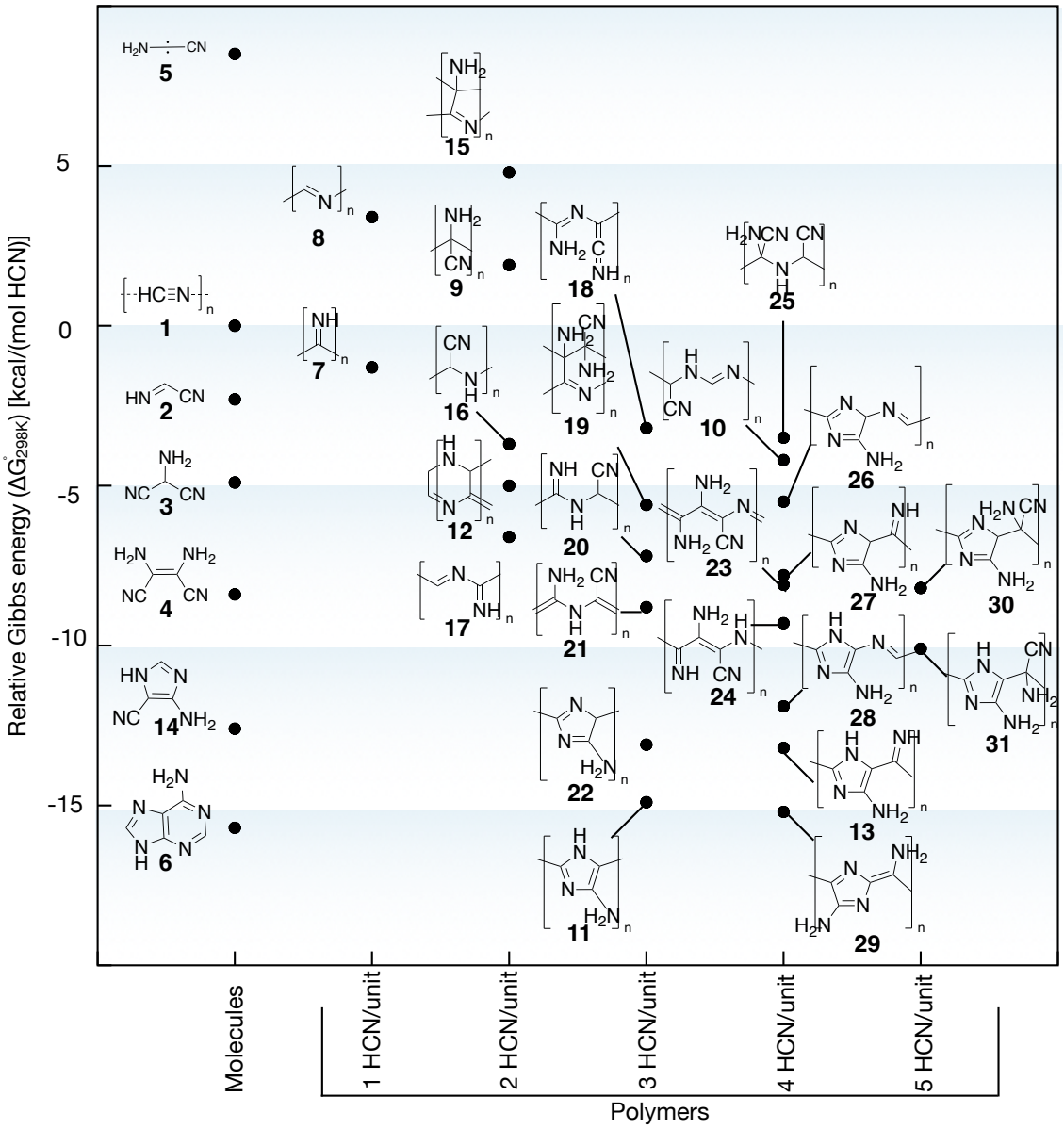

Figure 5.1. The thermodynamic landscape of HCN-derived molecules and polymers in aqueous solution at 298 K (reproduced from Paper I). Structures are organized by the number of HCN molecules in their smallest repeating unit. Relative Gibbs free energies were evaluated at the B3LYP-D3/aug-cc-pVDZ//B3LYP-D3/cc-pVDZ level of theory. The effects of water were described using a polarizable continuum model.

HCN's high-energy content, and potential reactivity, is evident in Figure 5.1. Almost all molecules and polymers lie below the small nitrile in energy. The relative free energy of the molecules and polymers varies between -15.7—+8.5 kcal/(mol HCN) relative to HCN. **11** and **29** are predicted to be the most stable out of all studied polymers. Only adenine (**6**) computes as more stable. Meanwhile, two of the more simple polymers **7** and **8**, lie very close in energy to HCN (-1.4 kcal/(mol HCN) and +3.4 kcal/(mol HCN) respectively). The spontaneity of their formation is likely sensitive to temperature (see section 5.2).

Experimental measurements of the heat released during HCN polymerization ($-\Delta H^o$) can be compared with our computed relative enthalpies of the polymers (see Paper I). Mozhaev et al. measured the heat released during radiation-induced HCN polymerization in the neat liquid to be 10.5 ± 1.4 kcal/(mol HCN) at ambient conditions (59). 6 out of the 24 (**12**, **17**, **19**, **20**, **23**, **26**) studied polymers have computed relative enthalpies of -10.5 ± 1.4 kcal/(mol HCN) at 298 K. The most stable polymers according to our calculations, **11** and **29**, lie below this interval with relative enthalpies of -19 kcal/(mol HCN) relative to HCN.

As the most stable structure in Figure 5.1, **6** should be the major product of HCN self-reactions under thermodynamic control. Yet **6** has a known low yield in HCN reaction mixtures in a range of temperatures (195-363 K, see ref. (69) for a review). These disparate results can have a few different explanations. One discernible possibility is that there exist additional structures, which are more stable than **6**, that were not included in this study. For example, condensation reactions, giving off ammonia to form other molecules and polymers, could yield more stable products. A second possibility is that formation of **6** is kinetically unfavoured compared to the formation of other products. The dichotomy between the stability of **6** and the low yield of the nucleobase in HCN reactions merits further investigation.

One use of our thermodynamics map is for classifying proposed polymerization pathways as thermodynamically favoured or unfavoured. For example, aminocyanocarbene (**5**) has a central role in mechanistic suggestions for radical-initiated polymerization of HCN (59, 124, 134). Another important proposed pathway is the formation of **5** from aminomalononitrile (**3**) or diaminomaleonitrile (**4**) (146, 192). This reaction has been invoked when suggesting that **5** is the monomer unit of polymer **20** (146, 192, 193). In Figure 5.1, **5** lies 8.5 kcal/mol higher in energy than HCN. The instability of **5** implies that the carbene is unlikely to act as an intermediate during polymerization. It is also unlikely that **3** or **4** would spontaneously form **5**. Nevertheless, it cannot be ruled out that at higher-temperatures, trace amounts of the carbene might form, allowing it to play a role as initiator in polymerization mechanisms (59, 124).

Polymer structures **9** and **15** were proposed by Völker (157) and Umemoto (156), respectively. These polymers lie higher in energy than HCN as well as iminoacetonitrile (**2**) which was proposed as a monomer. Owing to the high predicted energy of **9** and **15**, we can conclude that they should not form

spontaneously at 298 K. On the other hand, one of Völker other propositions, that the dimer **2** polymerizes to **10**, is estimated to be thermodynamically favoured. Others have suggested that *cis*- or *trans*-**4** forms **9**, **10** or **15**. Figure 5.1 shows that also these suggestions are unlikely to be true, based on the predicted thermodynamics. However, the thermodynamic map reveals much more about the chemistry of **4**.

4 is among the more stable structures in Figure 5.1. The tetramer's stability limits its possible polymerization products to only 8 polymers out of the 26 studied. The most stable polymers **4** *could* form spontaneously is **11** (proposed in ref. (127)) and **29** (proposed in ref. (121)). We have calculated the associated enthalpy of **4**→**11** and **4**→**29** to both be ~ -10 kcal/(mol HCN). These computed energies are larger than the experimentally measured value in neat the neat solid or melt at 423−463 K of ~3 −5 kcal/(mol HCN) (154). Thus, whereas **11** and **29** are the most likely polymer products of **4** judging by their stability, the other 6 plausible polymer products of **4** can tested for a better agreement with experiments. 3 out of the 6 polymers (besides **11** and **29**) have N-heterocyclic rings. These polymers have a relative enthalpy of ~ -8 − -5 kcal/(mol HCN) compared to **4**, close to the experimental values. Meanwhile, **21** and **24** are the only linear polymers studied that are stable with respect to **4**. These linear polymers have relative enthalpies of ~ -5 and -4 kcal/(mol HCN) compared to **4** respectively, matching the experimental values well.

Figure 5.2 shows proposed pathways from **4** via **21** and **24** intermediates. For example, polymers **21** and **25** have been proposed as intermediates in the formation of **11** and **20** (Fig. 5.2) (121, 149). In prebiotic chemistry, **20** is of special interest owing to Matthews and colleagues (193) who proposed the polymer as an intermediate in the formation of heteropolypeptides in HCN reaction mixtures – a controversial suggestion (194). In the **25** → **21** step, HCN is eliminated, as also observed in experiments (127, 149, 150, 154). Our calculations suggest that formation of **20** from **4** is slightly unfavoured (+1.2 kcal/mol). The equilibrium between **20** and **21** is expected to be shifted towards **21** (+1.5 kcal/mol). On the other hand, the **21** → **11** reaction is spontaneous. The **4**→**29** pathway via **24** (Fig 5.2) also computes as thermodynamically favoured. In summary, our computations support suggestions that **4** can polymerize to form **11** via intermediates **21** and **25**, and **29** via **23** and **24** intermediates. I conclude the topic of the potential reactivity of **4** by stating that further investigations into the reactivity of HCN with **4** are needed.

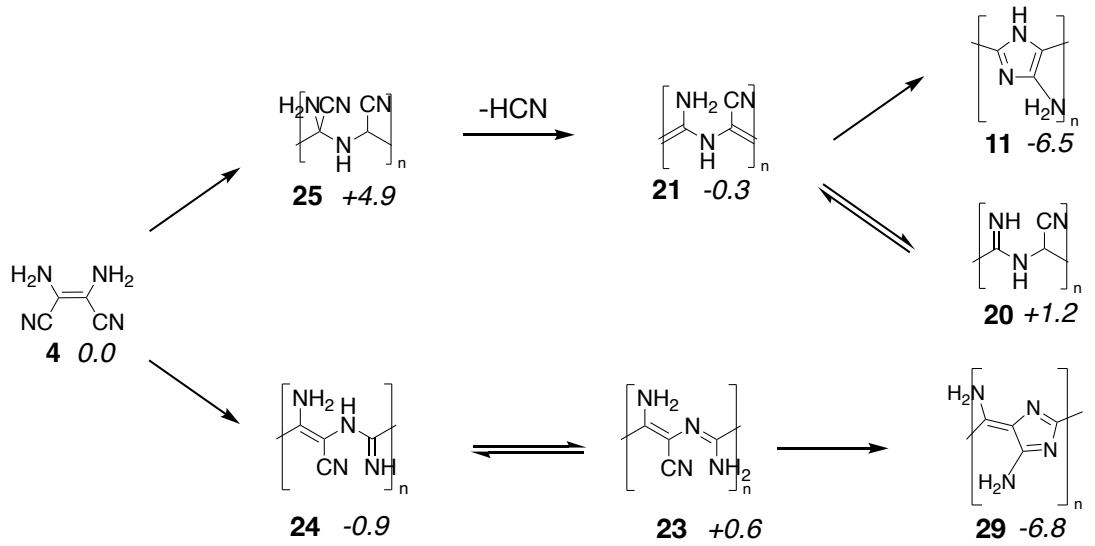

Figure 5.2. Proposed pathways of polymerization of diaminomaleonitrile (**4**). The free energy per HCN of each structure relative to **4** at 298 K is shown in italics. Polymers **21** and **24** have relative enthalpies which match the experimentally measured heat of polymerization of **4**. Meanwhile, the polymers **11** and **29** are the most stable products out of those investigated.

5.2 Temperature effects in the thermodynamic map

Our interest in HCN chemistry in a whole host of different astrochemical environments calls for considerations of temperature effects on the thermodynamic map. Polymerization reactions are generally unfavoured due to entropic effects. The reason is the loss of translational and rotational degrees of freedom during polymerization. The rotational and translational entropies are the main contributions to the total entropy correction of the free energy at ambient or colder conditions. At high temperatures this entropic correction ($-T\Delta S$) is magnified.

In Figure 5.3 the temperature effect is visualized by counting the number of thermodynamically favoured polymers for each HCN-derived molecule at 0 and 298 K, respectively. Two general trends can be observed. The first is that HCN-derived molecules become more stable per HCN as their size increases, reducing the number of possible products as a result. The second observable trend is that the number of viable polymer products of the molecules decreases with increasing temperature, as expected, and described above. The difference is largest for **4** which may transform into 16 possible polymer products at 0 K compared to only 8 at 298 K. Intermediate temperatures lie in-between the values at 0 and 298 K. Importantly, **11** and **29** are more stable than **6** at 0 K.

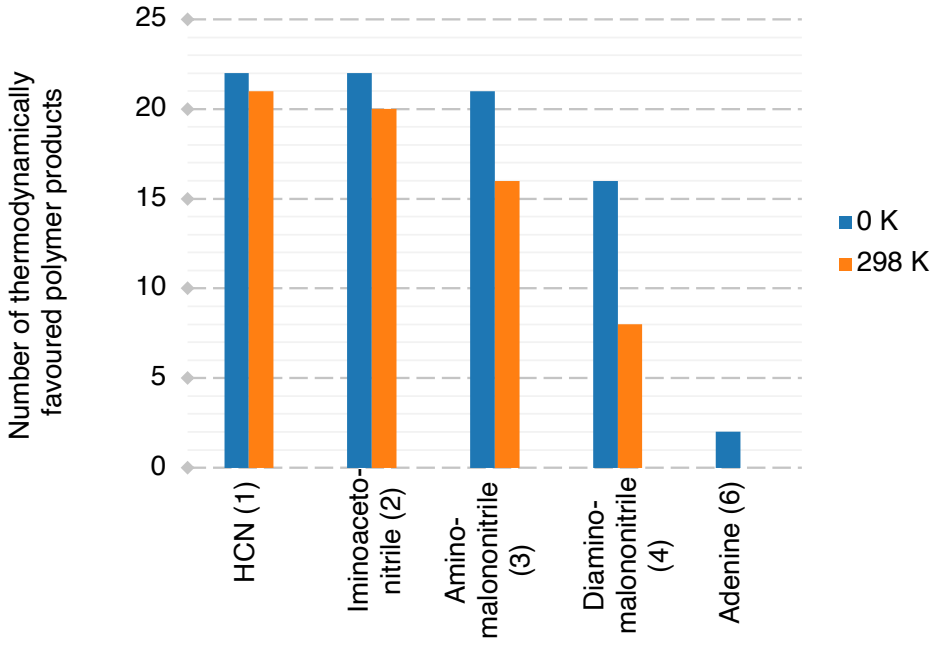

Figure 5.3. The HCN-derived molecules become increasingly stable with their size. Consequently, the number of polymer structures that can spontaneously form from such molecules also decreases with its size (data from Paper I). The effects vary with temperature. In total 24 polymers were included in the study.

The effect of temperature on a polymer's stability is well exemplified by studying polymer **7** (polyimine) (Fig. 5.4). At 298 K, **7** is slightly favoured by -1.4 kcal/mol compared to hydrogen bonded HCN. At ambient conditions, **2** – the monomer of **7** – is predicted to be *more* stable than **7**. On the other hand, **7** is the more stable structure at temperatures below the eutectic point of aqueous HCN solutions, such as temperatures found in the atmosphere and on the surface of Titan, as well as on some comets in the inner solar system (see Chapter 1 and 3). In Chapter 6 we study the growth of small chains of **7** in more detail.

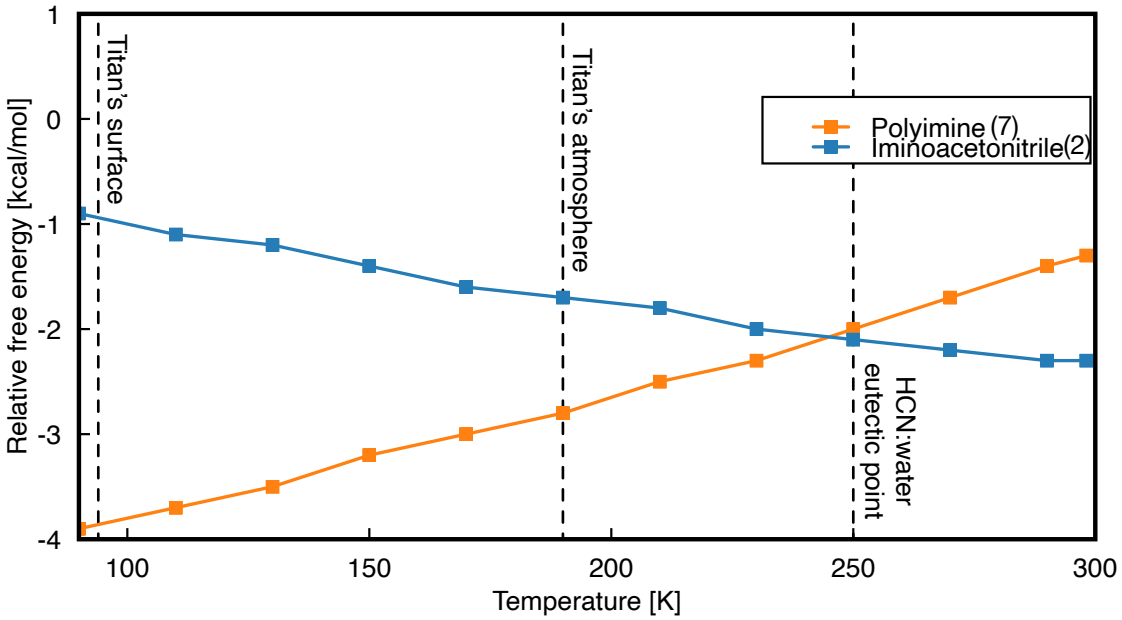

Figure 5.4. The temperature dependence of the free energy content of polyimine (**7**) and iminoacetonitrile (**2**) relative to HCN. The point where the lines intersect signifies a shift in relative stabilities.

5.3 Summary

The structural space of possible HCN-derived polymers is vast. In part, the lack of clarity regarding the structure of the formed HCN-derived polymer stems from the reactive nature of HCN, which makes many reaction processes favorable. Indeed, the formed polymer product is a complex and heterogenous one (e.g., ref (121)). Nevertheless, the long history of HCN polymer research has generated many suggestions of product structures and polymerization pathways. The thermodynamics map presented in this chapter is a tool for evaluating polymerization pathways. For example, the HCN tetramer **4** - central to the study of HCN chemistry - was found to be very stable with respect to most polymers. Our results suggest that the HCN polymerization products may consist of **11** or **29** under thermodynamic control (high temperature). Moreover, **6** is computed as the most stable HCN-derived structure. Unfavourable kinetics could account for the low yield of **6** in HCN reactions. If so, there may exist catalysts that favour the nucleobase's formation, over other products, in prebiotic environments. Searching for such catalysts is one possible future direction of research.

Temperature has a clear effect on the space of possible polymerization products. At low temperatures, such as those on Titan, the space of thermodynamically favoured products is larger. This chapter has shown which structures are thermodynamically allowed to form. However, polymer formation is also governed by reaction kinetics, the topic of next chapter.

6
The Reactivity of HCN and HCN-Derived Molecules

The HCN tetramer diaminomaleonitrile (**4**) is a stable product relative to most HCN-derived molecules and polymers, with thermodynamically feasible pathways to the most stable studied polymer structures (see Chapter 5). Thus, it would be reasonable to suggest that formation of more complex HCN-derived molecules and polymers occurs in two steps, first HCN forms **4** which then converts to more stable molecules or polymers. However, some experimental observations suggest that this is not the case. For example, polymerization of **4** in aqueous solution requires markedly higher temperatures compared to a typical HCN polymerization experiment to result in the same polymer yield (126). Moreover, solid crystals of **4** can be observed as a side-product during HCN reactions (e.g., refs. (59, 124)). In this chapter, I study and compare initial stages of **4** and polyimine (**7**) formation - both proposed products of base-catalyzed HCN reactions. I investigate how different the kinetics and thermodynamics of the first steps in the formation and **4** and **7** are. Moreover, **4** and **7** are both proposed to form with iminoacetonitrile (**2**) as a first intermediate. **2** has been suggested to be the rate-limiting step in the formation of **4** (138) – is this true for formation of **7** as well? Could formation of either **4** or **7** occur in astrochemical environments? Here, I focus on reactions in neat HCN – a simple starting point for studies on HCN reactivity. Molecules and polymers which form in aqueous solution can incorporate oxygen through reactions with water, expanding the space of possible reactions (121). The content of this chapter is based on Papers II and III.

6.1 Iminoacetonitrile – a first intermediate in HCN oligomerization

To explore HCN reactivity, we begin by studying the first proposed step of HCN oligomerization: the formation of **2**. Indirect evidence that **2** is an intermediate in HCN oligomerization comes from experimental studies which have shown that N-alkyliminoacetonitriles form N-alkyldiaminomaleonitriles – suggesting that the unsubstituted **2** forms **4** (139). Moreover, theoretical studies have shown that **2** is the most stable HCN dimer (145). **2** has also been detected in two molecular clouds, in Sagittarus B2(N) and G+0.693, making it additionally interesting within an astrochemistry context (195, 196).

Figure 6.1 shows the suggested mechanism for base-catalyzed formation of **2** (68, 138). A previous theoretical study has evaluated the kinetics and thermodynamics of base-catalyzed formation of **2** in an aqueous solution (197). In that study, the method used to describe the solvent was an implicit solvent model, and the focus was on evaluating the catalytic effect of ion-pairs. The study found that a simultaneous proton transfer and cyanide addition resulted in a lower-lying transition state than in a stepwise mechanism. Here, I address the question of the explicit influence of an HCN solvent environment on the reaction kinetics.

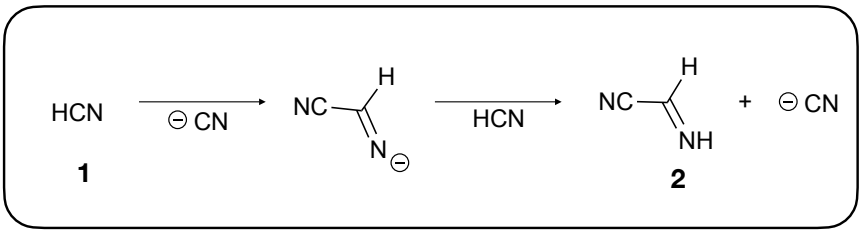

Figure 6.1. Suggested mechanism for base-catalyzed formation of iminoacetonitrile (**2**) (68, 138), which we evaluate in Paper II (198).

6.1.1 The free energy for formation of iminoacetonitrile in neat HCN

Figure 6.2A shows snapshots from a simulated formation of **2** in a liquid HCN environment. The snapshots are taken from a metadynamics simulation starting from a cyanide anion solvated in HCN (see Section 4.6.1). The simulation was steered towards making a carbon-carbon bond between the cyanide anion and an HCN solvent molecule. Simulations were set up so that we could identify transition states (via committor analysis, see Section 4.6.1) Two HCN molecules and the cyanide anion participate in the reaction. One HCN molecule coordinates and donates a proton to the other which is subjected to the attack by the cyanide anion.

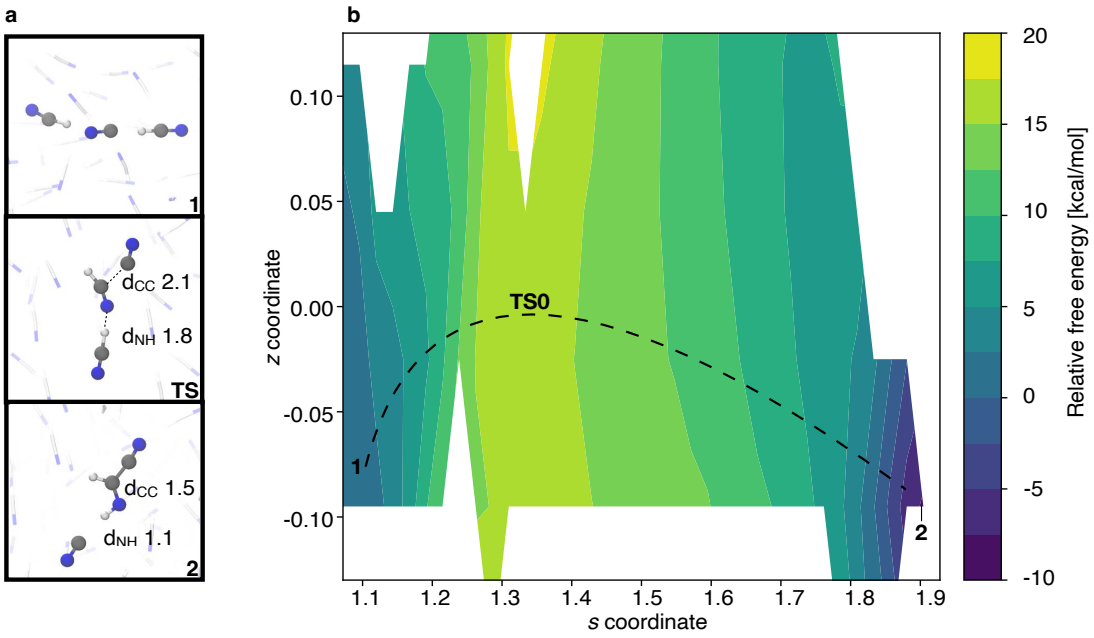

Figure 6.2. The free energy landscape of iminoacetonitrile (**2**) formation in path collective variable space. The white area corresponds to regions not sufficiently sampled during simulations. Here, the product is the *E*-form of **2**. We compute that the two isomers of **2** are almost equal in energy (within 1 kcal/mol). Figure is adapted from ref. (198).

Figure 6.2 shows the free energy profile of formation of **2** as a function of *s*- and *z*-coordinates (introduced in Section 4.6). At the level of theory used in the simulations, the barrier for formation of **2** is 15.5±1.2 kcal/mol and the reaction is exergonic by -7.1±0.8 kcal/mol. The reported uncertainty corresponds to a standard deviation estimated through block-averaging, and is a result of the limited sampling during the simulations. The computed barrier is low, likely due to method deficiency (see Section 4.6.3). We need to correct the free energy profile from simulations for a fair comparison between our computed free energy profile and experimental rates.

One way to improve upon the free energies obtained in simulations is to correct the electronic energy using a higher rung DFT method. In Paper II, we outline a procedure whereby hybrid DFT energy calculations are performed on structures taken from simulations of reactants, products, and the transition state (see Paper II for details). After correcting the free energy, the formation of **2** in HCN liquid is predicted to be marginally exergonic, ΔG^{o}_{278K}= -1.4±0.8 kcal/mol. Our corresponding best estimate for the activation energy is +21.8±1.2 kcal/mol.

Our predicted dimerization barrier is markedly lower than previously calculated for the same base-catalyzed formation of **2** in water (197). The previous study used the MP2 method, an implicit model for solvation, and considered only the electronic energy (not the free energy). The reason for the difference between the two results can in other words be several. An additional and potentially important

difference between the two models may be the presence of ion pairs. The concentration of ammonium in aqueous polymerization experiments is typically the same as the cyanide concentration (1 M, pH 9.2). However, in polar solution, the formation of ion-pairs in unlikely due to the efficient ion solvation by water.

Before we compare the implied reaction rate of our computed activation barrier with experiments, we need to verify whether formation of **2** is the rate-limiting step during base-catalyzed oligomerization of HCN in the neat liquid. Doing so requires investigating the corresponding activation barriers of subsequent reaction steps in the pathways towards **4** and **7**.

6.2 Two proposed products of base-catalyzed HCN reactions: diaminomaleonitrile and polyimine

4 and **7** make out two different proposed products of base-catalyzed HCN reactions. However, whereas **4** is a known thermodynamically favoured and experimentally detected product of HCN reactions (see Chapter 3 and Chapter 5), **7** is barely thermodynamically favoured at 298 K in aqueous solution (-1.4 kcal/(mol HCN)).[8] Still, **7** and fragments thereof have been proposed as a high-energy intermediates, hypothesized to form heterocycles in HCN reaction mixtures (154). Moreover, He et al. tentatively identified **7** as the major product of base-catalyzed polymerization of liquid HCN (123). **7** has been theoretically predicted to possess catalytic properties and be able to absorb wavelengths of light that are available on the surface of Titan (109). Such properties might enable the polymer to partake in dynamic chemistry on the otherwise cold and relatively energy-poor moon. In Paper III we explore the formation of HCN trimers and tetramers, proposed to lead either **4** or **7**.

Figure 6.3 shows proposed competing pathways starting from **2** and leading to **4** and **7** (Fig. 6.3). Pathway 1, leading to **4**, starts with an attack by a cyanide anion on the dimer's sp^2 carbon, forming aminomalononitrile (**3**) as a product (Fig 6.3, pathway 1) (138). In a subsequent step, **3** is proposed to form 2-amino-3-imino butanedinitrile (**32**), a tautomer of **4**, through a cyanide addition to one of its nitrile groups (122). In a last step, **32** tautomerizes to form **4**. In contrast, the proposed formation of **7** instead begins with a cyanide addition to the nitrile carbon of **2**, forming the trimer 2,3-diiminopropanenitrile (**33**) (Fig 6.3, pathway 2) (123). In the next step, **33** can form the tetramer 2,3,4-triiminobutanenitrile (**34**). Continued additions of cyanide anions to the nitrile group of the growing oligomer chain leads to formation of **7**. As a third alternative, He et al. proposed that **33** also can form **4** via **32** as an intermediate (Fig 6.3, pathway 3) (123).

[8] **7** is more favoured at lower temperatures, see Chapter 5.

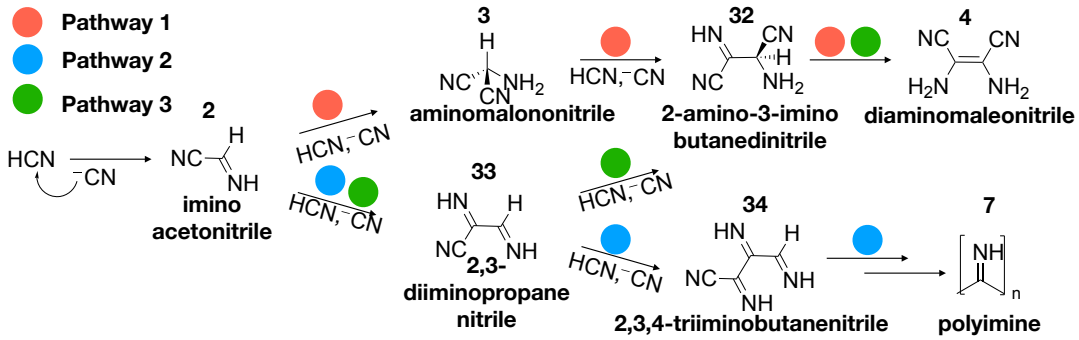

Figure 6.3. Three proposed reaction pathways for base-catalyzed HCN oligomerization and polymerization. Pathways 1 and 3 both lead to the commonly observed product diaminomaleonitrile (**4**). Polyimine (**7**) has been suggested to form in base-catalyzed HCN polymerization via successive cyanide additions onto the nitrile group of the growing polymer chain. Figure adapted from Paper III.

6.2.1 The free energy of base-catalyzed HCN oligomerization

Figure 6.4 shows the studied free energy profile of the first stage of formation of **4** and **7** in HCN liquid. Our derived free energy profile shows that pathway 1 (formation of **4**) is favoured under thermodynamic control. We did not study the last step of pathway 1 (**32**→**4**) but expect such imine-enamine tautomerization to proceed relatively fast in a protic polar solvent such as HCN. **4** computes as ~14 kcal/mol more stable than **32** at 278 K (refer to Paper III for details).

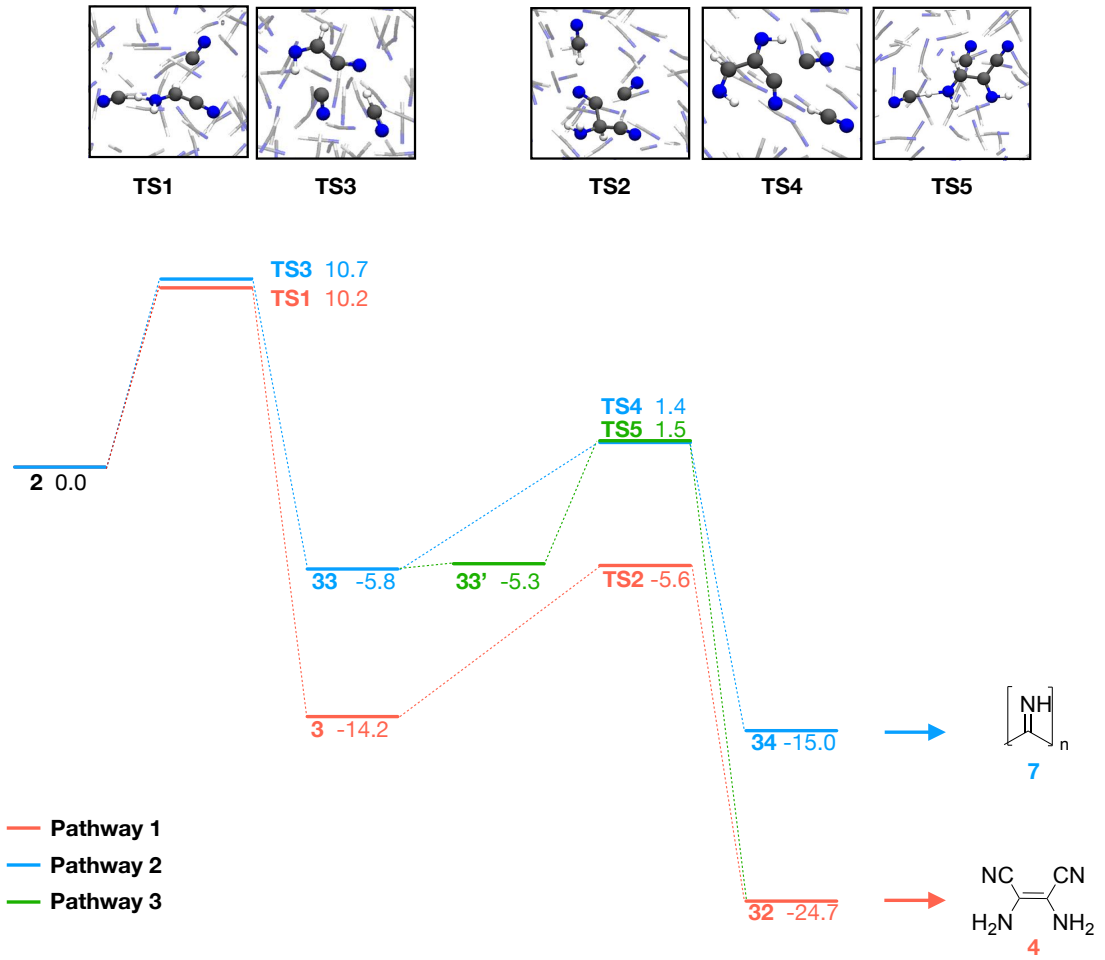

Figure 6.4. The computed energy profiles of the first steps of base-catalyzed formation of polyimine (**7**) and diaminomaleonitrile (**4**). The free energy was computed through umbrella sampling simulations at 278 K using PBE-D3 (see Paper III for details). Figure adapted from Paper III.

Our results in Figure 6.4 show that the first steps during formation of **4** and **7** are kinetically competitive. In the trimerization step, a cyanide addition onto the nitrile group occurs with very similar activation barrier as an addition to the sp^2 carbon (difference <0.5 kcal/mol). Moreover, the two lowest barriers of the tetramerization steps (**33**→**32** and **33**→**34**) compute as less than 0.1 kcal/mol apart. After correcting the trimerization and tetramerization barriers to account for the method deficiency in simulations, (see Section 4.6.3), all barriers in pathways 1-3 lie below ~16 kcal/mol. I also note that the barrier for continued polyimine growth remains low (similar to the barrier for trimerization and tetramerization) even for considerably longer polyimine structures (see Paper III for details). These results mean that polyimine oligomers likely form to a similar extent as **4** under kinetic control (i.e., low temperature). Polyimine oligomers could possibly have a role as high-energy intermediates in the formation of heterocycles, as proposed by Ruiz-Bermejo and colleagues (154). Whether or not there exist kinetically accessible routes from small polyimine fragments to heterocycles has not been explored yet.

The reliability of our computed barriers (setting the method deficiency aside) depends on the statistical error due to sampling (see Section 4.6.2). In this case, the statistical error due to insufficient sampling is between 1.1-2.9 kcal/mol (see details in Paper III). For context, an energy difference of 3 kcal/mol results in a relative Boltzmann population on the order of 10^{-4}-10^{-3} between 200 and 278 K. Hence, the uncertainty in the predicted ratio of **4** and **7** (based solely on the kinetics) can vary with several orders of magnitude.

6.2.2 Predicting the timescale for polyimine and diaminomaleonitrile formation in different environments

Taken together, our results show that formation of **2** is the rate limiting step, both in the formation of **4** and of **7** during base-catalyzed reactions of liquid HCN. The trimerization and tetramerization barriers compute as ~5-8 kcal/mol lower than the barrier for the rate-limiting step (Fig. 6.4). The rate-limiting step of a mechanism is most important in governing the apparent reaction rates. Thus, we can compare the predicted reaction rate of dimerization with the observed rate of diaminomaleonitrile or polymer appearance.

In Figure 6.5 we show the dimerization reaction half-life computed as a function of temperature (estimated with the Eyring equation and assuming a first-order reaction). At 278 K, the half-life of the reaction is about 5 hours. For comparison, in the studies of Mamajanov and Herzfeld the appearance of polymer (in pure HCN at 278 K) happened on the order of days (124). He et al. observed a similar experimental polymerization rate in pure HCN at room temperature (123). The difference between a half-life of hours and days at ambient temperature is large but corresponds to small difference in activation energy (~1 kcal/mol).

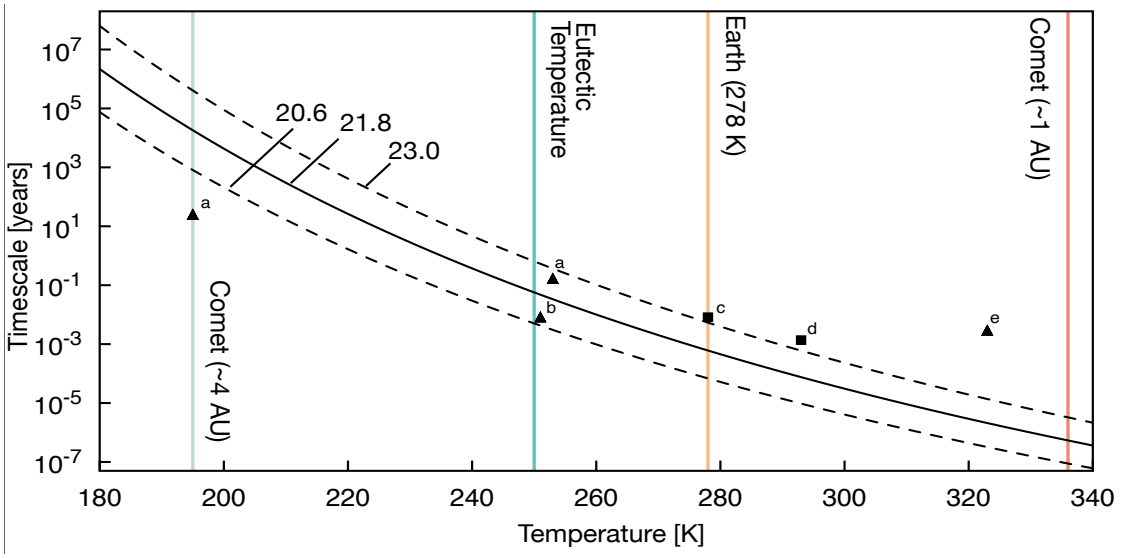

Figure 6.5. Our estimate of the half-life for iminoacetonitrile (**2**) formation as a function of temperature in pure HCN using a first-order Eyring equation. The three lines correspond to our barrier estimate 21.8±1.2 kcal/mol. The barrier height is temperature dependent. We have used the barrier calculated at 278 K to extrapolate reaction times to lower and higher temperatures. The half-life at the temperature of the simulations (278 K) is on the order of hours. At the eutectic temperate (250 K), the half-life is on the scale of days. Experimental data is shown as squares reactions in neat HCN liquid and triangles – reactions in aqueous solution. Data from [a]ref. (199), [b]ref. (138), [c]ref. (124), [d]ref. (123)and [e]ref. (132). Figure adapted from Paper II (198).

If one assumes a similar barrier in aqueous solution as in liquid HCN, we can expand our comparison to experimental rates observed in aqueous solution. Such a comparison is approximate. Yet, as water and HCN are both polar protic solvents, we might expect similar trends in the solvation effects from the two liquids. In Figure 6.5 we can see that the computed half-life agrees well with observed reaction rates close to the eutectic point of HCN and water mixtures. The agreement is worse at high (320 K) and low (190 K) temperatures. Our computed dimerization barrier is ~6 kcal/mol larger than the apparent polymerization barrier measured in aqueous solution by Mas et al. (measured between 323-363 K) (132). However, our barrier is in relative agreement with the computed activation barrier by Fernandez et al. measured in aqueous solution (11-20 kcal/mol) at 348-363 K (200), as well as the tetramerization rate observed by Sanchez et al. in aqueous solution at 273-313 K (122). In summary, our results support the premise in which formation of **2** is the rate-limiting step in base-catalyzed HCN polymerization, as proposed by Sanchez et al. (138).

Figure 6.5 reveals the reaction rate of the studied HCN dimerization in other temperatures relevant to astrochemical environments. For example, at temperatures around to 200 K (close to the warmest temperature in Titan's atmosphere) the reaction would proceed with a time scale of thousands of years.

Albeit exceedingly too slow to observe in experiments, the time frame is still interesting in an astrochemistry context. On the other hand, at temperatures close to those on the surface of Titan (94 K) the predicted reaction rate exceeds the age of the Universe, and the reaction is clearly not feasible. Future work could explore other compounds' potential to catalyze formation of **4** and **7** at low temperatures, as indicated in experiments (131). The subsequent steps after dimerization, formation of HCN trimers and tetramers, could occur relatively fast (timescale thousands of years) with activation barriers of ~ 16 kcal/mol at temperatures as low as ~150 K.

6.3 On the role of simulations when studying HCN reactions

We have used molecular dynamics simulations to better describe the solvent environment of HCN. However, such methods are time-consuming. The cost of studying a reaction profile using molecular dynamics is approximately three orders of magnitude larger compared to a standard molecular (static) calculation combined with an implicit solvent model (~1 CPU hour per reaction compared to thousands). After evaluating free energy profiles obtained using both approaches it is therefore crucial to ask: When is practically motivated to employ molecular dynamics simulations? Solvent effects are expected to be more accurately described using simulations (see section 4.4), while implicit models are markedly cheaper to use.

Figure 6.6 shows a comparison of reaction free energies (panel A) and reaction barriers (panel B) computed through simulations and molecular modelling. There is a good linear relationship between the two methods. The simulations compute the free energy barriers as slightly less than does the PCM model (MAD 1.2 kcal/mol). On the one hand, the agreement in Figure 6.6 is surprising, as we modelled HCN using a water PCM model with an adjusted dielectric constant. Moreover, hydrogen bonding and proton transfers are important parts of the studied reactions. Interactions with the HCN liquid is something implicit models cannot capture (see section 4.4). However, in this case, PCM predicts well the free energy profile of base-catalyzed HCN reactions, at least at the PBE-D3 level of theory.

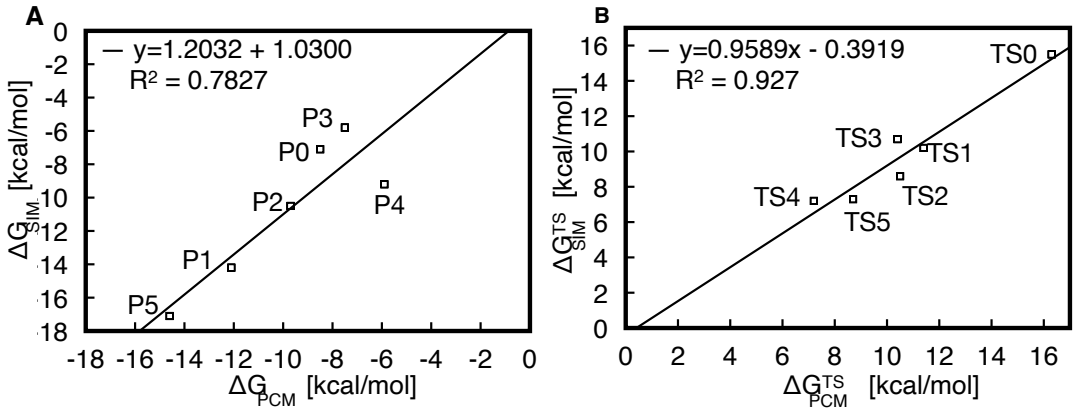

Figure 6.6. A comparison of the reaction free energy (Panel A) and barrier (Panel B) from simulations and molecular modelling in PCM. The trimerization and tetramerization transition states are named as in section 6.2.1. The dimer transition state is named TS0. In Panel A, the products corresponding to those formed from the transition state have been given the same numbers. The simulations were performed at the PBE-D3/GTH-DZVP level while the molecular modelling was done with PBE-D3/6-31G(d,p). The water PCM model was used with an adjusted dielectric constant to that of HCN at 278 K (144.8 (51)).

We cannot say how the comparison in Figure 6.6 would appear when using another, higher level of theory. On the other hand, the associated computational costs hinder us from making the same type of comparison. I note, however, that the corrected dimerization barrier (hybrid DFT level) computes at 2.5 kcal/mol less than the corresponding value in PCM. In future work, hybrid functionals could be used within a QM/MM description, in which different levels of theory are used for the surrounding and reacting complex. However, QM/MM is less practical for reaction exploration, when it is unknown which molecules participate in the reactions.

I note that one challenge in molecular modelling of reactions in solutions is the choice of reference state (the model of the reactant). The reference can either be a reactant complex or non-interacting solvated reactant molecules. Both reference models come with drawbacks. In this case, Figure 6.7 validates our choice of reference structure (the reactant complex) in the molecular modelling.

6.4 Conclusions

Simulations have been used to demonstrate very similar kinetics for the formation of **4** and **7**. For both products, formation of **2** is the rate-limiting step. Out of the two products, **4** is strongly thermodynamically favoured. We find that there is little kinetic preference for pathways leading to either product in the trimerization and tetramerization steps.

Our studies suggest that base-catalyzed HCN oligomerization via cyanide additions can proceed under relatively low temperatures (>200 K, assuming a temperature independent barrier) such as those on some cometary surfaces (< 4 AU) (112), or in concentrated solutions which might have existed on early Earth (like eutectic phases). However, barring heating through high-energy impact events, the conditions on Titan's surface are likely too cold to allow for formation of **2** through the investigated pathway.

One limitation of comparing our results with experiments is that our model describes a pure HCN system. In experiments, trace amounts of impurities may exist which could facilitate initiation of polymerization. Our model does not either describe the role of known oligometization catalysts, such as cyanohydrins, glyconitrile or ammonia (121, 131). Future studies might explore other small, readily available nitriles, as initiators for base-catalyzed HCN reactions. The subsequent steps following dimerization, forming HCN trimers and tetramers, have lower estimated activation barriers than formation of **2**, which suggests that once initiated, HCN oligomerization could proceed at lower temperatures than the dimerization can.

This chapter compared two computational approaches for exploring reactions in HCN solvent. On one hand, the simulations approach allows for a detailed description of the HCN solvent. On the other hand, the simulations are computationally costly and limit the level of theory. We can conclude that the free energy profiles for base-catalyzed HCN reactions obtained using an implicit solvent model and an explicit solvent model are very similar and show the same trends. Future studies can reduce computational cost associated with constructing a free energy profile by using smaller molecular models. Simulations are still useful in reaction exploration through metadynamics at fast and lower levels of theory. Automated reaction searches could allow for an efficient exploration of HCN reactions and products (201).

7
Evaluating the Potential for Cell Membranes in the Lakes of Titan

Is there life in other places of the Universe, and if so, where? Could cell membranes form in a world without liquid water? The imagination may not know any limits in answering this question, but chemistry might. In this chapter, I use computational chemistry to explore the plausibility of cryogenically operable membranes in the hydrocarbon lakes of Titan. The content of this chapter is based on Paper IV.

7.1 The role of the membrane

Life as we know it relies on membranes to concentrate, compartmentalize, and protect the molecular machinery inside the cell (202). It may be reasonable to assume that life in other places of the Universe would do the same, possibly under completely different environmental conditions, such as those of Titan. On Earth, cell membranes exist in a water surrounding, and most often close to ambient conditions. On Titan, on the other hand, the lakes, made of hydrocarbons, are very cold (<94 K) (203). Any membrane stable on Titan would need to adapt to these widely different conditions. Here, we investigate whether one certain type of membrane could be operable on Titan. Before I go on to describe the specific hypothesis to be explored, it helps to remind of how normal cell membranes form and operate.

A cell membrane has two domains: a hydrophilic external part which is exposed to water and an internal hydrophobic part (left panel, Fig 7.1). The membrane is composed of molecules which arrange to form a membrane through the process of self-assembly. The molecules are often amphiphilic with one polar and one non-polar end. Such amphiphilic molecules arrange in a bilayer structure (left panel, Fig 7.1). The driving force for the self-assembly is in part an entropic one and is called the hydrophobic effect. As a result of membrane formation, the number of possible interactions in the network of water molecules surrounding the membrane, is maximized (204). At the same time, the water molecules can create favourable enthalpic interactions through hydrogen bonds amongst themselves. The hydrophobic effect is not operable on Titan as its lakes and seas are made of non-polar hydrocarbon liquids which do not form hydrogen bond networks.

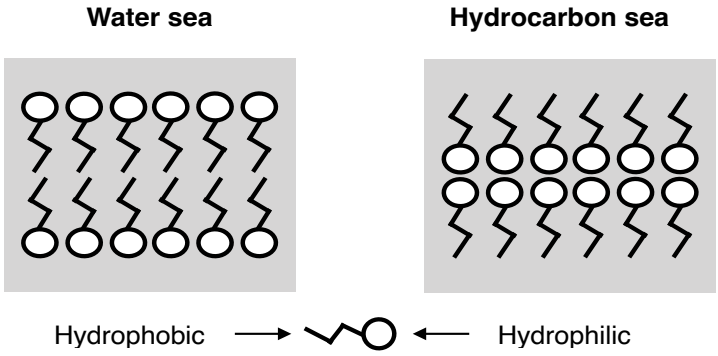

Figure 7.1 Two different types of membranes. Left panel: A bilayer membrane in an aqueous solution has an external hydrophilic part and an internal hydrophobic part. Right panel: An inverted membrane proposed to form in hydrocarbon solvents, like those in the lakes of Titan. The hydrophilic polar ends are oriented towards the internal part of the membrane, while the non-polar end is in contact with the hydrocarbon solvent. Figure reproduced from ref. (58).

In a non-polar liquid, like in the hydrocarbon seas of Titan, the self-assembly of amphiphilic molecules is hypothesized to result in an inverted membrane (Figure 7.1, Right panel) (205). The driving force for this type of self-assembly would not be entropic, but rather enthalpic. The polar hydrophilic end groups have the potential for the strongest type of interaction. However, to maintain membrane fluidity at low temperatures, interactions between molecules in the membrane cannot be too strong. Only small polar molecules would have weak-enough dispersion interactions in low-temperature environments.

With these prerequisites in mind, the azotosomes were proposed by Stevenson et al. as a type of membrane which could form on Titan (205). Azotosome is a name given to inverted membranes which are made up of polar nitrogen containing compounds. Stevenson et al. computationally tested the kinetic stability of several azotosomes made up of such small polar molecules. Membranes made from acrylonitrile showed the best promise, both in terms of kinetic stability and flexibility (205). Moreover, acrylonitrile has been detected in the atmosphere of Titan, increasing the relevance of the azotosome proposal (206). With the detection of acrylonitrile and the predicted kinetic stability of the azotosome, the idea of membranes on Titan appeared as a real possibility. But can azotosomes form in the first place?

7.2 Can membranes be operable on Titan?

The spontaneous self-assembly of acrylonitrile molecules into an azotosome depends on the thermodynamic landscape of acrylonitrile. To form, the azotosome needs to be the most stable form of acrylonitrile on Titan's cold surface. The competitors are solid acrylonitrile (f.p. 170 K) or acrylonitrile solvated in the Titan's hydrocarbon seas. To determine which structure is thermodynamically favoured, we computed the free energy of the azotosome and solid acrylonitrile. Doing so requires consideration of thermal effects as well methane solvation effects.

Figure 7.2 shows the computed relative free energies of the azotosome and the most stable form of solid acrylonitrile (see Paper IV for details). The relative energy of an acrylonitrile molecule in the azotosome compared to the crystal structure is 8-11 kJ/mol. The range corresponds to different ways of computing van der waals interactions (PBE(165)-D3(189) and vdw-df-cx (207), see Paper IV). The difference in energy per molecule between the structures is small. Yet, an azotosome is a supramolecular structure made from many acrylonitrile molecules. Multiple building blocks are required to form one mol of membrane. Hence, the larger the azotosome is, the larger is the energy difference between the azotosome and solid acrylonitrile. The probability of forming an azotosome structure can be computed via the Boltzmann distribution. As the energy difference between the azotosome and solid grows, the probability becomes diminishingly small. Hence, it follows that azotosome formation, proposed to occur in the seas of Titan, is unlikely.

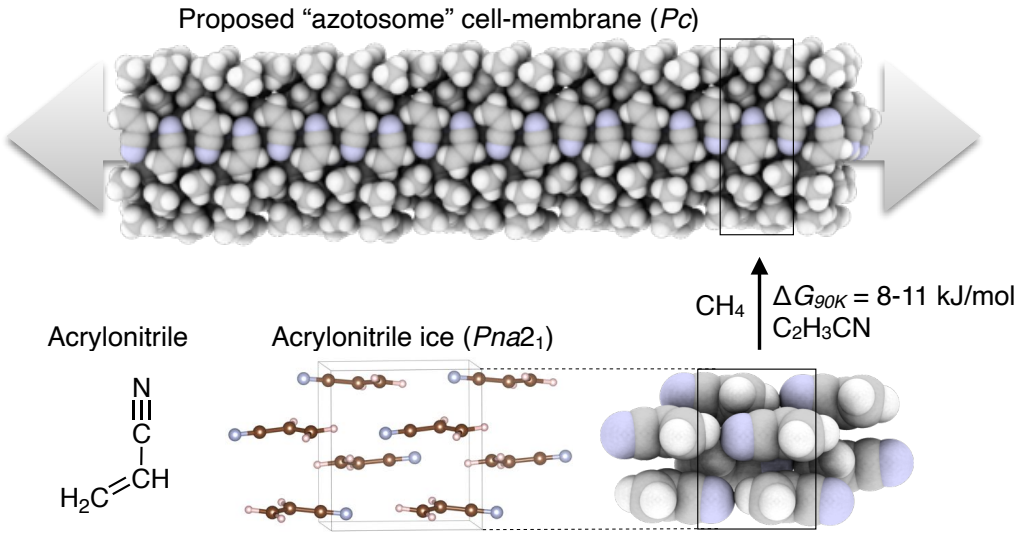

Figure 7.2. A thermodynamic evaluation of membrane stability. DFT calculations estimate that the azotosome is not thermodynamically favored to form in the seas of Saturn's moon Titan. The acrylonitrile molecules will instead form the solid form of the molecule. Crystal symmetries of the considered phases are shown within parenthesis. Figure is reproduced from Sandström and Rahm (Paper IV) (208).

7.3 Comparing acrylonitrile solubility and critical azotosome concentration

As outlined above, the azotosome will not form spontaneously from the solid in Titan's seas. Yet, one might wonder whether azotosomes could still form directly from acrylonitrile molecules solvated in methane. The answer to this question can be found by studying the solubility of acrylonitrile. The solubility corresponds to the maximum concentration of acrylonitrile in the methane solvent. The solubility, K_p, is related to the relative energy between the solid and the solvated molecule, $\Delta G_{solv-solid}$, according to

$$\Delta G_{solv-solid} = -RTlnK_p \qquad (7.1)$$

where R is the gas constant and T is the temperature. The solubility of acrylonitrile in methane has been estimated to 10^{-10} M (based on the methane molar density of 28 M (209) at 94 K, 1.5 bar and a solubility of 10^{-11} mole fractions (210)). This solubility corresponds to a $\Delta G_{solv-solid}$ of ~18 kJ/mol in favour of the solid. Furthermore, $\Delta G_{solv-solid}$ can be used to estimate the relative energy of the solvated azotosome compared to the solvated molecule, $\Delta G_{azotosome-solv}$, to -10 – -7 kJ/mol. In other words, acrylonitrile is more stable in the solvated azotosome than when solvated in methane. We can use $\Delta G_{azotosome-solv}$ to derive a measure analogous to critical micelle concentration (the acrylonitrile concentration required for azotosome formation), which we call $C_{critical}$ according to

$$\Delta G_{azotosome-solv} = RTlnC_{critical}. \quad (7.2)$$

Using Eq. (7.2), the critical azotosome concentration computes to 10^{-6}-10^{-4} M. Through a comparison of Equations (7.1) and (7.2), we can conclude that the critical azotosome concentration cannot be reached if the solvated azotosome is less stable than solid acrylonitrile. Hence, an azotosome more stable than the acrylonitrile solid is a requirement for azotosome self-assembly in the seas of Titan. The critical azotosome concentration and solubility are only approximate in the absence of experimental measurements. With the available solubility estimates and our computed relative energy of the azotosome, we derive a minimum four order of magnitudes difference between the solubility of acrylonitrile and the critical azotosome concentration at 94 K.

7.4 Conclusions

Can there be alternative forms of cell membranes on Titan? According to our estimates the answer is no. Whereas azotosomes have been shown to be kinetically stable, the unfavorable thermodynamics argues against this hypothesis. I note, however, that we focused our study to acrylonitrile azotosomes. We have not proven that other types of azotosomes, made from other molecules or at other temperatures, are thermodynamically unviable.

The proposal of the azotosome has inspired work on inverted membranes and micelles in hydrocarbon solvents. For example, inverted micelles, whose self-assembly would be driven by the same enthalpic principle as that proposed for the azotosome, were experimentally observed in a cyclohexane solvent at ambient conditions (211). Yet, theoretical simulations predict that the structure of such aggregates are not micellar (212).

The azotosome is an example of an astrobiology hypothesis which challenges our ideas of the limits of life. While in this case the results are not in favor of the hypothesis, this study is an example of the possible role of computational chemistry in astrobiology. Our work highlights the need to explore both the kinetic and thermodynamic stability to arrive at a conclusion regarding overall chemical stability. With the upcoming Dragonfly mission to Titan (14), studies like Paper IV continue to be important for furthering our understanding of the moon's hydrocarbon world.

8
Concluding Remarks and Outlook

This thesis has been devoted to an examination of central questions in prebiotic chemistry and astrobiology. I hope the research presented herein will be used to guide future studies in computational astrobiology, especially in the study of HCN reactivity. This chapter briefly presents key takeaways and discusses possible future research that could build upon the present work.

The calculated thermodynamic landscape of HCN-derived polymers reveals thermodynamic feasibility, or unfeasibility, of different proposed polymer products of HCN reactions. From a thermodynamics point of view, Paper I resolved some of the long-standing questions regarding HCN reactivity. A useful extension to the thermodynamic landscape could be adding considerations of polymers which form through condensation reactions, producing ammonia and resulting in polymers with unequal N, C, H ratios. Such reactions have been suggested as important during polymerization of diaminomaleonitrile (**4**) (149, 150, 154). The stability of oxygen-containing reaction products, which could form in aqueous solution, could also be considered in an extension to the thermodynamic map.

Improvements in our understanding of HCN reactions could be made through further studies focusing on mechanistic and kinetic aspects of polymer formation. Such studies might answer questions that arise when studying the thermodynamic landscape, such as: Do different polymerization mechanisms (radical or anionic) favour different products in the map? How selective is the reactivity of diaminomaleonitrile towards purines and polymers? Can HCN-derived polymers form polypeptides, as proposed by Matthews and Moser (146) but contended by others (194)? Why does adenine (**6**) form in such small quantities in HCN reaction experiments (69), when the purine is estimated as the most stable out of the studied structures? Does there exist chemical conditions or catalysts which would favour adenine formation?

This thesis included our first steps towards also unravelling the kinetics and selectivity of HCN reactions. Taking our study as a first indication, there exist multiple kinetically competitive processes during the beginning of HCN polymerization. Our mechanistic studies of base-catalyzed HCN oligomerization reveal that formation of iminoacetonitrile (**2**) is the rate-determining step when the oligomerization occurs in HCN liquid.

One of the central goals of the research in this thesis was to predict which HCN reactions can proceed in various astrochemical conditions. The predicted dimerization rate and observed experimental rate are in a fair agreement. Later steps of the oligomerization, the formation of HCN trimers and tetramers are estimated to proceed under astrochemically relevant timescales even at temperatures as low as ~150 K. These predictions suggest that base-catalyzed reactions could take place on cold worlds, albeit very slowly (timescale of thousands of years). The predicted reaction rates are exponentially dependent on errors in the methods used to compute activation barriers. Our results here show that we cannot exclude the possibility of low temperature HCN oligomerization.

Predictions of the chemical stability depend on the reliability of quantum chemical computation of relative free energies. Highly accurate methods, such as coupled cluster methods, are too computationally demanding to use for oligomer models and solvated systems studied herein. More computationally feasible are DFT methods such as B3LYP-D3 or PBE-D3, used in this thesis, which both have associated errors. The need for fast DFT methods in my studies was largely due to the large size of the models used to describe polymers and solution-state chemistry.

Investigating HCN reactions on surfaces of solids would be a natural extension to the work presented herein. In this work we only explore reactions in liquid HCN and extrapolate our results to predict reactivity at low temperatures. However, the solid environment might influence reactions occurring on its surface differently than the liquid. Solid HCN has been detected in clouds on Titan (97), and is predicted to exist on comets and in interstellar grains because of the high gas-phase abundance of HCN. Studying properties of and reactions on HCN crystal surfaces is a potential avenue to gain further insight into the astrochemistry of HCN.

In this thesis I examined the effect of using molecular dynamics for an accurate description of the solvent effects of liquid HCN. The mechanistic studies of base-catalyzed HCN oligomerization (Paper II and III) showed that there is a good agreement between results obtained with our adjusted implicit model of the HCN liquid and those obtained through simulations that include explicit solvent molecules. Based on the experience I have gathered while working on this topic, I suggest that molecular dynamics is used most effectively for reaction or configurational exploration in which case one could test semi-empirical methods (213) or reactive force fields (214) for this purpose – the latter especially with the improvements made possible in recent years by machine learning potentials (215). Obtaining reaction barriers and energies can then effectively be carried out using smaller molecular models or mixed levels of theory (QM/MM) that also allow for the use of more accurate methods. Such approaches would be similar to the environment-perturbed transition-state sampling developed to study enzyme catalyzis (216).

This thesis demonstrates the importance of evaluating both thermodynamics and kinetics of chemical processes when discussing and developing new hypotheses in astrobiology. Understanding HCN chemistry may prove essential for understanding prebiotic chemistry processes that could have occurred on early Earth, as well as the formation of complex organic molecules in the solar system. Research regarding the origin and ubiquity of life is the focus of ongoing and upcoming space exploration. This includes the planned Dragonfly mission on Titan (14), exploration of Venus (217, 218) and of Mars (e.g., refs. (219–221)) just in the current and coming decennia.

9
Acknowledgements

The research presented in this thesis has been carried out at the Division of Chemistry and Biochemistry at Chalmers University of Technology, Sweden, with financial support from Chalmers University of Technology and the Swedish Research Council (grant 2016-04127). This research relied on computational resources provided by the Swedish National Infrastructure for Computing (SNIC) at Chalmers Centre for Computational Science and Engineering (C3SE), Nationellt Superdatorcentrum (NSC) and Parallelldatorcentrum (PDC) partially funded by the Swedish Research Council through grant agreement no 2018-05973.

I would like to thank my supervisor Martin Rahm for inviting me to work on this project, for his supervision and our collaboration during these years. I am also thankful to my examiner Per Lincoln for providing help when needed. I would like to thank all my collaborators, in particular Rana, Siddhant and Fernando who co-authored Paper I presented in this thesis.

To the many researchers and students who have taken part in our research group during my studies, you have been inspirational scientists, incredibly helpful and supportive colleagues, and great fun. I am also thankful to everyone from the Division of Chemistry and Biochemistry for creating a wonderful work atmosphere.

10
Bibliography

1. Lazcano, A. Which way to life. *Orig Life Evol Biosph* **40**, 161-167 (2010).
2. Krishnamurthy, R. Giving rise to life: Transition from prebiotic chemistry to protobiology. *Acc. Chem. Res.* **50**, 455-459 (2017).
3. Wakelam, V. et al. A kinetic database for astrochemistry (KIDA). *Astrophys. J., Suppl. Ser.* **199**, 21 (2012).
4. Woon, D. E. The astrochymist. *http://www.astrochymist.org/* (2008).
5. Lindal, G. F. et al. The atmosphere of Titan: An analysis of the voyager 1 radio occultation measurements. *Icarus* **53**, 348-363 (1983).
6. Krasnopolsky, V. A. Chemical composition of Titan's atmosphere and ionosphere: Observations and the photochemical model. *Icarus* **236**, 83-91 (2014).
7. Desai, R. T. et al. Carbon chain anions and the growth of complex organic molecules in Titan's ionosphere. *The Astrophysical Journal* **844**, L18 (2017).
8. Lavvas, P. et al. Aerosol growth in Titan's ionosphere. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2729-2734, S2729/1 (2013).
9. Hörst, S. M. Titan's atmosphere and climate. *J. Geophys. Res. Planets* **122**, 432-482 (2017).
10. Thelen, A. E. et al. Spatial variations in Titan's atmospheric temperature: ALMA and Cassini comparisons from 2012 to 2015. *Icarus* **307**, 380-390 (2018).
11. Jennings, D. E. et al. Titan surface temperatures during the Cassini mission. *ApJ* **877**, L8 (2019).
12. Stofan, E. R. et al. The lakes of Titan. *Nature* **445**, 61-64 (2007).
13. Tokano, T. et al. Methane drizzle on Titan. *Nature* **442**, 432-435 (2006).
14. Barnes, J. W. et al. Science goals and objectives for the Dragonfly Titan rotorcraft relocatable lander. *Planet. Sci.* **2**, 130 (2021).
15. Mckay, C. P. & Smith, H. D. Possibilities for methanogenic life in liquid methane on the surface of Titan. *Icarus* **178**, 274-276 (2005).
16. Lunine, J. I. Titan and habitable planets around m-dwarfs. *Faraday Discuss* **147**, 405-18; discussion 527 (2010).

17. Solomonidou, A. et al. The chemical composition of impact craters on Titan. *A&A* **641**, A16 (2020).

18. Lunine, J. I., Cable, M. L., Glein, C. R., Hörst, S. M. & Rahm, M. in *Planetary Astrobiology* (eds Meadows, V., Arney, G. N., Schmidt, B., E. & Des Marais, D. J.) 247-266 (University of Arizona, Tuscon, 2020).

19. NASA astrobiology. About life detection. *https://astrobiology.nasa.gov/research/life-detection/about/*

20. Oparin, A. I. *The origin of life on the Earth.* (Oliver & Boyd, Edinburgh & London, 1957).

21. Drobot, B. et al. Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nature Communications* **9**, 3643 (2018).

22. Neveu, M., Kim, H.-J. & Benner, S. A. The "strong" RNA world hypothesis: Fifty years old. *Astrobiology* **13**, 391-403 (2013).

23. Altermann, W. & Kazmierczak, J. Archean microfossils: A reappraisal of early life on Earth. *Research in Microbiology* **154**, 611-617 (2003).

24. Tice, M. M. & Lowe, D. R. Photosynthetic microbial mats in the 3,416-myr-old ocean. *Nature* **431**, 549-552 (2004).

25. Pearce, B. K. D., Tupper, A. S., Pudritz, R. E. & Higgs, P. G. Constraining the time interval for the origin of life on Earth. *Astrobiology* **18**, 343-364 (2018).

26. Bowring, S. A. & Williams, I. S. Priscoan (4.00–4.03 ga) orthogneisses from northwestern Canada. *Contrib. Mineral. Petrol* **134**, 3-16 (1999).

27. Van Zuilen, M. A., Lepland, A. & Arrhenius, G. Reassessing the evidence for the earliest traces of life. *Nature* **418**, 627-630 (2002).

28. Bottke, W. F. & Norman, M. D. The late heavy bombardment. *Annu. Rev. Earth Planet. Sci.* **45**, 619-647 (2017).

29. Sagan, C. & Mullen, G. Earth and Mars: Evolution of atmospheres and surface temperatures. *Science* **177**, 52-56 (1972).

30. Kasting, J. F. Faint young sun redux. *Nature* **464**, 687-689 (2010).

31. Zahnle, K., Schaefer, L. & Fegley, B. Earth's earliest atmospheres. *Cold Spring Harb. Perspect. Biol.* **2**, a004895 (2010).

32. Von Paris, P. et al. Warming the early Earth—CO2 reconsidered. *Planet. Space Sci.* **56**, 1244-1259 (2008).

33. Cnossen, I. et al. Habitat of early life: Solar x-ray and uv radiation at Earth's surface 4–3.5 billion years ago. *J. Geophys. Res.* **112**, E02008 (2007).

34. Schaefer, L. & Fegley, B. Outgassing of ordinary chondritic material and some of its implications for the chemistry of asteroids, planets, and satellites. *Icarus* **186**, 462-483 (2007).

35. Miller, S. L. & Urey, H. C. Organic compound synthesis on the primitive Earth. *Science* **130**, 245-251 (1959).

36. Tian, F., Kasting, J. F. & Zahnle, K. Revisiting HCN formation in Earth's early atmosphere. *Earth Planet. Sci. Lett.* **308**, 417-423 (2011).

37. Chyba, C. & Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: An inventory for the origins of life. *Nature (London)* **355**, 125-132 (1992).

38. Parkos, D., Pikus, A., Alexeenko, A. & Melosh, H. J. HCN production via impact ejecta reentry during the late heavy bombardment. *J. Geophys. Res. Planets* **123**, 892-909 (2018).

39. Bada, J. L. New insights into prebiotic chemistry from stanley miller's spark discharge experiments. *Chem. Soc. Rev.* **42**, 2186-2196 (2013).

40. Miller, S. L. A production of amino acids under possible primitive Earth. *Science* **117 3046**, 528-529 (1953).

41. Miller, S. Production of some organic compounds under the possible primitive Earth condition. *J. Am. Chem. Soc.* **77**, 2351-2361 (1955).

42. Miyakawa, S., James Cleaves, H. & Miller, S. L. The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig Life Evol Biosph* **32**, 195-208 (2002).

43. Miller, S. L. & Orgel, L. E. *The origins of life on Earth* (Prentice-Hall Inc., New Jersey, 1974).

44. Baross, J. A. & Hoffman, S. E. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig Life Evol Biosph* **15**, 327-345 (1985).

45. Koga, T. & Naraoka, H. A new family of extraterrestrial amino acids in the murchison meteorite. *Sci. Rep.* **7**, 636 (2017).

46. Kvenvolden, K. et al. Evidence for extraterrestrial amino-acids and hydrocarbons in the murchison meteorite. *Nature* **228**, 923-926 (1970).

47. Cronin, J. R. & Chang, S. (eds Greenberg, J. M., Mendoza-Gómez, C. X. & Pirronello, V.) 209-258 (Springer Netherlands, Dordrecht, 1993).

48. Martins, Z. et al. Extraterrestrial nucleobases in the murchison meteorite. *Earth Planet. Sci. Lett.* **270**, 130-136 (2008).

49. Coates, J. E. & Davies, R. H. 246. Studies on hydrogen cyanide. Part xviii. Some physical properties of anhydrous hydrogen cyanide. *J. Chem. Soc.* 1194-1199 (1950).

50. Coates, J. E. & Harthorne, N. H. Studies on hydrogen cyanide. Part iii. The freezing points of hydrogen cyanide. *J Chem Soc* **0**, 657-665 (1931).

51. Coates, G. E. & Coates, J. E. Hydrogen cyanide. Part xiii. The dielectric constant of anhydrous hydrogen cyanide. *J Chem Soc* 77-81 (1944).

52. Rumble, J. R. *Crc handbook of chemistry and physics* (CRC Press/Taylor & Francis, Boca Raton, FL, 2020).

53. Malmberg, C. G. & Maryott, A. A. Dielectric constant of water from 0° to 100° c. *JJ. Res. Natl. Bur. Stand. (U. S.)* **56**, Research paper 2641 (1956).

54. Martiniano, H. F. M. C. & Costa Cabral, B. J. Structure and electronic properties of a strong dipolar liquid: Born-oppenheimer molecular dynamics of liquid hydrogen cyanide. *Chem. Phys. Lett.* **555**, 119-124 (2013).

55. De Oliveira, P. M. C., Silva, J. A. B. & Longo, R. L. Benchmark, DFT assessments, cooperativity, and energy decomposition analysis of the hydrogen bonds in HCN/HNC oligomeric complexes. *J. Mol. Model.* **23**, 1-10 (2017).

56. Sánchez, M., Provasi, P. F., Aucar, G. A., Alkorta, I. & Elguero, J. Theoretical study of HCN and HNC neutral and charged clusters. *J. Phys. Chem. B* **109**, 18189-18194 (2005).

57. Dulmage, W. J. L., W. N. The crystal structures of hydrogen cyanide HCN. *Acta cryst* **4**, 330-334 (1951).

58. Sandström, H. Quantum chemical exploration of nitriles in prebiotic chemistry and astrobiology [licentiate thesis]. Chalmers University of Technology. Göteborg. (2020).

59. Mozhaev, P. S., Kichigina, G. A., Kiryukhin, D. P. & Barkalov, I. M. Radiation-induced polymerization of hydrogen cyanide. *High Energy Chem. (Transl. of Khim. Vys. Energ.)* **29**, 15-18 (1995).

60. Giauque, W. F. & Ruehrwein, R. A. The entropy of hydrogen cyanide. Heat capacity, heat of vaporization and vapor pressure. Hydrogen bond polymerization of the gas in chains of indefinite length. *J. Am. Chem. Soc.* **61**, 2626-2633 (1939).

61. Esposti, C. D., Cazzoli, G., Damiani, D., Favero, P. G. & Strumia, F. Dipole moment of HCN by microwave-stark saturated absorption spectroscopy. *Infrared Phys.* **28**, 21-28 (1988).

62. Brandao, I., Rivelino, R., Fonseca, T. L. & Castro, M. A. An ab initio study of electric properties of linear (HCN)n and (HNC)n aggregates in gas phase. *Chem. Phys. Lett.* **580**, 9-13 (2013).

63. Rahm, M., Belanger-Chabot, G., Haiges, R. & Christe, K. O. Nitryl cyanide, ncno2. *Angew. Chem., Int. Ed.* **53**, 6893-6897 (2014).

64. Panda, M. & Robinson, N. C. Kinetics and mechanism for the binding of HCN to cytochrome c oxidase. *Biochemistry* **34**, 10009-10018 (1995).

65. Sutherland, J. D. The origin of life-out of the blue. *Angew. Chem., Int. Ed.* **55**, 104-121 (2016).

66. Yadav, M., Pulletikurti, S., Yerabolu, J. R. & Krishnamurthy, R. Cyanide as a primordial reductant enables a protometabolic reductive glyoxylate pathway. *Nat. Chem.* **14**, 170-178 (2022).

67. Oró, J. Synthesis of adenine from ammonium cyanide. *Biochem Biophys Res Commun* **2**, 407-412 (1960).

68. Oro, J. Mechanism of synthesis of adenine from HCN under possible primitive Earth conditions. *Nature* **191**, 1193-1194 (1961).

69. Ruiz-Bermejo, M., Zorzano, M.-P. & Osuna-Esteban, S. Simple organics and biomonomers identified in HCN polymers: An overview. *Life (Basel, Switz.)* **3**, 421-448 (2013).

70. Lv, K.-P., Norman, L. & Li, Y.-L. Oxygen-free biochemistry: The putative CHN foundation for exotic life in a hydrocarbon world. *Astrobiology* **17**, 1173-1181 (2017).

71. Marín-Yaseli, M. R., Mompeán, C. & Ruiz-Bermejo, M. A prebiotic synthesis of pterins. *Chem. Eur. J.* **21**, 13531-13534 (2015).

72. Miyakawa, S., Cleaves, H. J. & Miller, S. L. The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions. *Orig Life Evol Biosph* **32**, 209-218 (2002).

73. Molter, E. M. et al. ALMA observations of HCN and its isotopologues on Titan. *Astron. J.* **152**, 42/1-42/7 (2016).

74. Ferus, M. et al. High energy radical chemistry formation of HCN-rich atmospheres on early Earth. *Sci. Rep.* **7**, 6275 (2017).

75. Airapetian, V. S., Glocer, A., Gronoff, G., Hébrard, E. & Danchi, W. Prebiotic chemistry and atmospheric warming of early Earth by an active young sun. *Nat. Geosci.* **9**, 452-455 (2016).

76. Buhl, D. & Snyder, L. E. Unidentified interstellar microwave line. *Nature* **228**, 267-269 (1970).

77. Tanguy, L. et al. Stratospheric profile of HCN on Titan from millimeter observations. *Icarus* **85**, 43-57 (1990).

78. Rodgers, S. D. & Charnley, S. B. Hnc and HCN in comets. *Astrophys J Lett* **501**, L227-L230 (1998).

79. Wirström, E. et al. HCN observations of comets c/2013 r1 (Lovejoy) and c/2014 q2 (Lovejoy). *Astron. Astrophys.* **588**, A72 (2016).

80. Tokunaga, A. T., Beck, S. C., Geballe, T. R., Lacy, J. H. & Serabyn, E. The detection of HCN on Jupiter. *Icarus* **48**, 283-289 (1981).

81. Lellouch, E., Romani, P. N. & Rosenqvist, J. The vertical distribution and origin of HCN in neptune's atmosphere. *Icarus* **108**, 112-136 (1994).

82. Lellouch, E. et al. Detection of CO and HCN in Pluto's atmosphere with alma. *Icarus* **286**, 289-307 (2017).

83. Marten, A. et al. First observations of CO and HCN on Neptune and Uranus at millimeter wavelengths and their implications for atmospheric chemistry. *Astrophys J* **406**, 285-297 (1993).

84. Öberg, K. I. et al. The comet-like composition of a protoplanetary disk as revealed by complex cyanides. *Nature* **520**, 198-201 (2015).

85. European Southern Observatory(ESO) & NASA. The Galactic Centre and Sagittarius B2. *https://www.eso.org/public/images/eso0924e/* (2009).

86. NASA, Jet Propulsion Laboratory-Caltech & Space Science. Enhanced-color view of Titan taken by Cassini on Sept. 13, 2017. *https://www.nasa.gov/image-feature/jpl/pia21890/a-last-look-at-Titan* (2017).

87. European Space Agency (ESA). Image of 67P/Churyumov-Gerasimenko shows the diversity of surface structures on the comet's nucleus. *https://www.nasa.gov/jpl/rosetta/comet-surface-variations-20140815/* (2014).

88. Herbst, E. Chemistry in the interstellar medium. *Annu. Rev. Phys. Chem.* **46**, 27-54 (1995).

89. Linke, R. A., Goldsmith, P. F., Wannier, P. G., Wilson, R. W. & Penzias, A. A. Isotopic abundance variations in interstellar HCN. *Astrophys J* **214**, 50-59 (1977).

90. Loison, J.-C., Wakelam, V. & Hickson, K. M. The interstellar gas-phase chemistry of HCN and HNC. *Mon Not R Astron Soc* **443**, 398-410 (2014).

91. Herbst, E. & Klemperer, W. The formation and depletion of molecules in dense interstellar clouds. *Astrophys. J.* **185**, 505-534 (1973).

92. Herbst, E. & Van Dishoeck, E. F. Complex organic interstellar molecules. *Annu. Rev. Astron. Astrophys.* **47**, 427-480 (2009).

93. Pau, C. F. & Hehre, W. J. Heat of formation of hydrogen isocyanide by ion cyclotron double resonance spectroscopy. *J. Phys. Chem.* **86**, 321-322 (1982).

94. Deprince Iii, A. E. & Mazziotti, D. A. Molecular geometries and harmonic frequencies from the parametric two-electron reduced density matrix method with application to the HCN ↔ HNC isomerization. *J. Phys. Chem. B* **112**, 16158-16162 (2008).

95. Öberg, K. I. Photochemistry and astrochemistry: Photochemical pathways to interstellar complex organic molecules. *Chem. Rev.* **116**, 9631-9663 (2016).

96. Muñoz Caro, G. M. et al. Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **416**, 403-406 (2002).

97. De Kok, R. J., Teanby, N. A., Maltagliati, L., Irwin, P. G. J. & Vinatier, S. HCN ice in Titan's high-altitude southern polar cloud. *Nature* **514**, 65-67 (2014).

98. Burgdorf, M. et al. A tentative identification of HCN ice on Triton. *Astrophys J* **718**, L53-L57 (2010).

99. Lavvas, P. P., Coustenis, A. & Vardavas, I. M. Coupling photochemistry with haze formation in Titan's atmosphere, part ii: Results and validation with Cassini/huygens data. *Planet. Space Sci.* **56**, 67-99 (2008).

100. Teanby, N. A. et al. Vertical profiles of HCN, hc3n, and c2h2 in Titan's atmosphere derived from Cassini/cirs data. *Icarus* **186**, 364-384 (2007).

101. Coustenis, A., Schmitt, B., Khanna, R. K. & Trotta, F. Plausible condensates in Titan's stratosphere from voyager infrared spectra. *Planet. Space Sci.* **47**, 1305-1329 (1999).

102. Samuelson, R. E., Smith, M. D., Achterberg, R. K. & Pearl, J. C. Cassini cirs update on stratospheric ices at Titan's winter pole. *Icarus* **189**, 63-71 (2007).

103. Clark, R. N. et al. Detection and mapping of hydrocarbon deposits on Titan. *J. Geophys. Res. Planets* **115**, E10005 (2010).

104. Ennis, C., Cable, M. L., Hodyss, R. & Maynard-Casely, H. E. Mixed hydrocarbon and cyanide ice compositions for Titan's atmospheric aerosols: A ternary-phase co-crystal predicted by density functional theory. *ACS Earth Space Chem.* **4**, 1195-1200 (2020).

105. Sagan, C. & Khare, B. N. Tholins: Organic chemistry of interstellar grains and gas. *Nature* **277**, 102-107 (1979).

106. Perrin, Z. et al. An atmospheric origin for HCN-derived polymers on Titan. *Processes* **9**, 965 (2021).

107. Maillard, J. et al. Structural elucidation of soluble organic matter: Application to Titan's haze. *Icarus* **340**, 113627 (2020).

108. Cable, M. L. et al. Titan tholins: Simulating Titan organic chemistry in the Cassini-huygens era. *Chem. Rev.* **112**, 1882-1909 (2012).

109. Rahm, M., Lunine, J. I., Usher, D. & Shalloway, D. Polymorphism and electronic structure of polyimine and its potential significance for prebiotic chemistry on Titan. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8121-8126 (2016).

110. Morse, A. D. & Chan, Q. H. S. Observations of cometary organics: A post rosetta review. *ACS Earth Space Chem.* **3**, 1773-1791 (2019).

111. Capaccioni, F. et al. The organic-rich surface of comet 67p/churyumov-gerasimenko as seen by virtis/rosetta. *Science* **347**, aaa0628 (2015).

112. Suttle, M. D., Folco, L., Genge, M. J. & Russell, S. S. Flying too close to the sun – the viability of perihelion-induced aqueous alteration on periodic comets. *Icarus* **351**, 113956 (2020).

113. Bockelée-Morvan, D. & Biver, N. The composition of cometary ices. *Philos. Trans. R. Soc. A* **375**, 20160252 (2017).

114. Mumma, M. J. & Charnley, S. B. The chemical composition of comets - emerging taxonomies and natal heritage. *Annu. Rev. Astron. Astrophys.* **49**, 471-524 (2011).

115. Biver, N. et al. Spectroscopic monitoring of comet c/1996 b2 (Hyakutake) with the JCMT and IRAM radio telescopes. *Astrophys J* **118**, 1850-1872 (1999).

116. Paganini, L. et al. C/2013 r1 (lovejoy) at ir wavelengths and the variability of CO abundances among Oort cloud comets. *Astrophys. J.* **791**, 122 (2014).

117. Wickramasinghe, J. T., Wickramasinghe, N. C. & Wallis, M. K. Liquid water and organics in comets: Implications for exobiology. *Int. J. Astrobiol.* **8**, 281-290 (2009).

118. Rettig, T. W., Tegler, S. C., Pasto, D. J. & Mumma, M. J. Comet outbursts and polymers of HCN. *Astrophys. J.* **398**, 293-298 (1992).

119. Altwegg, K. et al. Organics in comet 67p - a first comparative analysis of mass spectra from rosina-dfms, cosac and ptolemy. *Mon. Not. R. Astron. Soc.* **469**, S130-S141 (2017).

120. Marin-Yaseli, M. R., Moreno, M., Briones, C., De, L. F. J. L. & Ruiz-Bermejo, M. Experimental conditions affecting the kinetics of aqueous HCN polymerization as revealed by uv-vis spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc* **191**, 389-397 (2017).

121. Ruiz-Bermejo, M., De La Fuente, J. L., Pérez-Fernández, C. & Mateo-Martí, E. A comprehensive review of HCN-derived polymers. *Processes* **9**, 597 (2021).

122. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Prebiotic synthesis. ii. synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **30**, 223-253 (1967).

123. He, C., Lin, G., Upton, K. T., Imanaka, H. & Smith, M. A. Structural investigation of HCN polymer isotopomers by solution-state multidimensional NMR. *J. Phys. Chem. A* **116**, 4751-4759 (2012).

124. Mamajanov, I. & Herzfeld, J. HCN polymers characterized by solid state NMR: Chains and sheets formed in the neat liquid. *J. Chem. Phys.* **130**, 134503/1-134503/6 (2009).

125. Bonnet, J.-Y. et al. Compositional and structural investigation of HCN polymer through high resolution mass spectrometry. *Int. J. Mass Spectrom.* **354-355**, 193-203 (2013).

126. Ruiz-Bermejo, M. et al. A comparative study on HCN polymers synthesized by polymerization of NH4CN or diaminomaleonitrile in aqueous media: New perspectives for prebiotic chemistry and materials science. *Chem. Eur J* **25**, 11437 - 11455 (2019).

127. Mamajanov, I. & Herzfeld, J. HCN polymers characterized by ssNMR: Solid state reaction of crystalline tetramer (diaminomaleonitrile). *J. Chem. Phys.* **130**, 134504/1-134504/5 (2009).

128. Vujošević, S. I., Negrón-Mendoza, A. & Draganić, Z. D. Radiation-induced polymerization in dilute aqueous solutions of cyanides. *Orig Life Evol Biosph.* **20**, 49-54 (1990).

129. Niketić, V., Draganić, Z. D., Nešković, S., Jovanović, S. & Draganić, I. G. Radiolysis of aqueous solutions of hydrogen cyanide (ph~6): Compounds of interest in chemical evolution studies. *J. Mol. Evol.* **19**, 184-191 (1983).

130. Draganić, Z. D., Niketić, V., Jovanović, S. & Draganić, I. G. The radiolysis of aqueous ammonium cyanide: Compounds of interest to chemical evolution studies. *J. Mol. Evol.* **15**, 239-260 (1980).

131. Schwartz, A. W. & Goverde, M. Acceleration of HCN oligomerization by formaldehyde and related compounds: Implications for prebiotic syntheses. *J. Mol. Evol.* **18**, 351-353 (1982).

132. Mas, I., De La Fuente, J. L. & Ruiz-Bermejo, M. Temperature effect on aqueous NH4cn polymerization: Relationship between kinetic behaviour and structural properties. *Eur. Polym. J.* **132**, 109719 (2020).

133. Cataldo, F., Patanè, G. & Compagnini, G. Synthesis of HCN polymer from thermal decomposition of formamide. *J. Macromol. Sci. A* **46**, 1039-1048 (2009).

134. Kliss, R. M. & Matthews, C. N. HCN dimer and chemical evolution. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 1300-1306 (1962).

135. Moser, R. E., Fritsch, J. M., Westman, T. L., Kliss, R. M. & Matthews, C. N. Hydrogen cyanide dimer. Aminocyanocarbene. *J. Am. Chem. Soc.* **89**, 5673-5661 (1967).

136. Budil, D. E., Roebber, J. L., Liebman, S. A. & Matthews, C. N. Multifrequency electron spin resonance detection of solid-state organic free radicals in HCN polymer and a Titan tholin. *Astrobiology* **3**, 323-329 (2003).

137. Eastman, M. P., Helfrich, F. S. E., Umantsev, A., Porter, T. L. & Weber, R. Exploring the structure of a hydrogen cyanide polymer by electron spin resonance and scanning force microscopy. *Scanning* **25**, 19-24 (2003).

138. Sanchez, R., Ferris, J. & Orgel, L. E. Conditions for purine synthesis: Did prebiotic synthesis occur at low temperatures. *Science* **153**, 72-73 (1966).

139. Ferris, J. P., Donner, D. B. & Lotz, W. Chemical evolution. IX. Mechanism of the oligomerization of hydrogen cyanide and its possible role in the origins of life. *J. Amer. Chem. Soc.* **94**, 6968-6974 (1972).

140. Ferris, J. P. & Orgel, L. E. An unusual photochemical rearrangement in the synthesis of adenine from hydrogen cyanide1. *J. Am. Chem. Soc.* **88**, 1074-1074 (1966).

141. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis: Iv. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. *J. Mol. Biol.* **38**, 121-128 (1968).

142. Al-Azmi, A., Elassar, A.-Z. A. & Booth, B. L. The chemistry of diaminomaleonitrile and its utility in heterocyclic synthesis. *Tetrahedron* **59**, 2749-2763 (2003).

143. Rivera, A., Ríos-Motta, J. R.-M. & León, F. Revisiting the reaction between diaminomaleonitrile and aromatic aldehydes: A green chemistry approach. *Molecules* **11**, 858-866 (2006).

144. Ferris, J. P. & Orgel, L. E. Aminomalononitrile and 4-amino-5-cyanoimidazole in hydrogen cyanide polymerization and adenine synthesis. *J. Am. Chem. Soc.* **87**, 4976-4977 (1965).

145. Evans, R. A., Lorencak, P., Ha, T. K. & Wentrup, C. HCN dimers: Iminoacetonitrile and n-cyanomethanimine. *J. Am. Chem. Soc.* **113**, 7261-7276 (1991).

146. Moser, R. E., Claggett, A. R. & Matthews, C. N. Peptide formation from aminomalononitrile (HCN trimer). *Tetrahedron Lett.* **9**, 1605-1608 (1968).

147. Thissen, H. et al. Prebiotic-chemistry inspired polymer coatings for biomedical and material science applications. *NPG Asia Mater.* **7**, e225 (2015).

148. Thissen, H., Evans, R. A. & Ball, V. Films and materials derived from aminomalononitrile. *Processes* **9**, e82 (2021).

149. Mas, I., Hortelano, C., Ruiz-Bermejo, M. & De La Fuente, J. L. Highly efficient melt polymerization of diaminomaleonitrile. *Eur. Polym. J.* **143**, 110185 (2021).

150. Hortelano, C., Ruiz-Bermejo, M. & De La Fuente, J. L. Solid-state polymerization of diaminomaleonitrile: Toward a new generation of conjugated functional materials. *Polymer* **223**, 123696 (2021).

151. Marin-Yaseli, M. R., Cid, C., Yagüe, A. I. & Ruiz-Bermejo, M. Detection of macromolecular fractions in HCN polymers using electrophoretic and ultrafiltration techniques. *Chem. Biodiversity* **14**, e1600241 (2017).

152. Mizutani, H., Mikuni, H., Takahasi, M. & Noda, H. Study on the photochemical reaction of HCN and its polymer products relating to primary chemical evolution. *Origins of life* **6**, 513-525 (1975).

153. Vuitton, V. et al. Very high resolution mass spectrometry of HCN polymers and tholins. *Faraday Discuss.* **147**, 495-508 (2010).

154. Ruiz-Bermejo, M., García-Armada, P., Mateo-Martí, E. & De La Fuente, J. L. HCN-derived polymers from thermally induced polymerization of diaminomaleonitrile: A non-enzymatic peroxide sensor based on prebiotic chemistry. *Eur. Polym. J.* **162**, 110897 (2022).

155. Ruiz-Bermejo, M. et al. New insights into the characterization of 'insoluble black HCN polymers'. *Chem. Biodiversity* **9**, 25-40 (2012).

156. Umemoto, K., Takahasi, M. & Yokota, K. Studies on the structure of HCN oligomers. *Orig Life Evol Biosph* **17**, 283-293 (1987).

157. Völker, T. Polymere blausäure. *Angew. Chem. Int. Ed.* **72**, 379-384 (1960).

158. Garbow, J. R., Schaefer, J., Ludicky, R. & Matthews, C. N. Detection of secondary amides in hydrogen cyanide polymers by dipolar rotational spin-echo nitrogen-15 NMR. *Macromolecules* **20**, 305-309 (1987).

159. Purvis, G. D., III & Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **76**, 1910-1918 (1982).

160. Born, M. & Oppenheimer, R. Zur quantentheorie der molekeln. *Ann. Phys.* **389**, 457-484 (1927).

161. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* **136**, 864-871 (1964).

162. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).

163. Thijssen, J. *Computational physics* (Cambridge university press, 2007).

164. Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372-1377 (1993).

165. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996).

166. Peterson, K. A., Feller, D. & Dixon, D. A. Chemical accuracy in ab initio thermochemistry and spectroscopy: Current strategies and future challenges. *Theor. Chem. Acc.* **131**, 1079 (2012).

167. Goerigk, L. et al. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **19**, 32147-32744 (2017).

168. Mangiatordi, G. F., Bremond, E. & Adamo, C. DFT and proton transfer reactions: A benchmark study on structure and kinetics. *J. Chem. Theory Comput.* **8**, 3082-3088 (2012).

169. Goerigk, L. & Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **13**, 6670-6688 (2011).

170. Zhao, Y., González-García, N. & Truhlar, D. G. Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods. *J. Phys. Chem. A* **109**, 2012-2018 (2005).

171. Zhao, Y. & Truhlar, D. G. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A* **109**, 5656-5667 (2005).

172. Schwerdtfeger, P. The pseudopotential approximation in electronic structure theory. *ChemPhysChem* **12**, 3143-3155 (2011).

173. Bloechl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter* **50**, 17953-17979 (1994).

174. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1-5 (2015).

175. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999-3093 (2005).

176. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378-6396 (2009).

177. Mennucci, B. Polarizable continuum model. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2**, 386-404 (2012).

178. Frisch, M. J. et al. *Gaussian 16, revision b.01* (Gaussian, Inc., Wallingford CT, 2016).

179. Ryu, H. et al. Pitfalls in computational modeling of chemical reactions and how to avoid them. *Organometallics* **37**, 3228-3239 (2018).

180. Tuckerman, M. E. & Martyna, G. J. Understanding modern molecular dynamics: techniques and applications. *J. Phys. Chem. B* **104**, 159-178 (2000).

181. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

182. Bussi, G. & Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2**, 200-212 (2020).

183. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *WIREs Comput Mol Sci* **1**, 826-843 (2011).

184. Branduardi, D., Gervasio, F. L. & Parrinello, M. From a to b in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).

185. Pietrucci, F. & Saitta, A. M. Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15030 (2015).

186. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291-318 (2002).

187. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J Comput Phys* **23**, 187-199 (1977).

188. Flyvbjerg, H. in *Advances in Computer Simulation* (eds Kertész, J. & Kondor, I.) 88-103 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1998).

189. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456-1465 (2011).

190. Hutter, J., Iannuzzi, M., Schiffmann, F. & Vandevondele, J. Cp2k: Atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 15-25 (2014).

191. Mangiatordi, G. F., Brémond, E. & Adamo, C. Dft and proton transfer reactions: A benchmark study on structure and kinetics. *J. Chem. Theory Comput.* **8**, 3082-3088 (2012).

192. Moser, R. E., Claggett, A. R. & Matthews, C. N. Peptide formation from diaminomaleonitrile (HCN tetramer). *Tetrahedron Lett.* **9**, 1599-1603 (1968).

193. Matthews, C. N. & Moser, R. E. Prebiological protein synthesis. *Proc Natl Acad Sci U S A* **56**, 1087-1094 (1966).

194. Ferris, J. P. Hydrogen cyanide did not condense to give heteropolypeptides on the primitive Earth. Comments. *Science (Washington, D. C.)* **203**, 1135-1136 (1979).

195. Rivilla, V. M. et al. Abundant z-cyanomethanimine in the interstellar medium: Paving the way to the synthesis of adenine. *Mon. Not. R. Astron. Soc. Lett.* **483**, L114-L119 (2018).

196. Zaleski, D. P. et al. Detection of e-cyanomethanimine toward Sagittarius b2(n) in the Green Bank Telescope primos survey. *Astrophys J* **765**, L10 (2013).

197. Kikuchi, O., Watanabe, T., Satoh, Y. & Inadomi, Y. Ab initio GB study of prebiotic synthesis of purine precursors from aqueous hydrogen cyanide: Dimerization reaction of HCN in aqueous solution. *J Mol Struct THEOCHEM* **507**, 53-62 (2000).

198. Sandström, H. & Rahm, M. The beginning of HCN polymerization: Iminoacetonitrile formation and its implications in astrochemical environments. *ACS Earth Space Chem.* **5**, 2152-2159 (2021).

199. Levy, M., Miller, S. L., Brinton, K. & Bada, J. L. Prebiotic synthesis of adenine and amino acids under Europa-like conditions. *Icarus* **145**, 609-613 (2000).

200. Fernández, A., Ruiz-Bermejo, M. & De La Fuente, J. L. Modelling the kinetics and structural property evolution of a versatile reaction: Aqueous HCN polymerization. *Phys. Chem. Chem. Phys.* **20**, 17353-17366 (2018).

201. Pracht, P., Bohle, F. & Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **22**, 7169-7192 (2020).

202. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387-387 (2001).

203. Cottini, V. et al. Spatial and temporal variations in Titan's surface temperatures from Cassini cirs observations. *Planet. Space Sci.* **60**, 62-71 (2012).

204. Anslyn, E. V. & Dougherty, D. A. *Modern physical organic chemistry* (University Science, Sausalito, CA, 2006).

205. Stevenson, J., Clancy, P. & Lunine, J. Membrane alternatives in worlds without oxygen: Creation of an azotosome. *Sci Adv* **1**, e1400067 (2015).

206. Palmer, M. Y. et al. ALMA detection and astrobiological potential of vinyl cyanide on Titan. *Sci Adv* **3**, e1700022 (2017).

207. Berland, K. & Hyldgaard, P. Exchange functional that tests the robustness of the plasmon description of the van der waals density functional. *Phys. Rev. B* **89**, 035412 (2014).

208. Sandström, H. & Rahm, M. Can polarity-inverted membranes self-assemble on Titan. *Sci Adv* **6**, eaax0272 (2020).

209. Setzmann, U. & Wagner, W. A new equation of state and tables of thermodynamic properties for methane covering the range from the melting line to 625 k at pressures up to 100 mpa. *J. Phys. Chem. Ref. Data* **20**, 1061-1155 (1991).

210. Stevenson, J. M. et al. Solvation of nitrogen compounds in Titan's seas, precipitates, and atmosphere. *Icarus* **256**, 1-12 (2015).

211. Facchin, M., Scarso, A., Selva, M., Perosa, A. & Riello, P. Towards life in hydrocarbons: Aggregation behaviour of "reverse" surfactants in cyclohexane. *RSC Adv.* **7**, 15337-15341 (2017).

212. Carrer, M. et al. Can polarity-inverted surfactants self-assemble in nonpolar solvents. *J. Phys. Chem. B* **124**, 6448-6458 (2020).

213. Bannwarth, C. et al. Extended tight-binding quantum chemistry methods. *WIREs Comput Mol Sci* **11**, e1493 (2021).

214. Liang, T. et al. Reactive potentials for advanced atomistic simulations. *Annu. Rev. Mater. Res.* **43**, 109-129 (2013).

215. Guo, F. et al. Intelligent-ReaxFF: Evaluating the reactive force field parameters with machine learning. *Comput. Mater. Sci.* **172**, 109393 (2020).

216. Yang, Z. et al. Influence of water and enzyme spnf on the dynamics and energetics of the ambimodal [6+4]/[4+2] cycloaddition. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E848-E855 (2018).

217. Cascioli, G. et al. The determination of the rotational state and interior structure of Venus with Veritas. *Planet. Space. Sci.* **2**, 220 (2021).

218. European Space Agency (ESA). EnVision, understanding why Earth's closest neighbour is so different. *https://www.cosmos.esa.int/documents/5763359/5763378/EnVision_YB_final.pdf/e9612355-67de-42a3-c25a-af1683f6fda3?t=1616679461807 (*(2021).

219. Osinski, G. R. et al. The canMars Mars sample return analogue mission. *Planet Space Sci.* **166**, 110-130 (2019).

220. Yan, J., Qingguo, D. & Chi, F. Y. C. W. Development progress of China's first Mars exploration mission: Its scientific objectives and payloads. *Chin. J. Space Sci.* 空间科学学报 **40**, (2020).

221. Farley, K. A. et al. Mars 2020 mission overview. *Space Sci. Rev.* **216**, 142 (2020).