Insights into Messy Chemistry Related to Cosmology and Origin of Life Obtained by Employing State-of-the-art Computational Methods

by

Tamal Das

10CC16J26001

A thesis submitted to the
Academy of Scientific & Innovative Research
for the award of the degree of
DOCTOR OF PHILOSOPHY
in
SCIENCE

under the supervision of
Dr. Kumar Vanka

CSIR-National Chemical Laboratory, Pune

Academy of Scientific and Innovative Research
AcSIR Headquarters, CSIR-HRDC campus
Sector 19, Kamla Nehru Nagar,
Ghaziabad, U.P. - 201002, India

April 2021

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, " *Insights into Messy Chemistry Related to Cosmology and Origin of Life Obtained by Employing State-of-the-art Computational Methods*" submitted by *Tamal Das* to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of *Doctor of Philosophy In Science* embodies original research work carried-out by the student. We, further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material(s) obtained from other source(s) and used in this research work has/have been duly acknowledged in the thesis. Image(s), illustration(s), figure(s), table(s) etc., used in the thesis from other source(s), have also been duly cited and acknowledged.

Tamal Das

Kumar Vanka

(Signature of Student)	(Signature of Co-Supervisor)	(Signature of Supervisor)
Name: Tamal Das	No	Name: Dr. Kumar Vanka
Date: 31-03-2021		Date: 31-03-2021

STATEMENTS OF ACADEMIC INTEGRITY

I, Tamal Das, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. 10CC16J26001 hereby undertake that, the thesis entitled "Insights into Messy Chemistry Related to Cosmology and Origin of Life Obtained by Employing State-of-the-art Computational Methods" has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on "*Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)*" and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)*".

Tamal Das

Signature of the Student

Date: 31-03-2021

Place: CSIR-NCL, Pune

It is hereby certified that the work done by the student, under my/our supervision, is plagiarism-free in accordance with the UGC Regulations on "*Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)*" and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)*".

Kumar Vanka

Signature of the Co-supervisor

no

Signature of the Supervisor

Name: Dr. Kumar Vanka

Date: 31-03-2021

Place: CSIR-NCL, Pune

Dedicated to My Family

Acknowledgement

No one knows what destiny a drop of water will fall to while it leaves the cloud. Whether it will get evaporated even before reaching the ground or will meet an ember or a dust mound to lose its existence. Whether it will fall on a seed to help that start a new life or will get into the mouth of a thirsty bird to quench his thirst. It is the speed and direction of the air and the temperature in the region, which determine the fate of the drop. Similarly, I knew nothing about what will I meet in my academic career when I left my home for the first time after MSc. for higher studies. My doctoral research journey was an unusual learning path in my life, which, finally, comes to a beautiful end. The path was composed of mixed emotions, and I met many supportive people, without whom it would be impossible to enjoy the journey and reach the destination.

First, I would like to express my gratefulness to my research supervisor, Dr. Kumar Vanka, for his constant guidance and support. Without him, it would have been impossible to reach this point. Furthermore, I am also grateful for the academic freedom he has given me, which motivated me to grow as an independent researcher. Also, his teaching, writing and communication skills have been inspirational to me throughout my doctoral research.

I would also thank my present and past Doctoral Advisory Committee members, Dr. Debashree Ghosh, Dr. Nayana Vaval, Dr. Sakya S. Sen, and DAC chairperson Dr. Ashok Giri for their insightful suggestions and feedback. I am also grateful to Dr. Ashwini Kumar Nangia, director, CSIR-NCL, and Dr. Sourav Pal, former director of CSIR-NCL. Moreover, I take the opportunity to thank the present and former heads of the Physical and Materials Chemistry Division for their support and providing all facilities during my Ph.D.

I want to extend my gratitude to all teachers who had taught me during my Ph.D. course work at CSIR-NCL: Dr. Kumar Vanka, Dr. Suman Chakrabarty, Dr. Syan Bagchi, Dr. Debashree Ghosh, Dr. Rajesh Gonnade and Dr. T.G. Ajithkumar, as well as other scientists in NCL. Let me extend my warm thanks to my research collaborators Dr. Sakya S. Sen, Dr. Santoshbabu Sukumaran, Dr. Ravi P. Singh and Dr. Pradip Maity Moreover, I owe to thank my University and school teachers for

their support and motivation.

I also acknowledge all the non-academic staff of CSIR-NCL, and AcSIR, for their support and help during my work. Without the funding, this Ph.D. journey would not have been possible; hence I would like to express my gratitude to the Council of Scientific & Industrial Research (CSIR) for the fellowship. Moreover, I want to thank the whole scientific community for being a source of inspiration and motivation.

No words are enough to thank my friends, who have helped and supported me at various stages. I feel myself to be lucky to have dearest friends and roommates Tapas, Subhrashis and Debu. I am immensely thankful to them for being there with me during my ups and downs during my whole Ph.D tenure. My initial days in NCL Pune was memorable with Somenath and Pooja. I thank them for their company and support.

I also extend my thanks to Jugal, Mrintyunjay, Amrita, Yuvraj for their help whenever needed during my Ph.D. Also, my special thanks to Siddarth for helping me a lot. Moreover, it is my pleasure to thank my past and present lab mates; Shantanu, Nisha, Manoj, Vipin, Shailja, Ruchi , Anagh and Soumya. I also thank Vrushali, Amit, Pranab, Turbasuda, Achintyada, Susantada, Himadrida, Aaryada, Sudip, Deepak, Ravi, Avijit, ashis, Rahul, Pronoy, Milan, Sutanu, Saibal, Sibaprasad, Subrata, Abdul, Sanjukta, Gargi and many more.

I want to thank the most important people in my life, my family. No words are sufficient to describe their love, affection, and support. I am highly indebted to my Father (Chandra Sekhar Das) and Mother (Mamata Das) for their unconditional love and the sacrifices they have made for me in their lives. Also, I wish to thank my only brother (Tanmoy) for taking responsibility for my family, which helped me focus on my research work. Furthermore, I would like to thank my all family members cousin brothers and sisters for their mental and emotional support. I consider myself blessed to have such a beautiful family around me.

Finally, I express my gratitude to the Almighty for the blessings and for providing special challenges.

- **Tamal Das**

“No one who does good work will ever come to a bad end, either here or in the world to come”

- Bhagavad Gita

“In a day, when you don't come across any problems - you can be sure that you are travelling in a wrong path”

— Swami Vivekananda

Table of Contents

List of Figures

List of Tables

List of Schemes

Abbreviations

AIMD	*Ab initio* Molecular Dynamics
AINR	*Ab initio* Nanoreactor
B3LYP	Becke, 3-parameter, Lee-Yang-Parr
BOMD	Born-Oppenheimer Molecular Dynamics
CPMD	Car-Parrinello Molecular Dynamics
COSMO	Conductor-like Screening Model
DFT	Density Functional Theory
EMD	Ehrenfest Molecular Dynamics
GGA	Generalized Gradient Approximation
IRC	Intrinsic Reaction Coordinate
KS	Kohn-Sham
LDA	Local Density Approximation
MP2	Second-order Møller–Plesset Perturbation Theory
M06	Minnesota 06
NPA	Natural Population Analysis
PBE	Perdew, Burke and Ernzerhof
PCM	Polarizable Continuum Model
RI	Resolution of Identity SCF Self-consistent Field
TZVP	Triple Zeta Valence plus Polarization
TS	Transition state

Physical Constants

Avogadro's Constant	$(NA) = 6.02214129 \times 10^{23}\ mol^{-1}$
Atomic Mass Unit	$(u) = 1.660538921 \times 10^{-27}\ kg$
Boltzmann's Constant	$(k) = 1.3806488 \times 10^{-23}\ JK^{-1}$
Bohr Radius	$(a_0) = 5.291772109 \times 10^{-11}\ m$
Elementary Charge	$(e) = 1.602176565 \times 10^{-19}\ C$
Gas Constant	$(R) = 8.3144621\ JK^{-1}mol^{-1}$
Mass of Electron	$(m_e) = 9.10938291 \times 10^{-31}\ kg$
Mass of Proton	$(m_p) = 1.672621777 \times 10^{-27}\ kg$
Mass of Neutron	$(m_n) = 1.674927351 \times 10^{-27}\ kg$
Rydberg Constant	$(R) = 1.097373157 \times 10^{7}\ m^{-1}$
Speed of Light	$(c) = 2.99792458 \times 10^{8}\ ms^{-1}$
Planck's Constant	$(h) = 6.62606957 \times 10^{-34}\ Js$

Chapter 1

A Brief Introduction about the Approach of Messy Chemistry Towards Cosmology and Prebiotic Chemistry

Chapter 1

A Brief Introduction about the Approach of Messy Chemistry Towards Cosmology and Prebiotic Chemistry

Abstract

How life began on the early earth, as well as how small molecules and ions were first formed at the beginning of the universe are some of the biggest unsolved questions in science that have intrigued researchers over time. However, no clear answers have yet been received for researchers in the fields of cosmology, prebiotic chemistry, interstellar chemistry as well as astrochemistry. From previous studies in the areas of astronomy, astrochemistry and astrobiology it has become clear that the creation of the universe and the origin of life are not the result of any single event. Rather, they are likely the product of highly complex or "messy" chemical processes. In this chapter, an overview has been provided on how messy chemistry plays a role in cosmology, as well as in prebiotic chemistry.

1.1 Introduction:

The way the universe, and all the elements, came into being is one of the fascinating questions of science. Attempts to answer this question has led to the Big Bang Theory, and an understanding of the early universe and the entities that it was made up of.[1] Based on a series of calculations and observations through telescopes, the best explanation is that the Big Bang occurred 13.8 billion years ago. Due to a brutal explosion, all the matter in the universe emerged from a single, minute point. Hydrogen and helium nuclei formed due to the combination of protons and neutrons, which formed within the very first second. After 0.3 million years, by capturing the electrons, nuclei could form atoms and finally, the universe filled with helium and hydrogen gas. Advancements in science and technology have resulted in greater understanding, which led NASA's Stratospheric Observatory for Infrared Astronomy (SOFIA) to the detection of HeH^+, the first molecule formed after the Big Bang,[2] 94 years after its discovery in the laboratory in 1925.[3]

Origin of life on the early earth[4-10] is just as fascinating a question for the human race as the origin of the universe. It has been extremely difficult for the specialists in the field of cosmology, prebiotic science, interstellar chemistry and astrochemistry to offer a comprehensive and satisfactory explanation for the origin of life on earth. It is well known that life on earth is made up of cells, which are composed of small organic molecules made up of the atoms of oxygen, hydrogen, nitrogen, phosphate, carbon and sulphur. Therefore, it stands to reason that, on today's Earth, all life arose from very small molecules that were present on the early earth atmosphere or arrived on earth from outer, celestial bodies. Unfortunately, till date, it has not been possible to generate or synthesize a living cell from organic molecules in the laboratory. So, how could life have appeared on Earth? The age of the Earth is now 4.6 billion years. The time taken for chemical evolution is estimated to have been almost a billion years and ultimately biological evolution took place over a very long time span. When scientists and the researchers ask questions on how chemical evolution took place, they do not directly address how life arose in the early earth. Rather, they are focused on the intermediate steps of chemical evolution, based on different theories and hypotheses. This journey was begun more than a half-century ago by Stanley Miller and Harold Urey, when they conducted their famous Urey-Miller experiment in 1959.[11] After this epoch-making year, a lot of experimental studies[12-19] and some computational

investigations[20-23] have been performed by various research groups all over the world. From their studies, it has been clear that the creation of the universe and origin of life is not any single event – rather, it is highly complex or "messy". In this thesis, by employing state-of-the-art computational methods, we have tried to tackle the "messy" chemistry behind the origin of life, as well as the formation of the universe's first and simple small molecules.

1.2 What is messy-chemistry?

More than six decades ago, a famous experiment[11] was conducted by Stanley Miller and Harold Urey. They took a gaseous mixture, which was supposed to have been present in the atmosphere of the primitive Earth, into a flask. They sparked the mixture with an electric discharge. After several days they observed that amino acids and other chemical building units of life were produced in the flask. The experiment was addressed as a path-breaking production of how the important building units of life may have been produced from rather simple components. Now, a lot of messy substances were produced during the experiment, which covered the inside part of the beaker, and would have been considered unimportant residue from the experiment, by most researchers. However, in recent years, some plucky researchers have begun looking at the messy residue from a different point of view. They contend that the messy or sticky tar – formed by the reaction between small organic molecules in the presence of an energy source — may be offering a pathway that could lead to many advances in the field of prebiotic chemistry. These studies, focused on taking off the tar and investigating it carefully, is called "messy chemistry," today, as opposed to the "clean" or "clear" chemistry that focused on the well known organic compounds that were considered precursors to proteins and sugars. Messy chemistry is now considered as an ignored but an up-and-coming way forward. In typical synthetic chemistry and biology studies, one considers a particular reaction and characterizes its results. However, in real life, such an approach would be flawed: there are no characterizing reactions, but a rather complex set of chemical processes taking place. Therefore, we can ask: 'why not look at the entire complex system?'"

1.3 Cosmology

The branch of astronomy connected with the studies of the origin and evolution of the universe from the Big Bang[2] to the current day, and even the future is known as cosmology.

Cosmology, in short, is the scientific study of the origin, evolution, and inevitable fate of the universe. Physical cosmology[24] deals with the scientific study of the origin of the universe, its structures and dynamics as well as its ultimate fate, as well as the laws of science that govern these areas. NASA's definition of cosmology is "the scientific study of the large scale properties of the universe as a whole."

Thomas Blount was the first to use in English the term cosmology in 1656.[25-26] The body of beliefs based on religious, mythological and abstruse literature and the custom of creation myths is known as religious or mythological cosmology. On the other hand, scientists, such as physicists and astronomers, as well as the philosophers of space and time, studied the physical cosmology in detail. Cosmology is different from astronomy in that the former deals with the whole universe, while the latter is concerned with individual celestial bodies. Modern physical cosmology, which attempts to bring together astronomy and particle physics,[27-28] is influenced by the Big Bang theory.

1.4 Prebiotic Chemistry

If one can define prebiotic chemistry as the study of the elementary chemical reaction steps, which lead to the formation of first-ever organisms, then what is also needed is a clear definition of a "living organism". But unfortunately, there is no such explicit or universally accepted definition for "living things". Under such circumstances, probably the best way to be handle questions connected to "the origin of life" is to use the word "messy" or "fuzzy".[29] For researchers who are interested in the elementary reaction steps and try to find the mechanistic pathways towards the formation of first living cell, the inadequacy of definition of the final goal is a little annoying: it offers ambiguity. Therefore, prebiotic chemistry is a furry or messy field! One could also be define prebiotic chemistry as the study of the evolution from "nonliving systems" to "living systems."[30] Therefore, prebiotic chemistry deals with the "how" of the formation of organic compounds and the self-organization of small organic molecules that lead to the origin of life on earth or any outer celestial bodies[31] of the universe.

1.5 Literature precedence towards cosmology

The study of the role of molecular hydrogen as a crucial cooling agent began in the late 1960s for the formation of the first neutral species in the early universe. Saslaw and Zipoy[32] first

pointed out the role of molecular hydrogen for the thermal and dynamical transformation of pre-cosmic gas clouds in the post-recombination period. This was unlike earlier experiments that had avoided the possible presence of molecular H_2 due to the lack of dust grains and the slower three-body reactions rate and radiative association processes. The proposed probability of charge transfer reactions and radiative association was considered by Saslaw and Zipoy, for reactions such as[33]

$$H + H^+ \rightarrow H_2^+ + h\nu \qquad (1)$$

and

$$H_2^+ + H \rightarrow H_2 + H^+ \qquad (2)$$

Bates calculated[34] the reaction (1) rate constant. At the same time, the cross-section for reaction (2) was predicted to have a comparatively higher value ($\sim 10^{-15}$ cm^2). Therefore, an H_2^+ ion could be transformed to H_2 very quickly - as soon as it formed, and as a result, the H_2^+ concentration remains very low.

In astronomy the interstellar medium (ISM) is the matter and radiation that exist in the space between the star systems and a galaxy. This matter includes gas in ionic, atomic and molecular form. The ISM is composed primarily with hydrogen followed by helium. The thermal pressures of these phases are in equilibrium with one another. In all phases, the interstellar medium is extremely tenuous by terrestrial standards. In cool, dense regions of the ISM, matter is primarily in molecular form, and reaches number densities of 10^6 molecules per cm^3 (1 million molecules per cm^3). In hot, diffuse regions of the ISM, matter is primarily ionized, and the density may be as low as 10^{-4} ions per cm^3. This can be called the low density environment. Compare this with a number density of roughly 10^{19} molecules per cm^3 for air at sea level, and 10^{10} molecules per cm^3 (10 billion molecules per cm^3) for a laboratory high-vacuum chamber. The abundance of H_2 in the interstellar medium was calculated by McDowell[35] subsequent to an earlier communication by Dalgarno. According to their scheme, the formation of molecular H_2 advances through the reactions

$$H + e \rightarrow H^- + h\nu \qquad (3)$$

and

$$H^- + H \rightarrow H_2 + e \qquad (4)$$

By taking the cross-section evaluated by Chandrasekhar,[36] the rate constant for reaction (3) was calculated. On the other hand, based on the experiments by Ferguson *et al.*[37], the rate constant for reaction (4) was determined.

The incorporation in the chemical network of the three reactions

$$H + H + H \rightarrow H_2 + H \qquad (5)$$

and

$$H + H + H_2 \rightarrow H_2 + H_2 \qquad (6)$$

altered this picture appreciably.[38] It was observed that, over a wide range of primitive conditions, when the densities are in the order of 10^{12} cm^{-3}, gases can be converted to their molecular form. Major studies[39-54] in this area have been done by several research groups.

1.5.1 The Chemistry of Helium

Molecular helium ion (He_2^+) and helium hydride ion (HeH^+) have been argued for as the first-ever molecular species formed after the Big Bang[2] in the universe. HeH^+ was detected in space, either in high-redshift absorbers[55] or in gaseous nebulae.[56-58] HeH^+ and He_2^+ were formed in the primitive universe by the radiative association process of atomic He with H^+ and He^+, respectively. In 1978 Saha *et al*. determined[59] the rate of the former process for the first time. Several groups have determined this theoretically[59-65]. In 2007, Pedersen *et al*.[66] employed very high photon energy (38.7 eV) free-electron laser FLASH to measure the absolute photodissociation cross-section, although at high photon energy (38.7 eV). The enhancement of the rate of photodissociation produced by high-energy photons has also been addressed by several studies[67-68]. In 1993, Stancil *et al.* determined[69] the rate of He_2^+ formation and found that it depended on temperature - like that of the radiative association of H and H^+.[69] Since He^+ briskly recombines, an abundance of He_2^+ reaches a maximum value and then is readily destroyed by dissociative recombination and photodissociation. In comparison, the abundance of HeH^+ steadily rises due to the partial recombination of H^+. Photodissociation, collisions with

H,[70-71] and dissociative recombination[72] processes are majorly responsible for the choppy behavior of the HeH^+ evolution. As mentioned in the previous section, the annihilation of HeH^+ by collisions with H atoms is an extensive source of H_2^+ and accounts for its enhancement at the smallest redshifts.

After the Big Bang, nucleosynthesis recombined the ions of the light elements formed in a reversal of their ionization potential. Due to the higher ionization potentials, He^+ (24.6 eV) and He^{2+} (54.5 eV) quickly combined with free electrons to form the universe's first neutral atom, preceding the recombination process of H^+ (13.6 eV). During that period, in a low-density environment, neutral helium atoms made a bond with a proton, leading to the formation of the universe's first molecule helium hydride ion HeH^+, by a radiative association process

$$He + H^+ \rightarrow HeH^+ + h\nu \qquad (7)$$

As this recombination process advanced, the annihilation of HeH^+ shown in equation 8, occurred:

$$HeH^+ + H \rightarrow He + H_2^+ \qquad (8)$$

which led towards the formation of the hydrogen molecule, as is posited today. Though this makes it very important for the evolution of the primitive universe, the HeH^+ ion has so far only been detected in interstellar space. The ion was first discovered in the laboratory in 1925 by Hogness *et al.*,[3] and its astrophysical existence was discussed[73-76] in the late seventies. It was thought that the conditions in planetary nebulae are probably convenient for the formation of HeH^+. The HeH^+ is supposed to form by the radiative association process of He^+ and H in the Stromgren sphere. The rotational ground state transition for HeH^+ was recently accessed by the GREAT spectrometer[77-78] onboard SOFIA.[79] Very recently, NASA's spacecraft SOFIA detected HeH^+ in the planetary nebula. In 2019 Novotny *et al*. reported[80] that the dissociative reaction mechanism rate of HeH^+ in presence of electrons is highly rotational state-specific. They obtained a significant decrease in the dissociation rate for the lowest rotational level by using a cryogenic ion storage ring merged with the electron beam.

1.5.2 Radiative association of H_3^+

Though the reaction

$$H_2 + H^+ \rightarrow H_3^+ + h\nu \qquad (9)$$

is the major pathway for H_3^+ formation, there is uncertainty up to four orders of magnitude in the value of the rate constant. The ion trap measurement and a classical trajectory analysis by Gerlich and Horning[81] endorsed the rate constant value of 1×10^{-16} cm^3 s^{-1}. In most of the chemical models,[82-83] this value has been accepted. However, on the basis of quantum mechanical calculations by Stancil *et al.*[84] it was argued that the rate should be much smaller for other diatomic species, which has also been supported in recent experiments.[54] The generation of H_3^+ is certainly affected by such a large uncertainty. Therefore, the effective role of H_3^+ in the above-mentioned reaction (9) needs to be checked further.

1.6 Literature precedence towards prebiotic chemistry

Finding the solution for the origin of life is a major task that belongs to the field of chemistry, biology as well as physics. Extensive research has led to two hypotheses behind the origin of life: according to "RNA World" advocates,[85-89] life on earth originated from the molecules of RNA , which is the polymeric form of activated ribonucleotides . RNA not only acts as a carrier of genetic information, but also behaves as a chemical catalyst. According to the "metabolism-first" principle, [90] instead of protein-based enzymes, there were simple metal catalysts that were present on the early earth and they may have created a soup of organic building blocks that could have formed the necessary biomolecules.

According to Oparin's hypothesis and some other experiments, it has been proposed that the primitive earth atmosphere was reducing in nature. It contained H_2, CH_4, N_2, NH_3, H_2S, HCN, PH_3, hydrocarbons, and smaller amounts of CO_2, CO, O_2, NO_X, SO_X - not more than a few ppm in a highly reducing atmosphere.

In the year of 1959, Stanley Miller and Harold Urey[11] investigated how inorganic compounds led to the formation of organic molecules in the laboratory. Their experiment produced a number of various small organic molecules along with amino acids, which are very important precursors that lead to the formation of complex life building living units. They have started the reaction in a gaseous chamber containing H_2, CO, NH_3, CH_4 and H_2O and employed a very high energy electric spark, which served as a replacement for UV radiation and lightning that was present in the prebiotic earth atmosphere. This was a revolutionary experiment that led

to researchers getting involved in this field. After this experiment, scientists have, in the last few decades, tried to synthesize amino acids[11,91-94] under different conditions.

In 1960, Oro synthesized[13] purine nucleic acid base, and adenine *via* oligomerization of hydrogen cyanide (shown in Figure1.1).

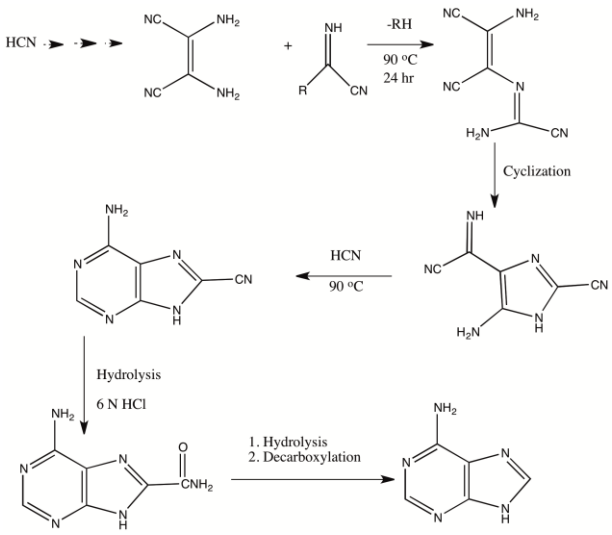
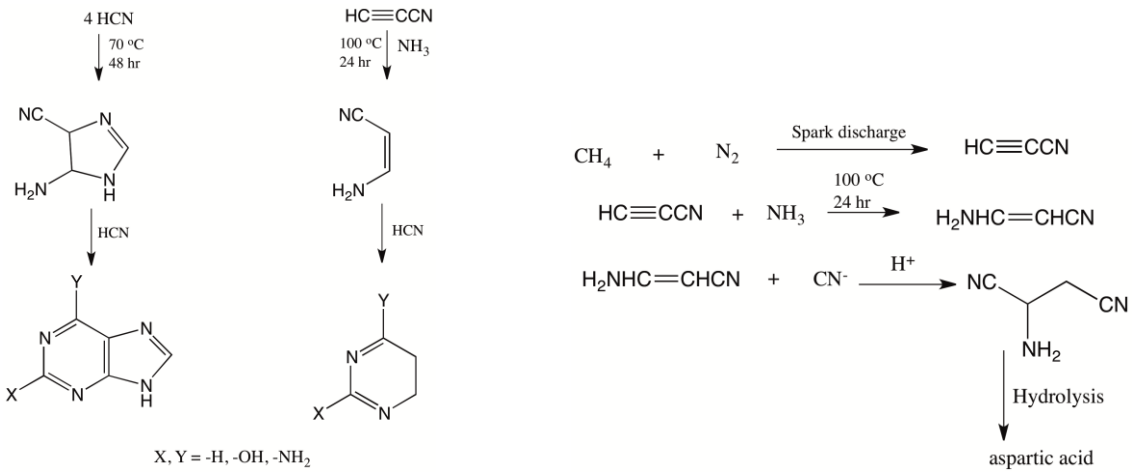


Figure 1.1 Oro synthetic route of purine nucleic acid base (adenine).

In 1965, Oro[95] had prebiotically synthesized a pyrimidine and purine base (shown in figure 1.2A) from a β-aminoacrylonitrile compound. After that, scientists had thought that cyano compounds might be the source of almost all the life building block precursors. In 1966, Orgel *et al.*[12] synthetically showed that cyanoacetylene is a major nitrogen-containing product formed (shown in Figure 1.2B) during the reaction between a mixture of methane and nitrogen by employing an electric discharge. Then, cyanoacetylene was seen to react with simple inorganic substances in an aqueous medium to produce cytosine, aspartic acid and asparagine, which are important precursors of the life building blocks.



(A) **(B)**

Fig 1.2. **(A)** Proposed synthetic pathway for pyrimidine and purine bases and **(B)** Synthetic pathways for cyanoacetylene.

Subsequently, after two years, in 1968 Sanchez *et al.*[96] synthesized pyrimidine nucleic acid bases (shown in Figure 1.3) from cyanoacetylene and cyanate.

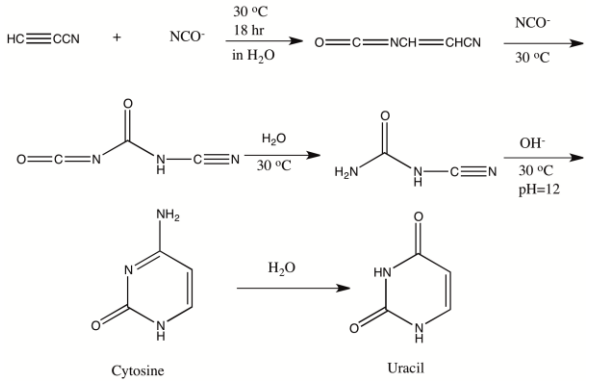


Figure 1.3 Pyrimidine nucleic acid base from cyanoacetylene and cyanate.

Several chiral amino acids had been synthesized (shown in Figure 1.4B) by Miller and coworkers[97] in 1969 from the reaction between glycine nitrile and ketones, though Strecker had synthesized an amino acid (shown in Figure 1.4A) the first time in 1854 *via* the well known Strecker pathway.[98]

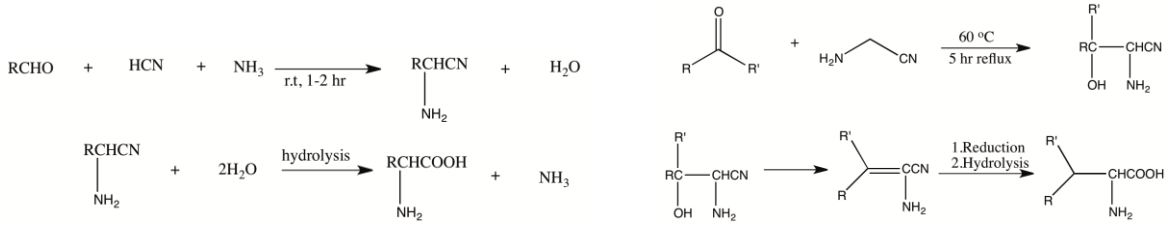


Figure 1.4 (A) Strecker pathway for amino acid synthesis. **(B)** Miller amino acids synthesis pathway.

In the year 1984, Hagan *et al*[99] showed the possible role of cyano compounds in prebiotic synthesis. Mainly, the role of HCN, which is the least molecular weight cyano compound was highlighted. It has been proposed that the appearance of HCN on early earth is due to the bombardment of organic matter with excess energy.[100] Also, it could have been widely available in the outer celestial bodies in the interstellar medium,[101-104] the solar system and beyond in comets, as well as in the atmosphere of planets and their satellites. Moreover, HCN plays a crucial role in prebiotic chemistry, mainly because of its involvement in the Strecker-type synthesis of amino acids[98] and in the Oro´ synthesis of adenine.[13]

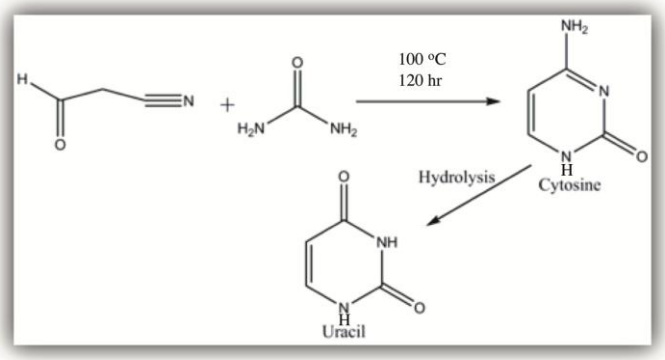


Figure 1.5 Miller synthesis of cytosine and uracil nucleic acid bases under prebiotically plausible conditions.

In the year 1995, Miller *et al*[105] prebiotically synthesized cytosine and uracil pyrimidine nucleobases (shown in Figure 1.5). Previously, cytosine had been synthesized from

cyanoacetylene and cyanate,[106] but this reaction required a relatively high concentration of cyanate, which was unlikely in the aqueous medium, because in the aqueous medium, cyanate hydrolyzed to give CO_2 and NH_3. In their work, they have used a concentrated urea solution and cyanoacetaldehyde (produced from the hydrolysis of cyanoacetylene) to produce cytosine up to 50.0% yield. Uracil can be produced from this reaction by further hydrolysis. These reactions contribute a plausible mechanistic pathway towards the pyrimidine bases required to support the RNA world hypothesis.[107]

Before John D. Sutherland, there has been no such prebiotically plausible synthetic route for the synthesis of ribonucleotides, the RNA precursor which supports the "RNA World" hypothesis.[85-89] The ribonucleotides unit made up of three different parts; ribose sugar, a phosphate group, and pyrimidine (cytosine and uracil) or purine (adenine and guanine) base. The previously assumed pathway for the formation of ribonucleotides from sugar, nucleobase, and phosphate is implausible, because the condensation that occurs between cytosine and the ribose sugar base is highly unlikely.[108] After a long time interval in 2009, Sutherland's group came up with an idea that there may be another feasible pathway by which one could synthesize ribonucleotides starting from the same small molecule (shown in Figure 1.6) precursors but in a systematic way. According to Szostak, this is characterized as "system chemistry".[109] Sutherland *et al.* first managed to synthesize activated ribonucleotides[110] in a completely different and feasible pathway, starting from five different starting materials – cyanamide, cyanoacetylene, glycolaldehyde , glyceraldehydes, and inorganic phosphate under photoredox conditions and a phosphate buffer solution in order to mimic prebiotically plausible conditions. In a one pot synthesis, they got a very low yield and therefore they did a stepwise synthesis for the improvement of yield and used phosphate from the beginning of the reaction to prevent the formation of unwanted side products. The key findings of their work was the separation of the carbon-oxygen chemistry (formation of sugar) and carbon-nitrogen chemistry (formation of nucleobase) as long as possible so that to avoid the condensation reaction, which is highly unfavorable. Their synthesis of activated ribonucleotides strongly supports the "RNA world" hypothesis.[85-89]

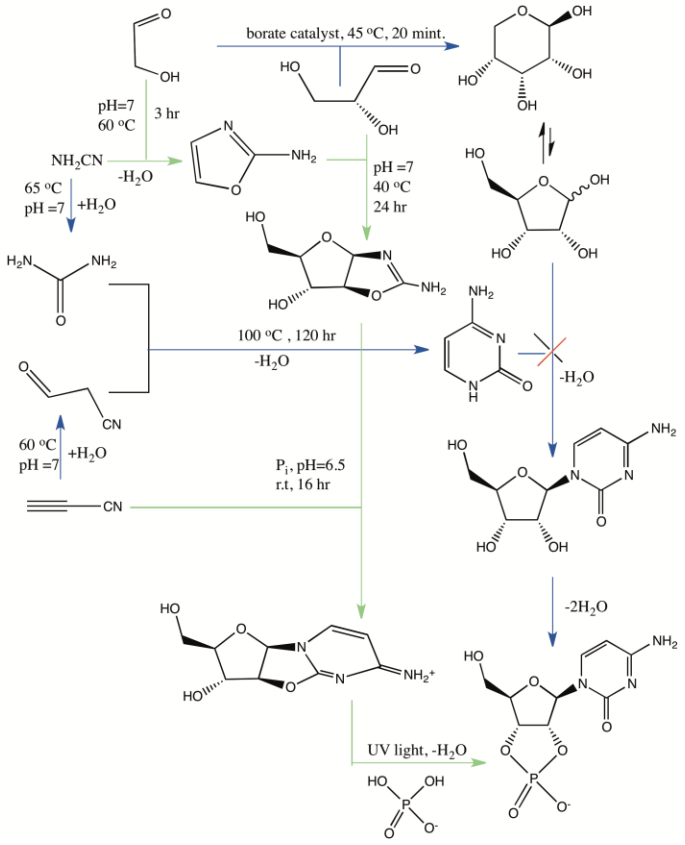


Figure 1.6 Mechanistic route to the formation of pyrimidine ribonucleotide. Previously proposed synthetic pathway of β-ribocytidine-2`,3`-cyclic phosphate (shown by blue arrow) and the successful new synthesis described by Sutherland *et al.* (green arrow) in 2009.

In 2012 Sutherland *et al.*[111] tried to synthesize two (glycolaldehyde) or three carbon (glyceraldehyde) sugars, which are the ribonucleotide precursors starting from even very small molecules, which were proposed to be present on the early earth. They synthesized glycolaldehyde started from a single carbon source molecule HCN in water through the photoredox system chemistry in presence of copper cyanide complex as catalyst (shown in figure 1.7). Previously, sugars had been synthesized by the formose reaction[112]– where formaldehyde was homologated in a basic medium. But this homologation is not fully accepted because it needs umpolung in the first step. There is an alternative homologation process known as the Killiani-Fischer synthesis[113-114] where formaldehyde reacts with HCN to give cyanohydrins glycolonitrile. But there was a problem in the next step where catalytic hydrogenation occured in glycolonitrile, which needs the poisoned catalyst to stop the further unwanted reduction. Also, their starting material is formaldehyde, which was not supposed to be present in the early earth. But Sutherland *et al.* started the reaction only with HCN, which is acceptable in prebiotic chemistry conditions.

The key findings of Sutherland were that they started the reaction only with the sole carbon source molecule HCN and in the intermediate step they got formaldehyde by selective reduction of HCN. In their sugar synthesis reaction sequence, they had used a copper cyanide complex as a catalyst in the photoredox condition to selectively reduce HCN to form formaldimine and the key intermediate glycolonitrile. Here, they have also followed the "systems chemistry" as characterized by Szostak.[109] There were several experimental studies[115-120] towards the formation of glycoaldehyde and other simple sugar molecules.

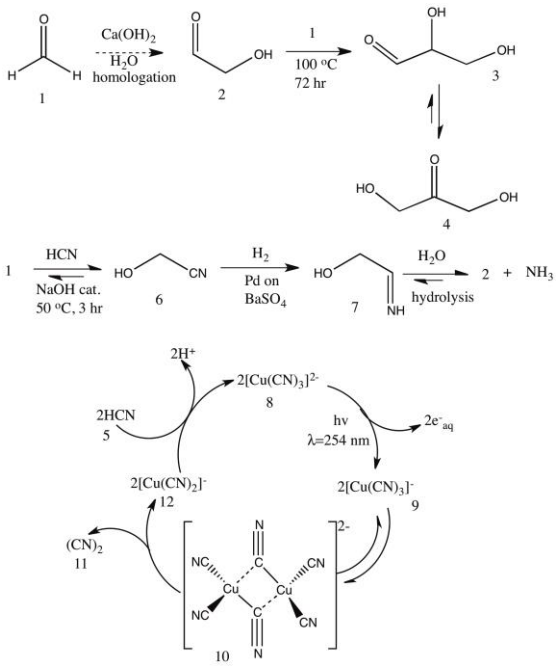


Figure 1.7 Formose and Killiani-Fischer synthesis of sugar and photoredox metal catalysed cycle proposed by Sutherland *et al.* in 2012.

In the year 2013, Sutherland and coworkers[115] showed that the effective production of aldehyde antecedents to the building units of RNA and proteins by irradiation with UV light on a system containing copper (I) cyanide to an aqueous solution of glycolonitrile, hydrogen sulfide and sodium phosphate. In 2015 they again came up with novel work where they showed that all the precursors of RNA, proteins, and lipids[116] were formed from the cyanosulfidic protometabolism process. They proposed that in the presence of a hydrogen sulfide catalyst, UV radiation and the photoredox cycle accelerated by Cu(I)-Cu(II) reductive homologation of HCN leads to the formation of proteins and lipids precursors. The Sutherland group stated that the

condition for the formation of ribonucleotides precursors also formed the starting materials of proteins and lipids. They also said that the different sets of building blocks are separate from each other and required different metal catalysts for their synthesis.

So far, we have discussed the synthetic routes of different life building block precursors and their experimentally obtained mechanistic pathways. This is due to the lack of innovative computational tools where the theoretical methods not only support the experimental findings but also lead to the discovery of new chemistry. There have been very few theoretical and computational studies in the field of origin of life in the last couple of decades. In 2010, Goldman *et al.* synthesized glycine containing compounds[20] using the interstellar molecular composition (NH_3, H_2O, CO, CO_2, CH_3OH) by shock wave molecular dynamics simulation. This was a computational follow-up of work done by Greenberg *et al.*,[121] where they had synthesized different amino acids by using the interstellar ice composition and irradiating with UV radiation without using any metal catalyst. In 2014, Saija and Saitta showed how formamide molecules play an important role towards the formation of amino acids by using the reactive molecular dynamics method.[22] In the same year, a completely new computational approach was developed by Martin*ez and co-workers*[21] for discovering new molecules and the reaction mechanism without employing a predefined reaction coordinate. Using the *ab initio* nanoreactor (AINR) approach they found completely new pathways of formation of glycine starting from very small molecule compositions, which had previously been experimentally studied by Urey and Miller. These results provide not only new insights into the previously done experiments, but also highlight the necessity of computational chemistry as a tool for discovering and finding new mechanisms and pathways that had not been explored yet by any experiments.

Therefore, from the above discussed experimental and theoretical work in the field of prebiotic chemistry, it seems to be a big challenge for the researchers to give a bold statement about how life had begun due to chemical evolution that occurred 4.6 billion years ago. Lots of questions are still unanswered. Researchers are still trying to find solutions to the unsolved problems, and maybe at some point in the future, they will be able to finally find all the answers to all the questions regarding the evolution of chemistry on the early earth.

1.7 Statement of Problem:

When we are entering into the field of the origin of life or the origin of the universe, a lot of questions and thoughts come to our mind. Described below are the chief objectives for further progressing into the field of such "messy" chemistry.

1. Determining the requirements for the origin of life, such as the carbon source, the source of energy, and the molecular composition in the early earth.

2. Determining the stepwise mechanistic pathways that lead to chemical evolution of the origin of life.

3. Determining the formation of several organic molecules such as amino acids, simple sugars, nucleotides, as well as the building blocks of living cells.

4. Determining the formation of proteins, RNA, and the polymeric form of organic molecules that can act as catalysts to carry out metabolic reactions.

5. Determining the formation of the very simple and small molecules during the origin of the universe.

1.8 Objectives of the thesis

Though the experimental studies have opened some windows towards the mystery of the origin of life as well as the universe based on the existing theories and hypotheses, still a lot of unsolved questions exist, such as (i) What was the starting point for the chemical evolution towards the origin of life? (ii) Was the origin of life a local event or a global phenomenon? (iii) What are the plausible mechanistic pathways that lead towards the formation of RNA, proteins, nucleic acid bases *via* elementary steps in the presence of a lot of constraints present in the prebiotic era? (iv) Is there any role of metal ions acting as catalysts towards the formation of different life building units such as RNA, proteins and nucleic acid bases? (v) What is the energy source to cross the reaction barrier for each elementary reaction step; are the processes thermally or photochemically driven? (vi) According to the theory and hypothesis, the early earth atmosphere was reducing in nature - therefore it might be interesting to know what would be the effect of an oxidizing atmosphere towards the formation of the important life building blocks

such as RNA and proteins. (vii) How did the early universe's small molecules come into the picture after the Big Bang? In this thesis, full quantum chemical calculations using *ab initio* molecular dynamics (AIMD), coupled-cluster (CC), second-order Moller-Plesset perturbation theory (MP2) as well as density functional theory (DFT) have been employed to provide some interesting insights and address the questions mentioned above. We have investigated the origin of life on the prebiotic earth and also looked at the small molecules that would have existed at the beginning of the universe. The thesis titled "**Insights into Messy Chemistry Related to Cosmology and Origin of Life Obtained by Employing State-of-the-art Computational Methods**" is divided into seven different chapters. A brief introduction to each chapter is provided below with the chapter titles.

1.9 Organization of Thesis

Chapter-1: A Brief Introduction about the Approach of Messy Chemistry Towards Cosmology and Prebiotic Chemistry

In this chapter, a brief introduction regarding messy chemistry, which is the complex chemistry that happened on prebiotic earth or interstellar space during the origin of life. Also discussed is the formation of small molecules at the beginning of the universe. The effect of oxidizing and reducing atmospheres towards the formation of life building units on the prebiotic earth as well as interstellar space has also been discussed.

Chapter-2: Fundamentals of Computational Chemistry

This chapter deals with the fundamentals of quantum mechanical methods (QM), density functional theory (DFT), *ab initio* molecular dynamics (AIMD) as well as *ab initio* nanoreactor dynamics (AINR), and volume correction methods.

Chapter-3: Insights into Chemical Reactions at the Beginning of the Universe: From HeH^+ to H_3^+

In this chapter, we have tried to computationally mimic the conditions in the early universe to show how the recombination process would have led to the formation of the first ever formed diatomic species of the universe: HeH^+, as well as the subsequent processes that would have led

to the formation of the simplest triatomic species: H_3^+. We have also studied some special cases: higher positive charge with fewer numbers of hydrogen atoms in a dense atmosphere, and the formation of unusual and interesting linear, dicationic He chains beginning from light elements He and H in a positively charged atmosphere. For all the simulations, the *ab initio* nanoreactor (AINR) dynamics method has been employed.

Chapter-4: Insights Into the Origin of Life: Did life begin from HCN and H_2O? - An *ab initio* Nanoreactor Dynamics Approach

In this chapter, we have conducted full quantum mechanical molecular dynamics (MD) simulations employing the AINR approach on systems containing a mixture of molecules of HCN and H_2O. The goal has been to follow the chemical reactions that can occur through collisions between the molecules and observe what new species are formed as a result. In short, our objective has been to perform the equivalent of an experimental study while satisfying the conditions present on the prebiotic earth. Remarkably, we have found, that just the interaction of HCN and H_2O was sufficient to eventually lead to the formation of the experimentally reported precursor molecules to RNA and proteins: cyanamide, glycolaldehyde, an oxazole derivative, and the amino acid glycine

Chapter 5: From Messy Chemistry to the Origin of Life: Insights from complete Quantum Mechanical (QM) Perspective

In this chapter, analysis of the data is presented, which allowed us to determine the mechanistic pathways by which HCN and H_2O reacted together to yield intermediates and, eventually, the RNA and protein precursors. We subsequently subjected these pathways to a full static quantum chemical study with density functional theory (DFT) and thus obtained all the barriers ($\Delta G^{\#}$) for the reactions involved in these processes, as well as the energies (ΔG) of the reactions. As will be discussed in this chapter this has led to results that not only reveal interesting pathways for the formation of the precursor molecules beginning from aqueous HCN but also indicate that these mechanistic routes would have been thermodynamically and kinetically feasible.

Chapter 6: The Effect of Oxidizing Atmosphere on the Origin of Life in Prebiotic Earth and Interstellar Space

In this chapter, we have shown how an oxidizing atmosphere inhibits the formation of sugars and amino acids, which are the precursors to RNA and proteins in the prebiotic earth, as well as in interstellar space. We have done the *ab initio* nanoreactor dynamics simulations by taking the Urey-Miller gaseous composition as the prebiotic earth atmosphere which contains NH_3, CO, CH_4, H_2O, H_2 species. In this, we have further added CO_2 and O_2 molecules, both at a time as well as individually, and observed the effect towards the formation of precursors to the life building blocks. Furthermore, we have performed similar reactive dynamics by taking the bare interstellar ice composition (CH_3OH, NH_3, CO, H_2O) and similarly added the CO_2 and O_2 molecules and checked the outcomes of taking such mixtures from the AINR dynamics. It has been observed that the effect of O_2 is more compared to CO_2 towards the inhibition of sugar and amino acids formation. We have also discussed the pathways for glycoaldehyde, and glycine formation from the AINR simulations.

Chapter 7: Summary and Future Aspects

In this chapter, the conclusion and future aspects of the thesis work have been provided.

1.10 References:

1. Meyer, B. S. *ACS Symp. Series.* **2008**, *981*, 39
2. Gusten, R.; Wiesemeyer, H.; Neufeld, D.; Menten, K. M.; Garf, U. U.; Jacobs, K.; Klein, B.; Ricken, O.; Risacher, C.; Stutzki, J. *Nature* **2019**, *568*, 357.
3. Hogness, T. R.; Lunn, E. G. *Phys. Rev*. **1925**, *26*, 44.
4. Ganti, T.́ *Oxford Univsersity Press*: Oxford UK, **2003**.
5. Dyson, F. *Cambridge University Press:* Cambridge UK, **1999**.
6. Sutherland, D. *Nat. Rev*. **2017**, *1*, 1
7. Steel, M.; Penny, D. *Nature* **2010**, *465*, 168.
8. Oparin, A. I. *World Publishing: Cleveland*, **2003**

9. Bernal, J. D. *World Publishing*: *Cleveland*, **1967**.
10. Sutherland, J. D. *Angew. Chem., Int. Ed*. **2016**, *55*, 104.
11. Miller, S. L., Urey, H. C. *Science*, **1959**, *130*, 245.
12. Orgel, L. E. *Science*, **1966**, *154*, 784.
13. Oro´, J. *Biochem. Biophys. Res. Commun*., **1960**, *2*, 407.
14. Miller S. L. *Nature*, **1995**, *375*, 772.
15. Ferris, J. P., Sanchez, R. A., Orgel, L. E. *J. molec. Biol.,* **1968**, *33*, 693.
16. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.*, **2004**, *39*, 99.
17. Powner, M.W., Gerland, B., Sutherland, J. D. *Nature*, **2009,** *459*, 239.
18. Ritson D, Sutherland J. D. *Nature Chem.*, **2012**, *4*, 895.
19. B. H. Patel, C. P., D. J. Ritson, C. D. Duffy, and J. D. Sutherland , *Nat. Chem*. **2015**, *7*, 301.
20. Goldman, N., Reed, E. J., Fried, L. E., Kuo, I. F. W., Maiti, A. *Nat. Chem*. **2010**, *2*, 949.
21. Wang, L. P., Titov, A., Liu, F., Martinez, T. J. *Nat. chem*., **2014**, *6*, 1044.
22. Saitta, A. M.; Saija, F. *Proc. Natl. Acad. Sci.* U. S. A. **2014**, *111*, 13768.
23. Ayyappan A., Nandi S. , Bhattacharyya D. *Chemistry - A European* Journal, **2018**, *24*, 4885**.**
24. Introduction:Cosmology-space *New Scientist.* 4 September **2006**
25. Hetherington, Norriss S. (**2014**). Routledge. p. 116. 978-1-317-67766
26. Luminet, Jean-Pierre (**2008**). CRC Press. p. 170. ISBN 978-1-4398-6496-8.
27. Cosmology Oxford Dictionaries
28. Overbye, Dennis *(25 February* **2019***). The New York Times.* Retrieved 26 February 2019.
29. G. Bruylants, K. Bartik, J. Reisse ; **2011**, *14*, 388.
30. G. Bruylants, K. Bartik, J. Reisse *Orig. Life Evol. Biosph*, **2010**, *40*, 137
31. M. Dresden, H.A. Kramers (Eds.), **1987**, *Springer-Verlag*, *NewYork*
32. Saslaw WC, Zipoy D.. *Nature*, **1967**, *216*, 976
33. Abel T, Anninos P, Zhang Y, Norman, ML. 1997. New Astron. 2:181
34. Bates D. R. *MNRAS*, **1951,** *111*, 303
35. McDowell M. R. C. *Observatory*, **1961,** *81*,240
36. Chandrasekhar S.. *Ap. J*. **1958**,*128*,114
37. Schmeltekopf A.L, Fehsenfeld F.F, Ferguson E.E.. *Ap. J.* **1967**, *148*,155.

38. Palla F, Salpeter E.E, Stahler S.W. *Ap. J*. **1983**, *271*, 632
39. Lepp S, Shull J. M. *Ap. J*., **1984**, *270*, 578
40. Black, J. H. *Molecular Astrophysics*, (Cambridge: University Press), **1990**, 473
41. Puy D, Alecian G, Le Bourlot J, Leorat J, Pineau Des Forˆets G.. *Astron Astrophys*. **1993**, *267*, 337.
42. Stancil P. C, Lepp S, Dalgarno A., *Ap. J*. **1996**, *458*,401
43. Stancil P.C, Lepp S, Dalgarno A.. *Ap. J*. **1998**, *509*,1
44. Galli D, Palla F.. *Astron. Astrophys*. **1998**, *335*, 403
45. Galli D, Palla F., eds. A Weiss et al. (Berlin: Springer), **2000**, 229
46. Galli D, Palla F. *Planet. Sp. Sci*. **2002,** *50*,1197
47. Puy D, Signore M.. *New Astron. Rev*. **2002**, *46*, 709
48. Puy D. Signore M.. *New Astron*. **2007**, *51*, 411
49. Lepp S, Stancil PC, Dalgarno A. *J. Phys. B*, **2002**, *35*, 57
50. Glover S. C. O, Abel T. *MNRAS*, **2008**, *388*, 1627
51. Schleicher D. R. G, Galli D., Palla F., Camenzind M., Klessen R.S.,. *Astron. Astrophys*. **2008**, *490*, 521
52. Glover S.C.O, Savin D.W. MNRAS, **2009**, *393*,911.
53. Vonlanthen P, Rauscher T, Winteler C, Puy D, Signore M, Dubrovich V.. *Astron. Astrophys*., **2009**, *503*,47.
54. Gay C.D, Stancil P.C, Lepp S., Dalgarno A., *Ap. J*., **2011,** *737*,44.
55. Zinchenko I., Dubrovich V., Henkel C., *MNRAS*, **2011**, *415*,78.
56. Moorhead J.M., Lowe R.P., Wehlau W.H., Maillard J.P., Bernath P.F. *Ap. J*., **1988**, *326*, 899.
57. Liu X.W., Barlow M.J., Dalgarno A., Tennyson J., Lim T., *MNRAS* , **1997**, *290*,71.
58. Dinerstein H.L., Geballe T.R. *Ap. J*., **2001**, *562*,515
59. Saha S., Datta K.K., Barua A.K., *J. Phys. B*, **1978**, *11*, 3349
60. Flower D.R., Roueff E., *Astron. Astrophys*., **1979**, *72*, 361.
61. Roberge W, Dalgarno A., *Ap. J*., **1982**, *255*,489
62. Basu D., Barua A.K., *J. Phys. B*, **1984**, *17*, 1537
63. Kimura M., Lane N.F., Dalgarno A., Dixson R.G., *Ap. J*. **1993,** *405*, 801.
64. Dumitriu I, Saenz A. J. *Phys. Conf. Ser*., **2009**, *194*,152026

65. Sodoga K., Loreau J., Lauvergnat D., Justum Y., Vaeck N., *Phys. Rev. A,* **2009**, *80*, 033417.
66. Pedersen H.B., Altevogt S., Jordon-Thaden B., Heber O., Lammich L., *Phys. Rev. Lett*., **2007**, *98*, 223202.
67. Juˇrek M., Spirko V., Kraemer W.P., *Chem. Phys*., **1995**, *193*, 287
68. Zygelman B, Stancil PC, Dalgarno, A., *Ap. J*., **1998**, *508*,151
69. Stancil P.C., Babb J. F., Dalgarno A., *Ap. J*., **1993**, *414*, 672.
70. Linder F., Janev R. K., Botero J., **1995.**
71. Bovino S., Tacconi M., Gianturco F.A., Galli D., *Astron. Astrophys*., **2011**, *529*, 140
72. Guberman S.L., *Phys. Rev. A*., **1994**, *49*, 4277.
73. I. Drabowski, G. Herzberg *Ann. N.Y Acad. Sci*. **1977**, *38*, 14.
74. Black, J. H. *Astrophys. J*., **1978**, *222*, 125.
75. Flower, D. R., Roueff, E. *Astron. Astrophys*., **1979**, *72*, 361.
76. Roberge, W., Dalgarno, A., *Astrophys. J*., **1982**, *255*, 489.
77. Heyminck, S., Graf, U. U., Güsten, R., Stutzki, J., Hübers, H. W., Hartogh, P., *Astron. Astrophys*., **2012**, 542,1
78. Risacher, C. *IEEE Transactions on Terahertz Science and Technology*, **2016**, *6*, 199
79. Young, E. T., Becklin, E.E., Marcum, P.M., *Astrophys. J.*, **2012**, 749,17
80. Becke, A; Saurabh, S; Kalosi, A; Paul, D; Wilhelm, P; Novotny, O; *Science*, **2019**, *365*, 676.
81. Gerlich D, Horning S. *Chem. Rev.*, **1992**, *92*, 1509.
82. Schleicher D.R.G., Galli D., Palla F., Camenzind M., Klessen R.S.,. *Astron. Astrophys*., **2008**, *490*, 521
83. Glover S.C.O., Savin D.W., *MNRAS*, **2009**, *393*,911
84. Stancil P.C., Lepp S., Dalgarno A., *Ap. J*., **1998**, *509*,1
85. Woese, C.,*179* (Harper & Row, **1967**).
86. Crick, F. H. C., *J. Mol. Biol.*, **1968**, *38*, 367–379.
87. Orgel, L. E., *J. Mol. Biol*., **1968**, *38*, 381–393.
88. Joyce, G. F., *Nature*, **2002**, *418*, 214–221.
89. Joyce, G. F. & Orgel, L. E. (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 23–56 (Cold Spring Harbor Laboratory Press, **2006**).

90. Hartman, H. J. *Mol. Evol*, **1975**, *4*, 359–370.

91. Miller, S. L., *Science*, **1953**, *117*, 528−529.

92. Strecker, A. *Liebigs Ann. Chem*. **1854**, *91*, 349− 351.

93. Bar-Nun, A.; Bar-Nun, N.; Bauer, S. H.; Sagan, C. *Science*, **1970**, *168*, 470−473.

94. Matthews, C. N.; Moser, R. E. *Nature*, **1967**, *215*, 1230−1234.

95. B.Basile, A. Lazcano, J. Oro, *Adv. In Space Research*, **1965**, *4*, 125-131.

96. R. A. Sanchez, J. P. Ferris, and L. E. Orgel, *Mol. Biol*., **1968**, *38*, 121-125.

97. N. Friedman and S. L. Miller, *Nature*,**1969**, *221*, 152-156.

98. Strecker, A. *Liebigs Ann. Chem*.,**1854**, *91*, 349–351.

99. Ferris, J. P. & Hagan, W. J. Jr., *Tetrahedron*,**1984**, *40*, 1093–1120.

100. Eschenmoser, A. & Loewenthal, E. *Chem. Soc. Rev*., **1992**, *21*, 1–16.

101. Donn, B., *J. Mol. Evol.*, **1982**, *18*, 157–160.

102. Tokunaga, A. T., Beck, S. C., Geballe, T. R., Lacey, J. H. & Serabyn, E. *Icarcus*, **1981**, *48*, 283–289.

103. Hanel R. *et al*. *Science*, **1981**, 212, 192–200.

104. Snyder, L. E. & Buhl, D. *Astrophys. J. Lett.*, **1971**, *163*, L47–L52.

105. Miller S. L *Natur*, **1995**, *375*, 772–774.

106. Ferris, J. P., Sanchez, R. A. & Orgel, L. E. *J. molec. Biol.*, **1968**, *33*, 693–704.

107. The RNA World (eds Gesteland, R. F. & Atkins, J. F.) (*Cold Spring Harbor Lab.* Press, Massachusetts, **1993**).

108. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.*, **2004**, *39*, 99–123.

109. Szostak, J. W., *Nature*, **2009**, *459*, 171–172.

110. Powner, M.W., Gerland, B. & Sutherland, J. D., *Nature*, **2009**, 459, 239–242.

111. Ritson D., Sutherland J.D., *Nature Chem*.; **2012**, 4; 895–899.

112. Socha, R. F., Weiss, A. H. & Sakharov, M. M., *J. Catal*., **1981**, *67*, 207–217.

113. Fischer, E. *Ber. Dtsch Chem. Ges*., **1889**, *22*, 2204–2205.

114. Morrison, J. D. & Mosher, H. S. 133–141 (*Prentice-Hall*, **1971**).

115. Ritson, D. J.; Sutherland, J. D. *Angew. Chem., Int. Ed.*, **2013**, *52*, 5845−5847.

116. Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D., *Nat. Chem*., **2015**, *7*, 301−307.

117. Ritson, D. J.; Battilocchio, C.; Ley, S. V.; Sutherland, J. D. *Nat. Commun.* **2018**, *9*, 1821−1830.

118. Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. *Science,* **2004**, *303*, 196-200.

119. Fischer, E.¨*Ber. Dtsch. Chem. Ges.*,**1889**, *22*, 2204−2205.

120. Benner, S. A.; Kim, H. J.; Carrigan, M. A., *Acc. Chem. Res*. **2012**, *45*, 2025−2034.

121. G. M. Mun˜ oz Caro, U. J. Meierhenrich, W. A. Schutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W. H.-P. Thiemann, A. Brack & J. M. Greenberg. *Nature.*, **2002**, *416*, 403-406.

Chapter 2

Fundamentals of Computational Methods

Chapter 2

Fundamentals of Computational Methods

Abstract

The reactive *ab initio* molecular dynamics (AIMD), specifically the state-of-the-art *ab initio* nanoreactor dynamics (AINR), has been used as a tool to discover new and unknown chemistry that would have happened in the prebiotic era, as well as investigate chemical reactions of interest in cosmology. Furthermore, a full quantum mechanical (QM) and density functional theory (DFT) approach has been employed to investigate mechanistic pathways of a series of reactions that would have happened during the "messy chemistry" of the prebiotic age. In this chapter, the fundamentals of these computational methods: AIMD, QM and DFT have been discussed in brief. The full quantum mechanics and DFT methods have allowed us to compute the minimum energy pathways of systems as a function of energy and electron density. AINR has become a powerful tool to study and discover new chemistry, and DFT and full QM methods deal with reactivity, structures, transition states, spectra and other properties of chemical systems with accuracy. After establishing the favorable reaction mechanism (lowest energy pathways), volume correction in the entropy have been carried out in this thesis work. In this chapter, the theoretical background of these methods has been described in brief.

2.1 Introduction

Solving chemical problems on a computer using existing algorithms rather than developing new algorithms or theoretical methods is generally known as computational chemistry. Douglas Hartree, father of the Hartree-Fock theory, was the first scientist who used computers in 1946 for theoretical calculations[1] on chemical systems. Computational methods can be employed today for the geometry optimizations of molecules, the energy determination of molecules, transition state searches, IR, Raman, NMR, and UV-Visible spectrum analysis, bond and orbitals and bond analysis, the determination of polarizability, ionization potential, electron affinities, charge and population analysis, and to study the effect of dispersion interactions. They can also lead to the discovery of new chemistry that has not been explored yet by experimental methods. Elementary step and multistep mechanistic pathways for various reactions can be represented as energy profile diagrams from the analysis of potential energy surfaces.

Computational chemistry consists of the following methods for studying diverse chemical problems:

i) Molecular mechanics or Newtonian mechanics, which does not use a wave function, and is based on force field calculations. Therefore, in order to see the electronic effects in the chemical system, classical mechanics or molecular mechanics methods cannot be applied. As this method does not deal with the number of electrons present in the system, it is faster than quantum chemical methods.

ii) *Ab initio* quantum chemistry methods are based on the Schrödinger equation. *Ab initio* methods means first principles or from the beginning. They include Hartree Fock, post Hartree Fock, and Quantum Monte Carlo methods.

iii) Semi-empirical methods, which use the modified Hartree-Fock equation with a simpler Hamiltonian rather than the exact molecular Hamiltonian, by introducing functions with empirical parameters. The method is highly ambitious, and is exclusively for larger systems. Hence, these methods are inexpensive methods due to the parameterization of the two-electron integrals, making the computations faster.

iv) Density functional theory (DFT), provides an alternative method to solve the many-electron

problems by using the probability of the electron density. In DFT, the density of the electron, which can be represented with three coordinates, is the basis of determining the energy and other ground state properties of the system.

v) The *ab initio* molecular dynamics (AIMD) method consists of Newton's equation of motion along with the Schrödinger equation, where the force calculated from Newton's equation and the energy is determined from Schrödinger equation. Nowadays, the reactive dynamics tools are highly relevant, where the bond breaking and bond formation can also be studied, as a special part of AIMD.

Apart from these computational approaches, there are highly computationally expensive a*b initio* quantum mechanical methods known as post Hartree-Fock methods such as Moller-Plasset perturbation theory (MP2),[2] configuration interaction (CI)[3] and coupled-cluster (CC)[4]. In this thesis, quantum mechanical calculations, especially with state-of-the-art reactive *ab initio* molecular dynamics (AIMD), CC, DFT have been performed, as these are known as the best theoretical methods to study chemical systems accurately.[5] In Section 2.3 and Section 2.5 of this chapter, the fundamentals of DFT and AIMD have respectively been provided. Furthermore, elementary quantum mechanics has been considered in the next section of this chapter to understand the basis of AIMD and DFT.

2.2 Elementary Quantum Mechanics

2.2.1 The wave function, Hamiltonian Operator and the Schrödinger Equation

The basic goal of advance quantum mechanics is to solve the time-dependent non-relativistic Schrödinger equation, which was first put forward in 1926 by the Austrian-Irish physicist Erwin Rudolf Josef Alexander Schrödinger. Quantum mechanics depends on solving the Schrödinger equation, which allows us to calculate the wave function at any time. However, the goal of quantum mechanical approaches towards any chemical system is to deal with the time-independent Schrödinger equation (TISE), because, in most of the cases, time-dependent interactions are not compellingly relevant in chemical problems. For any chemical system with n number of electrons and m number of nuclei, the many-body TISE is given by:

$$\hat{H}\Psi\,(x_1, x_2, x_3 \ldots\ldots x_n, r_1, r_2 \ldots r_m) = E\Psi\,(x_1, x_2, x_3 \ldots\ldots x_n, r_1, r_2 \ldots r_m) \qquad (2.11)$$

x is the variable used for electronic co-ordinates, and r for the nuclei. This is an eigenvalue equation where, the eigenvalue E represents the total energy of the system, eigenfunction $\Psi(x_n, r_m)$ is the wave function for the system, which is function of the 3n spatial coordinates and n spin coordinates of electrons and 3m spatial coordinates of nuclei.

Furthermore, a wave function contains all the possible information regarding the system. Therefore, it is also called the state function of the system. By applying a suitable operator on the wave function, all the information regarding the system can be obtained. The wave function ($\Psi(x_n,r_m)$) has no such physical meaning but the square of it, called the probability density ($|\Psi(x_1, x_2, x_3....x_n)|^2$) provides the basis of the physical interpretation of the wave function. The wave function is a function of both electrons and nuclei. After separating the electronic and nuclear wavefunction one can arrived at probability density.

Equation 2.12 represents the probability of finding the electrons within a given volume element

$$\int|\Psi(x_1, x_2, x_3....x_n)|^2 dx_1.dx_2.....dx_n \qquad (2.12)$$

Equation 2.12 represents the probability that electrons 1, 2, ….., n are found simultaneously in volume $dx_1.dx_2…….dx_n$.

The Hamiltonian operator $\hat{H}(x_n, r_m)$, can be represented as;

$$\hat{H} = -\frac{\hbar^2}{2}\sum_{a=1}^{m}\frac{1}{m_a}\nabla_a^2 - \frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2 + \sum_a\sum_{a>b}\frac{z_a z_b e^2}{r_{ab}} - \sum_a\sum_i\frac{z_a e^2}{r_{ia}} + \sum_i\sum_{i>j}\frac{e^2}{r_{ij}}$$

(2.13)

In equation (2.13), The first two terms in the Hamiltonian operator represent the kinetic energy of electrons and the nuclei respectively. Nuclear-electron attraction, nuclear-nuclear repulsion and electron-electron repulsion respectively are represented by the last three terms. Indices i and j indicate a total number of n electrons, whereas indices a and b denote a total of the m nuclei of the system. Other associated terms in the equation have their usual meaning. $\hbar=h/2\pi$ and h is the Planck's constant, m_e and m_a are the masses of the electrons and nuclei respectively, Z_a and Z_b represent the charges of nuclei, r_{ab} is the distance between the nuclei a and b, r_{ia} represents the

distance between the i^{th} electron and the a^{th} nuclei and r_{ij} is the distance between the i^{th} and the j^{th} electron.

For addressing real life chemical problems by employing QM theory, we need to solve the SE for multi-nuclear multi-electron molecular systems. However, the exact solution to the SE is finite and it is limited to simple systems only, such as a particle in a box, the harmonic oscillator, the hydrogen atom and the rigid rotor. Hence, to make this theory applicable to larger systems, approximate methods have been proposed over the years. In quantum mechanics, the adiabatic approximation refers to those solutions to the Schrödinger equation that make use of a time-scale separation between fast and slow degrees of freedom, and use this to find approximate solutions as product states in the fast and slow degrees of freedom. For instance, in the study of vibrational dynamics when the bond vibrations of molecules occur much faster than the intermolecular motions of a liquid or solid. It is also generally implicit in a separation of the Hamiltonian into a system and a bath, a method we will often use to solve condensed matter problems. As widely used as the adiabatic approximation is, there are times when it breaks down, and it is important to understand when this approximation is valid, and the consequences of when it is not. This will be particularly important for describing time-dependent quantum mechanical processes involving transitions between potential energy sources. Perhaps the most fundamental and commonly used approximation that comes into the consideration is the Born-Oppenheimer approximation, which is believed to be a good approximation for stationary point calculations.

2.2.2 The Born-Oppenheimer Approximation

In 1927 Börn and Oppenheimer (BO) proposed an approximation for simplifying the Schrödinger equation (S.E.). According to the BO approximation, the nuclei move much slower than electrons, as nuclei are much heavier than electrons. Therefore, we can presume that all electrons are present in the field of fixed nuclei, *i.e.*, the nuclear kinetic energy is zero, and the potential energy of nucleus is simply a constant. Therefore, the molecular Hamiltonian operator represented in equation 2.14 adequately reduces to the electronic Hamiltonian.

$$\hat{H}_{el} = -\frac{\hbar^2}{2}\sum_{a=1}^{m}\frac{1}{m_a}\nabla_a^2 - \frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2 - \sum_a\sum_i\frac{z_a e^2}{r_{ia}} + \sum_i\sum_{i>j}\frac{e^2}{r_{ij}} \qquad (2.14)$$

Hence, the Schrödinger equation with the electronic Hamiltonian would be:

$$\hat{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \qquad (2.15)$$

Now, the total energy of the system can be written as:

$$E_{tot} = E_{el} + E_{nn} \qquad (2.16)$$

where E_{el} is the electronic energy and E_{nn} is the nuclear-nuclear repulsion, which is given by:

$$E_{nn} = \sum_a \sum_{a>b} \frac{z_a z_b e^2}{r_{ab}} \qquad (2.17)$$

2.2.3 The Variational Principle

First, we need to design a Hamiltonian operator of the molecule for which the S.E needs to be solved. $\Psi(x_n, r_m)$ and E are the unknown quantities in Equation 2.11.To solve the S.E., we need to find the eigenfunctions and eigenvalues for the Hamiltonian operator. However, there is no such technique to solve the S.E accurately for any system. Therefore, to get rid of this problem we need a trial wave function. The variational principle gives us the opportunity to solve this problem. In a system with wave function Ψ, the average value of the energy for that precise system can be obtained by

$$E[\Psi] = \frac{<\Psi|\hat{H}|\Psi>}{<\Psi|\Psi>} \geq E_0 = \frac{<\Psi_0|\hat{H}|\Psi_0>}{<\Psi_0|\Psi_0>} \qquad (2.18)$$

where $$<\Psi|\hat{H}|\Psi> = \int \Psi^* \hat{H} \Psi d\tau \qquad (2.19)$$

The variational principle states that *"the energy calculated using a trial wave function is always an upper bound to the original ground state energy (E_0) of the system of interest."*

Here, Ψ_0 is the true ground state wave function. The trial wave functions must follow certain criteria that would ensure that these functions have a physical significance. For instance, to be well behaved, the wave function, Ψ must be continuous and be square integrable. This can be represented as,

$$E_0 = min_{\Psi \to n}\, E[\Psi] = \; min_{\Psi \to n} < \Psi \left| \hat{T} + \; \hat{V}_{ne} + \; \hat{V}_{ee} \right| \Psi > \qquad (2.20)$$

2.3 Density Functional Theory

Wave function (Ψ) based quantum mechanical strategies are computationally more costly, as the wave function is a 4n variable: 3n for space variables and n for spin variables for the electrons. Additionally, the multifaceted nature of the wave function increases with the increase in the number of electrons in the framework. To decrease this intricacy in the calculation of fundamental properties of any framework, it is important to consider an option for a wave function. Density functional theory (DFT) has distinct advantages in comparison to other approaches of wave function based strategies, with regard to efficiency and applicability. By utilizing the electron density as the main entity of interest, DFT allows the determination of energy and different properties of the system. Dissimilar to the wave function, the electron density is an observable and quantifiable amount, and relies upon three spatial directions. Hence, electron density based strategies are more effective than wave function based techniques. One can apply DFT for figuring out different types of properties for different systems, including significantly larger systems than handled by wave function methods. In the accompanying subsections, the DFT approach is discussed further.

Although the journey towards modern DFT began just a few decades ago, the first attempt to use electron density to obtain information about atomic and molecular systems has been dated back to the early days of quantum chemistry, just shortly after the introduction of the Schrödinger equation (1926). The first approximation of such a kind was proposed by Llewellyn Thomas (1927) and Enrico Fermi (1927). They introduced electron density in place of the wave function as a means to understanding the electronic structure of many-body systems. This model is known as the Thomas– Fermi (TF) model. In this quantum statistical model, Thomas and Fermi used the concept of the uniform electron gas with a constant electron density. Electron gas can defined as a population of free electrons in a vaccum or in a metallic conductor. They further derived the expression of the kinetic energy of a quantum mechanical system. This model possesses some limitations. It gives a very rough estimate of the actual kinetic energy of the system, as the electrons in this model have been considered to be part of a gas of a constant electron density. Furthermore, the exchange and correlation effects are totally ignored.

2.3.1 The Electron Density

The electron density is an essential segment of the density functional hypothesis. It is characterized as the likelihood of discovering one electron inside a specific volume. The electron density is shown by the following integral.

$$\rho(\overrightarrow{r}) = N \int \dots \int |\Psi(\overrightarrow{x_1} \dots \dots . \overrightarrow{x_n})|^2 ds_1 d\overrightarrow{x_2} \dots \dots . d\overrightarrow{x_n} \qquad (2.21)$$

where, $\vec{x} = \vec{r} \, . s$ and $\rho(\vec{r})$ is the likelihood of finding any of the N electrons in volume dr_1, with an arbitrary spin value. The leftover N-1 electrons will have arbitrary positions and spins in the state represented by Ψ. Moreover, $\rho(\vec{r})$ (the electron density), can be estimated by test methods; for example, X-ray diffraction, scanning tunneling microscope and electron diffraction. Additionally, $\rho(\vec{r})$ is a non-negative function of three spatial directions that integrates to the absolute number of electrons, i.e., $\rho(\vec{r}) \geq 0$ and $\int \rho\,(\vec{r}) d(\vec{r}_1)$ = N where $\rho(\vec{r} \longrightarrow \infty) = 0$. To evaluate and examine the topology of ρ, the slope of ρ (first derivative, $\nabla(\rho)$) must be consider. Since at the inflection points the slope disappears, the Hessian of ρ (second derivative, $\nabla^2(\rho)$) is also a useful quantity for getting a picture of where the electron density can build up and where there is a paucity of electron density. Besides, the likelihood of discovering electrons pair with spins 1 and 2 all the while inside two distinctive volume components $d(\vec{r}_1)$ and $d(\vec{r}_2)$, though other N-2 electrons have subjective spins and positions, is known as the pair density that can be given as

$$\rho(\overrightarrow{x_1}, \overrightarrow{x_2}) = N(N-1) \int \dots \int |\Psi(\overrightarrow{x_1}, \overrightarrow{x_2} \dots . . \overrightarrow{x_n})|^2 d\overrightarrow{x_3} \dots . . d\overrightarrow{x_n} \qquad (2.22)$$

2.3.2 The Pair Density

The pair density is the likelihood of finding a couple of electrons with spins σ_1 and σ_2 at the same time inside two distinctive volume components $d\overrightarrow{r_1}$ and $d\overrightarrow{r_2}$ while the other n-2 electrons have random positions and spins. It is given as

$$\rho(\overrightarrow{x_1}, \overrightarrow{x_2}) = n(n-1) \int \dots \int |\Psi(\overrightarrow{x_1}, \overrightarrow{x_2}, \dots \dots, \overrightarrow{x_n})|^2 \, d\overrightarrow{x_3} \dots \dots . d\overrightarrow{x_n} \qquad (2.23)$$

Like the electron density, the pair density is additionally a positive amount and is normalized to the complete number of unmistakable sets of electrons, i.e., n(n-1)*. The pair density is vital since it contains data about electron correlation.

2.3.3 The Hohenberg-Kohn Theorems

In 1964 Pierre Hohenberg and Walter Kohn[6] was first proposed density functional theory. They have demonstrated two hypotheses with respect to the electron density on which DFT has been developed. All in all, these hypotheses are the premise of Kohn-Sham density functional theory. The subtleties and verification of the two H.K. theorems have been given underneath.

Theorem 1:

'The external potential $\hat{V}_{ext}(\vec{r})$ *is within a trivial additive constant a unique functional of the electron density* $\rho(\overrightarrow{r})$*. Since* $\hat{V}_{ext}(\vec{r})$ *fixes* Ĥ *we see that the full many-particle ground state is a unique functional of* $\rho(\overrightarrow{r})$*'.*

Proof

Allow us to expect two external potentials $\hat{V}_{ext}$ and $\hat{V}'_{ext}$ (these contrast from one another by an additive constant), giving a similar ground-state electron density $\rho(\overrightarrow{r})$. These two diverse external potentials will relate to two individual electronic Hamiltonian operators Ĥ and Ĥ′ respectively, and Ĥ and Ĥ′ will correspond to two distinctive ground state wave functions, Ψ and Ψ' individually. E_0 and E_0' are the ground state energies corresponding to the two wave functions Ψ and Ψ' individually.

i) E_0 and E_0' ii) Ĥ and Ĥ′ can be expressed as follows;

$$E_0 = <\Psi|\hat{H}|\Psi>, E_0' = <\Psi'|\hat{H}'|\Psi'>, \text{ where } E_0 \neq E_0' \qquad (2.24)$$

$$\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} \qquad (2.25)$$

$$\hat{H}' = \hat{T}' + \hat{V}_{ee} + \hat{V}'_{ext} \qquad (2.26)$$

Since both wave functions give rise to the same electron density, overall we can represent them as;

$$\hat{V}_{ext} \Rightarrow \hat{H} \Rightarrow \Psi \Rightarrow \rho(\vec{r}) \Leftarrow \Psi' \Leftarrow \hat{H}' \Leftarrow \hat{V}'_{ext}$$

Now, if we apply the variational principle, where Ψ' as a trial wave function for

$$E_0 << \Psi_0 |\hat{H}|\Psi_0> = < \Psi_0 \left|\hat{H}'\right|\Psi_0 > + < \Psi_0 \left|\hat{H} - \hat{H}'\right|\Psi_0 > \qquad (2.27)$$

After putting the value of $\hat{H}$ and $\hat{H}'$,

$$E_0 < E_0' + < \Psi_0 |\hat{T} + \hat{V}_{ee} + \hat{V}_{ext} - \hat{T} - \hat{V}_{ee} - \hat{V}_{ext}'|\Psi_0 > = E_0 < E_0' + \int \rho(r)\{\hat{V}_{ext} - \hat{V}_{ext}'\}dr$$

(2.28)

Right-hand side of equation (2.28) can also be written as;

$$E_0' < E_0 - \int \rho(r)\{\hat{V}_{ext} - \hat{V}_{ext}'\}dr$$

Addition of equations (2.27) and (2.28) leads following results;

$$E_0 + E_0' < E_0' + E_0 \qquad (2.29)$$

Therefore, it has been provided that there cannot be two different $\hat{V}_{ext}$ that provides the same ground-state. The ground state energy is therefore a functional of ground-state electron density that can be represented as;

$$E_V[\rho_0] = T[\rho_0] + V_{ne}[\rho_0] + V_{ee}[\rho_0] \qquad (2.30)$$

$$E_V[\rho_0] = \int \rho_0(\vec{r})\, v(\vec{r}) d\vec{r} + T[\rho_0] + V_{ee}[\rho_0] \qquad (2.31)$$

Where, $\int \rho_0(\vec{r})\, v(\vec{r}) d\vec{r}$ is the system-dependent part, and $T[\rho_0] + V_{ee}[\rho_0]$ is the system independent part. The later part called the Hohenberg-Kohn functional $F_{HK}(\rho_0)$ (given below).

$$F_{HK}(\rho_0) = T[\rho_0] + V_{ee}[\rho_0] \qquad (2.32)$$

F_{HK} is universal functional of electron density, which is independent of external potential (V_{ext})

Up to this point, we have perceived that the ground state electron density alone is adequate to take care of all properties of the framework picked. Nonetheless, it doesn't illuminate us on the most proficient method to guarantee that the electron density utilized to assess the

properties is the genuine ground-state electron density for the framework. In the second Hohenberg-Kohn hypothesis, the remedy for this has been given.

Theorem 2:

"F_{HK} [ρ_0], the functional that delivers the ground state energy of the system, delivers the lowest energy, in the case that the input density is true ground state density, ρ_0"

This theorem can be represented as

$$E_0 \leq E[\rho'] = T[\rho'] + E_{ne}[\rho'] + E_{ee}[\rho'] \quad (2.33)$$

Proof

Allow us to expect that surmise density that will have its own Hamiltonian □′ and wave function. This wave function can be utilized as a wave function for the Hamiltonian that is created from the outside possible V_{ext}. Thus, we can address;

$$< \Psi|\hat{H}|\Psi > = T[\dot{\rho}] + V_{ee}[\dot{\rho}] + \int \rho(\vec{r})\hat{V}_{ext}d\vec{r} = E[\dot{\rho}] \geq E_0[\rho_0] =< \Psi_0|\hat{H}|\Psi_0 > \quad (2.34)$$

2.3.4 The Kohn-Sham approach

As examined in the past segment, the electron density is the major entity for figuring out the ground state properties of any nuclear or atomic framework. The Hohenberg-Kohn hypothesis proposes that the ground state energy of a nuclear or molecular framework can be written as

$$E_0 = min_\rho \left(F_{HK}[\rho] + \int \rho(\vec{r})V_{Ne}(\vec{r})d\vec{r}\right) \quad (2.35)$$

Where $F_{H.K}$ [ρ] is a universal functional; it contains the kinetic energy, the classical Coulomb, and a non-classical contribution.

$$F[\rho(\vec{r})] = T[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{ncl}[\rho(\vec{r})] \quad (2.36)$$

Here, $J[\rho]$ is known, yet the kinetic energy term (which is a significant part of complete energy) $T[\rho(\vec{r})]$ is unknown.

To figure out the kinetic energy precisely, Kohn and Sham published one paper in 1965, where the idea of the arrangement of non-interacting electrons has been referenced. The non-interfacing reference framework is built from a bunch of orbitals in the Hohenberg-Kohn formalism with the end goal that the kinetic energy can be gotten with sufficient exactness. Hence,

$T_s = -\frac{1}{2}\sum_i^N < \Psi_i|\nabla^2|\Psi_i >$ and $\rho_s(\vec{r}) = \sum_i^N \sum_s |\Psi_i(\vec{r}, s)|^2 = \rho(\vec{r})$

Where T_s is the kinetic energy, and ψ_i is the wave function of the reference framework. By utilizing this, the kinetic energy term can be determined. Regardless of whether both the frameworks (associating, just as non-interacting) have a similar electron density, it is recognizable that $T_s \neq$ T. In any case, the significant segment of the kinetic energy $T[\rho(\vec{r})]$ is recovered through T_s. To address this blunder, Kohn and Sham presented the partition of the widespread functional, where the exchange-correlation energy $E_{XC}[\rho]$ has been added. Hence, the all universal functional can be composed as;

$$F[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] \quad (2.37)$$

where, $E_{XC}[\rho] = (T[\rho] - T_S[\rho] + E_{ee}[\rho] - J[\rho]$ (2.38)

The segments of energy that are obscure or hard to acquire from other hypothetical techniques, (for example, the commitment of electron correlation and electron exchange, the adjustment for the self-interaction, the remaining part of the kinetic energy (excluded from the term T_s)), are associated with this functional ($E_{XC}[\rho]$).

The following inquiry is how we can locate the likely $V_S(\vec{r})$ for the non–collaborating reference framework, which is related with a similar density as the interacting framework? To address this issue, we will rewrite the energy of the framework as given beneath:

$$E[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \quad (2.39)$$

The further expansion of terms involved in equation (2.39) will lead to

$$E[\rho] = T_S[\rho] + \frac{1}{2}\frac{\iint \rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{EX}[\rho] + \int V_{Ne}\,\rho(\vec{r}) d\vec{r}$$

$$= -\frac{1}{2}\sum_i^N < \Psi_i |\nabla^2| \Psi_i > + \frac{1}{2}\sum_i^N \sum_j^N \iint |\Psi_i(\vec{r}_1)|^2 \frac{1}{r_{12}} |\Psi_j(\vec{r}_1)|^2 d\vec{r}_1 d\vec{r}_2 + E_{EX}[\rho] -$$

$$\sum_i^N \int \sum_A^M \frac{Z_A}{r_{MA}} |\Psi_i(\vec{r}_1)|^2 d\vec{r}_1 \qquad (2.40)$$

In equation (2.40) the only unknown term is $E_{XC}[\rho]$. Now, according to the variational principle, upon minimization of the energy under the constraint

$<\Psi_i|\Psi_j>=\delta_{ij}$ leads to the *Kohn-Sham equation*.

$$-\frac{1}{2}\nabla^2 + [\int \frac{\rho(r_2)}{r_{12}} dr_2 + V_{XC}(r_1) - \sum_A^M \frac{Z_A}{r_{MA}}])\Psi_i = \varepsilon_i \Psi_i \qquad (2.41)$$

Where, $V_{XC}(\vec{r})$ can be defined as;

$$V_{XC}(\vec{r}) = \frac{\partial E_{XC}[\rho(\overrightarrow{r})]}{\partial \rho(\overrightarrow{r})}$$

New, improved potential $V_{eff}(\vec{r})$ can be determined from these equations, which will ultimately give self-consistency. Besides, the hypothetical mapping in the middle of the kinetic energy and the density can be done by utilizing ψ_i s. Be that as it may, ψ_i s isn't comparable to the genuine orbitals of the framework. Besides, the Kohn-Sham wave function is a solitary determinant approach that fails in the case of multiple determinants. DFT is subsequently utilized to discover better approximations to these two quantities (V_{xc} and E_{XC}) that will give functionals and permit the calculation of the energy of the frameworks.

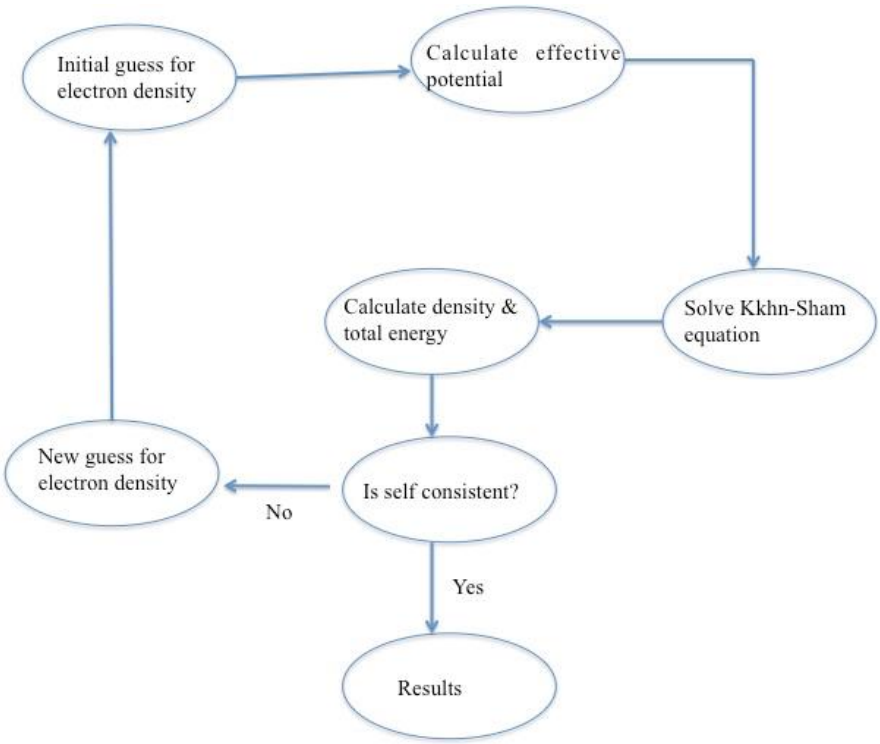


Figure 2.1 The flow chart of the self-consistent cycle for Kohn-Sham iterations for the single-step during optimization.

2.3.5 Functional

The functional is defined as the function of another function. The functional takes a function as an input and provides an output, whereas a function takes a number as an input to give an output. For instance, the variational integral $F[\phi] = <\phi|\hat{H}|\phi>/<\phi|\phi>$ is a functional of the variation function Φ. Like a function, one can find the derivative of a functional. The differentiation of a functional F[g] can be expressed as;

$$\partial F[g] = F[(g + \partial x)] - F[g] = \int \frac{\partial F}{\partial g(x)} \partial g(x).dx \qquad (2.42)$$

the total differential of a function F (g_1,g_2,……..,g_n) $dF = \sum_{i=1}^{n} \frac{\delta F}{\delta g_i} dg_i$

where g_1,g_2,….,g_n are independent variables. The functional derivative $\delta F / \delta g(x)$ has a role

similar to that of partial derivative $\partial F/\partial g_i$, here x is the integration variable used in the function as part of the functional. The rules of differentiation are also similar to the functions. Different types of functional such as generalized gradient, local-density, hybrid approximations and meta-generalized gradient, can be employed in the DFT method.

2.3.5.1 Local-Density approximation (LDA)

LDA is pertinent to the homogeneous gas framework, where the electron density $\rho(\overrightarrow{r})$ varies gradually with the position. Specifically, the electron density can be dealt with locally as a uniform electron gas in this estimate. LDA is the simple estimate to discover the exchange-correlation. The exchange-correlation term can be written as

$$E^{LDA}[\rho] = \int \rho(\overrightarrow{r}) \in_{XC}(\rho(\overrightarrow{r}))\, dr \qquad (2.43)$$

where $\in_{XC}$ is the exchange-correlation energy per particle of a homogeneous electron gas of electron density ρ. The exchange-correlation energy can be written as the addition of correlation and exchange terms.

$$\in_{XC}(\rho) = \in_{X}(\rho) + \in_{C}(\rho) \qquad (2.44)$$

For closed-shell systems, the LDA approximation is applicable, while LSDA (local spin density approximation) is relevant for open-shell systems (free radical) calculations.

2.3.5.2 Generalized Gradient Approximation (GGA)

LDA does not give reliable energies when the density changes significantly with distance, as in atoms. In order to tackle this issue, the approach is to consider the gradient of the electron density ($\nabla(\rho)$), and think of the exchange-correlation term in the energy expression. $E_{XC}^{GGA}[\rho]$ can be split into the exchange and correlation terms:

$$E_{XC}^{GGA}[\rho] = E_{X}^{GGA}[\rho] + E_{C}^{GGA}[\rho] \qquad (2.45)$$

In 1988, Becke presented the B88, Bx88, and Becke88 GGA exchange functionals, which

improve DFT calculations. The normally utilized GGA correlation functionals are Perdew (P86 or Pc86), Perdew-Wang (PW91 or PWc91), and Lee-Yang-Parr (LYP functional, for example, B88LYP or BLYP). Besides, we can join any exchange functional with the correlation functional in DFT estimations. For instance, the BLYP functional addresses the blend of the Becke exchange functional and the LYP correlation functional. At present, both observational and nonempirical GGA functionals are accessible for use. Perdew-Burke Ernzerhof (PBE), a widely utilized GGA functional, does not include experimentally determined parameters.

2.3.5.3 Meta-Generalized Gradient Functionals

Meta-GGA functionals are an augmentation of GGA functionals that rely upon the second derivatives of the electron density (kinetic energy density). In the meta-GGA estimate, the exchange correlation energy can be determined as;

$$E_{XC}^{MGGA}\left[\rho^{\alpha},\rho^{\beta}\right] = \int f(\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta},\nabla^{2}\rho^{\alpha}\nabla^{2}\rho^{\beta},\tau^{\alpha},\tau^{\beta})\,dr \qquad (2.46)$$

where, $\tau_{\alpha}, \tau_{\beta}$ are the non-interacting kinetic energy density terms that can be defined as $\tau_{\alpha} = \frac{1}{2}\sum_{i}|\nabla\theta_{i\alpha}^{KS}(r)|^{2}$ where the $\theta_{i\alpha}^{KS}$ are the Kohn-Sham orbitals for the electrons that have α spin. Meta-GGA functionals are more exact, yet computationally more costly than GGA functionals. For instance, M06-L is a meta-GGA functional that gives a superior presentation as it contains 37 streamlined parameters. Another illustration of a meta-GGA functional is TPSS, a nonempirical meta-GGA functional that gives preferable outcomes over the PBE functional.

2.3.5.4 Hybrid Functionals

These days, generally utilized functionals are hybrid exchange-correlation functionals, which give precise outcomes. These useful functionals incorporate an accurate exchange term from Hartree-Fock, alongside the exchange correlation energy from different sources; for example, observational and *ab initio*. The HF exchange energy, given underneath, is the specific exchange energy for an arrangement of non-interacting electrons, with electron density equivalent to the genuine framework.

$$E_{X}^{HF} = -\sum_{I=1}^{n}\sum_{j=1}^{n} < \Psi_{i}^{KS}(1)\Psi_{j}^{KS}(2)\left|\frac{1}{r_{ij}}\right|\Psi_{i}^{KS}(2)\Psi_{j}^{KS}(1) > \qquad (2.47)$$

In 1993 Axel Becke presented a hybrid approach for DFT calculations, which adjusted by Stevens *et al.* in 1994. The most mainstream hybrid DFT functional is B3LYP or Becke3LYP. It is trying to say what DFT functional is best on the grounds that the utilization of functional relies upon the sort of framework and properties that are being determined.

2.4 The Volume Corrections Method for Determining the Translational Entropy

Determination of entropy is an important task to calculate the free energy of any reaction. It is well known that the total entropy is the summation of translational, rotational, vibrational, and electronic components.

$$S_{tot} = S_{trans} + S_{rot} + S_{vib} + S_{el} \quad (2.48)$$

The translational entropy can be calculated by using the Sackur-Tetrode equation,[7] which can be presented as;

$$S_{trans} = Rln[(\frac{10^{-15/2}}{N_A^4[X]})(\frac{2\pi MRTe^{5/3}}{h^2})^{3/2}] \quad (2.49)$$

Where [X] is the concentration of the molecule, M is the mass of the particle, T is the temperature N_A is the Avogadro number, k is the Boltzmann constant and h is the Planck constant. However, in the solution phase, the Sackur-Tetrode equation overestimates the translational entropy, as it avoids the molecular volume (V_{mol}). Mammen *et al.* introduced the *free volume* correction to address this problem, which calculates the exact translational entropy.[8] According to this method, it has been assumed that the volume of molecules in solution is smaller than the total volume.

$$V_{free} = C_{free}\left(\left(\sqrt[3]{\frac{10^{27}}{[X]N_0}} - \sqrt[3]{V_{mol}}\right)\right)^3 \quad (2.50)$$

Where C_{free} is 8 (for cube), N_0 is the Avogadro's number. After the free volume correction to Equation 2.48, the translational entropy is represented by

$$S_{trans}^{analyte} = Rln\left[\left(\frac{10^{-15/2}V_{free}^{solvent}}{N_0^4[analyte]}\right)\left(\frac{2\pi MRTe^{5/3}}{h^2}\right)^{3/2}\right] = 11.1 + 12.5\ln(T) + 12.5\ln(M) +$$

$$8.3lnV_{free}^{solvent} \qquad (2.51)$$

Where, $V^{solvent}$ is the free volume, T is the temperature. Therefore, this equation contributes the free translational entropy in solution.

2.5 *Ab initio* Molecular Dynamics

The essential basis of the well known *ab initio* molecular dynamics (AIMD) approach is to tackle the electronic SE so as to to calculate the nuclear potential. Nuclear forces can be determined from the obtained potential, and on the other hand Newton's equation of motion can be enforced for determining the nuclear movement. This methodology in its various variations charts the road for the Ehrenfest MD (EMD), Car-Parrinello MD (CPMD) and Born-Oppenheimer MD (BOMD) methods.

We will provide a brief introduction of every one of these *ab initio* techniques in the following three sections. The objective is not give a point by point depiction of these techniques - just an outline without talking about all the subtleties. The peruser is alluded to Ref. [9] for subtleties and more comprehensive portrayal.

2.5.1 Ehrenfest Molecular Dynamics

EMD involves solving the following equations of motion:

$$M_n\ddot{R}_n = -\nabla_n < \Psi|H_N|\Psi >,$$

$$i\hbar\frac{\partial\Psi_{R(t)}}{\partial t} = H_N \qquad (2.52)$$

here H_N is time-dependent through the coordinates of nuclei $\{R_n\}$. This technique is known as a self-consistent mean field strategy, where the time-dependent electronic SE is calculated for each time step of Newton's equation of motion.

A typical extension, which is regularly presented in EMD, is to initially expand the

electronic wave function in an adiabatic basis,[10] and from there on confine the entire electronic wave function to a ground state in that basis. The time-independent electronic SE in the adiabatic basis is represented by

$$H_N\Psi_k(r;R) = E_k(R)\Psi_k(r;R) \qquad (2.53)$$

Therefore, in order to construct the electronic wave function, the solution Ψ_k of equation 2.53 is the linear combination of complex, time-dependent coefficients. Mathematically, it can be represented by

$$\Psi_{R(t)}(r;t) = \sum_{j=0}^{\infty} c_j(t)\Psi_k(r;R), \qquad (2.54)$$

where the coefficients portray how the inhabitance of the various states advances over the long run. It only comprises of the initial term from the overall sum for confining the ground state electronic wave function in this basis. As the time development of the wave function relates to a unitary spread,[9] the average value of H_N was first minimized by the ground state wave function, remaining in its corresponding minima with the motion of the nuclei.

The electronic motion is used to determine the timescale and hence the timestep to integrate the equations of motion in EMD. Accordingly, as the nuclei are much heavier than the electrons, the motion of the electrons are much faster than the nuclei. This increases the computational expense of the EMD, and thus one is bound to work with a lot more modest time scales than the one given by the motion of the nuclei. Due to this reason, the EMD is not in common use, except if the framework has numerous degrees of freedom, despite the fact that it has been utilized to study collision and scattering-type problems.[9]

2.5.2 Born-Oppenheimer Molecular Dynamics (BOMD)

The basis of the BOMD is the Born-Oppenheimer estimate (shown in Section 2.2.2), the fact that nuclei are substantially more enormous than electrons. Due to the large contrast in mass among nuclei and electrons, the latter can behave like particles which follow the nuclear movement. It implies that the electrons react quickly to the motion of the nuclei and relax to the ground-state. Therefore, we may subsequently think that with respect to the motion of the

electrons, the nuclei are fixed. This signifies that we can solve the electronic wave function by fixing the nuclear configuration.

In BOMD, the calculation of the electronic structure is reduced to the arrangement of the TISE, which is then utilized for calculations of forces following up on the nuclei at each time step. The classical movement of the nuclei governs the time-dependency of the electrons. In comparison to the EMD, where the dynamics is resolved from the time-dependent electronic SE. Along these lines, BOMD time step is dictated by the motion of the nuclei. Therefore, there is an immense preference of BOMD in comparison with EMD, where the time step is dictated by the motion of the electrons. However, the major drawback of the BOMD method is that at each time step, minimization is needed.

2.5.3 Car-Parrinello Molecular Dynamics (CPMD)

A major development in AIMD was the improvement of the Car-Parrinello strategy,[11] which made it possible to treat enormous scope issues through *ab initio* MD. CPMD endeavors to consolidate the upsides of EMD and BOMD, and simultaneously stay away from their weaknesses. The advantage of EMD is the initial energy minimization of the wave function, so that the system remains in its corresponding minima with the nuclear movement. However the major impediment is the determination of time step by the motion of the electrons. On the other hand in BOMD, though the nuclear movement determines the time, in each step minimization is required. The Car-Parrinello strategy integrates the equations of motion on the large time scale set by the motion of the nuclei, while simultaneously, at each time step, the energy minimization is avoided. In the CPMD method, to keep the electrons close to the ground state, the electrons are unequivocally included as active degrees of freedom. Also in each time step there is no requirement of an electronic minimization. Because of an underlying, standard minimization, the imaginary electronic dynamics keep the electron close to the ground state. An extensive discussion about the CPMD strategy is represented in Ref. [16].

2.6 *Ab Initio* Nanoreactor Dynamics

Experimental chemistry regularly plays the primary role in finding new species and proposing new mechanistic pathways, and computational science offers significant support by

mediating between contending prospective mechanistic pathways. Ongoing computational advances, including those that influence graphics processing unit (GPU) architectures,[12-15] could make the way for utilizing calculations not exclusively to help decide between various speculations, but also, in addition, to uncover new chemical reaction mechanisms. The experimentally inspired[16] *ab initio* nanoreactor (AINR) achieves this utilizing an *ab initio* molecular dynamics (AIMD) simulation approach, combined with programmed investigation and clarification strategies to assemble a quantitatively exact reaction network. By employing the AINR with different reactants accessible in different conditions, for example, in the prebiotic earth or in interstellar space, one can investigate reactivity and find new mechanistic pathways. This methodology will help direct investigations by presenting new speculations and proposing novel trials.

The statistical uncommonness of enacted chemical reactions limits most AIMD studies to explicit changes along a picked reaction coordinate.[17-19] A new way to deal with conquering the uncommonness of reactive occasions has been the utilization of preordained heuristic rules[20-22] or geometry searching[23-24] to produce new species and networks of chemical reactions. Depending upon the essential conditions of quantum and classical mechanics, the AINR finds new molecules and reactions. Reactions happen openly without predefined reaction coordinates or rudimentary steps.

Although recent advances in AIMD have given a lot of computational help, these simulations remain exorbitant for inspecting enormous quantities of reactive events. To surmount this problem, Martinez and co-workers have incorporated new speeding up methods in the nanoreactor. A virtual piston intermittently pushed molecules towards the focal point of the nanoreactor and thus upgrades it reactivity, which enormously increases the frequency of collision and the crossing of the activation energy barrier. This summons thoughts from high-pressure experiment and shock-wave simulations.[25-27]

The nanoreactor accomplishes its objective of extensively investigating reaction pathways by taking a moderate position between genuinely reasonable dynamics and rule-based list approaches. The *AINR* simulation guarantees that the reaction trajectories follow the Newtonian

mechanics and stays away from a combinatorial blast of potential outcomes. This methodology is substantial as long as the important reactions are inspected in any event and remembered for the information base.

2.7 References

1. A. Paul Medwick, *Ann. Hist. Comp.* **1988**, *10*, 105.

2. C. Moller, M.S. Plesset, *Phys. Rev*. **46** (7), 618-622.

3. C. J. Cramer, Chichester: *John Wiley & Sons, Ltd*. **2002**, 191.

4. H. G. Kümmel, *Int. J. Mod. Phys. B*, **2003**, *17*, 5311.

5. a) M. H. N. Assadi, D. A. H. Hanao *J. Appl. Phys*. **2013**, *113*, 233913; b) M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, M. C. Payne, *J. Phys: Condens. Matter*, **2002**, *14*, 2717-2743; c) F. R. Somayeh, H. Soleymanabadi, *J. Mol. Model*. **2014**, *20*, 2439-2445; d) F. R. Somayeh, H. Soleymanabadi, *J Mol Model*, **2013**, *19*, 3733-3740; e) D. Music, R. W. Geyer, J. M. Schneider, *Surface & Coatings Technology*, **2016**, *286*, 178.

6. (a) P. Hohenberg, W. Kohn, *Physical review*, **1964**, *136*, B864. (b) W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, 1133.

7. R.W. Gurney, 1st ed.; McGraw Hill, 1949.

8. M. Mammen, E. I. Shakhnovich, J. M. Deutch, G. M. Whitesides, *J. Org. Chem.***1998**, *63*, 3821.

9. D. Marx and J. Hutter. *Forschungszentrum Julich*, **2000**. *2000*, 329.

10. M. Griebel, S. Knapek, and G. Zumbusch. *Texts in Computational Science and Engineering. Springer*, **2007.**

11. D. Marx and J. Hutter. *Cambridge University Press*, **2009.**

12. Ufimtsev, I. S. & Martinez, T. J. *J. Chem. Theory Comput*., **2009**, *5*, 2619–2628.

13. Ufimtsev, I. S., Luehr, N. & Martinez, T. J. J. Phys. Chem. Lett. 2, 1789–1793 (2011).

14. Luehr, N., Ufimtsev, I. S. & Martinez, T. J. J. Chem. Theory Comput. 7, 949–954 (2011).

15. Kulik, H. J., Luehr, N., Ufimtsev, I. S. & Martinez, T. J. J. Phys. Chem. B 116, 12501–12509 (2012).

16. Yin, Y. *et al*. *Science*, **2004**, *304*, 711–714.

17. Ensing, B., De Vivo, M., Liu, Z. W., Moore, P. & Klein, M. L. Acc. Chem. Res. 39, 73–81 (2006).

18. Pietrucci, F. & Andreoni, W. Phys. Rev. Lett. 107, 085504 (2011).

19. Iannuzzi, M., Laio, A. & Parrinello, M. Phys. Rev. Lett. 90, 238302 (2003).

20. Zimmerman, P. M. J. Comput. Chem. 34, 1385–1392 (2013).

21. Rappoport, D., Galvin, C. J., Zubarev, D. Y. & Aspuru-Guzik, A. J. Chem. Theory Comput. 10, 897–907 (2014).

22. Virshup, A. M., Contreras-García, J., Wipf, P., Yang, W. & Beratan, D. N. J. Am. Chem. Soc. 135, 7296–7303 (2013).

23. Maeda, S. & Morokuma, K. J. Chem. Theory Comput. 8, 380–385 (2012).

24. Wales, D. J., Miller, M. A. & Walsh, T. R. Nature 394, 758–760 (1998).

25. Goldman, N., Reed, E. J., Fried, L. E., Kuo, I. F. W. & Maiti, A. Nature Chem. 2, 949–954 (2010).

26. Goldman, N. *et al*. J. Chem. Phys. 130, 124517 (2009).

27. Bernasconi, M., Chiarotti, G. L., Focher, P., Parrinello, M. & Tosatti, E. Phys. Rev. Lett. 78, 2008–2011 (1997).

Chapter 3

Insights into Chemical Reactions at the Beginning of the Universe: From HeH^+ to H_3^+

Chapter 3

Insights into Chemical Reactions at the Beginning of the Universe: From HeH^+ to H_3^+

Abstract

At the dawn of the universe, the ions of the light elements produced in the Big Bang nucleosynthesis recombined with each other. In our present study, we have tried to mimic the conditions in the early universe to show how the recombination process would have led to the formation of the first ever formed diatomic species of the universe: HeH^+, as well as the subsequent processes that would have led to the formation of the simplest triatomic species: H_3^+. We have also studied some special cases: higher positive charge with fewer number of hydrogen atoms in a dense atmosphere, and the formation of unusual and interesting linear, dicationic He chains beginning from light elements He and H in a positively charged atmosphere. For all the simulations, the *ab initio* nanoreactor (*AINR*) dynamics method has been employed.

3.1 Introduction

The way the universe, and all the elements, came into being is one of the fascinating questions of science. Attempts to answer this question has led to the big bang theory, and an understanding of the primeval universe and the entities that it was made up of.[1] Further advancement of science and technology unfolded subtle but emphatic nuances, which led NASA's Stratospheric Observatory for Infrared Astronomy (SOFIA) to the detection of HeH^+, the first molecule formed after the big bang,[2] after 94 years since its discovery in the laboratory in 1925.[3]

As the first molecule, the significance of the role of HeH^+ in the evolution of other species cannot be overstated. One of these species, and perhaps the most important one, is the simplest polyatomic molecule H_3^+, which has always intrigued researchers ever since its discovery in 1911 by J.J. Thomson.[4a,b] However, the importance H_3^+ in astrochemistry was realized only after it was detected on Jupiter in the 1980s.[5-6] High abundance of H_3^+ in the universe and its ability to donate a proton established this triatomic cation as the interstellar acid of utmost importance for many extra-terrestrial reactions.[7-10] While there are many reports of H_3^+ formation from doubly ionized organic molecules,[11-19] our focus is on its origin and the role played by HeH^+ on its formation.

H_3^+ formation was first reported to be formed primarily from the combination of H_2^+ and H_2, where H_2^+ is formed from the ionization of H_2.[3] There are other reports which state that H_2^+ is more likely to be formed from the combination of HeH^+ and H.[20-24] At the same time, the possibility of HeH^+ combining with H_2 to produce H_3^+ cannot be overlooked.[25] Thus, many factors can influence the origin of H_3^+, but there have not been any conclusive studies yet.

In this work, we have employed the *ab initio* nanoreactor (AINR) method to carry out full quantum mechanical molecular dynamics (MD) simulations on systems containing atoms/ions of helium and hydrogen, and have obtained reaction profiles by varying their mixture ratio and the charge. The *AINR* method, developed by Martinez and co-workers allows the determination of new reaction pathways and products, without the need of controlling the chemical system.[26-28] Our primary goal was to gain insight into the formation of different species from the combination of He and H in the presence of a positively charged atmosphere, as well as their further dissociation and recombination. As the Results and Discussion section will show, our studies provide interesting new insights into HeH^+ formation, and sheds light on various short-lived

intermediates that could have formed *en route* to obtaining H_3^+– the stable species that was known to exist in the early universe.[8]

3.2 Computational Methods:

3.2.1 *Ab Initio* Molecular Dynamics (AIMD) Simulations. TeraChem 1.9[29-35] software package has been employed for performing the AIMD simulations where Born−Oppenheimer potential energy surface calculated by using the Hartree−Fock (HF)[36] electronic wave function and the 6-311g[37] Gaussian basis set. This method has been implemented in TeraChem by Martinez and co-workers. This approach was deemed acceptable because the HF method is well-known for predicting chemically reasonable structures.[38] Also, it should be noted that HF was not employed to determine the thermodynamics and reaction rates of the reactions: its only role was in the discovery process. This was also the approach employed by Martinez and co-workers in their original *AINR* paper (employing HF/3-21g), where they replicated the results obtained from the Urey−Miller experiment, as well as from the interaction of acetylene molecules. The similar method (HF) was also employed by us in our previous report on reaction pathways leading to the formation of precursors of RNA and sugars.[39]

The results were obtained from the *AINR* simulations by varying both the He to H ratio as well as the positive charge of the system. The system was constrained in a spherical boundary of 4.0 and 2.0 A radii, so that the atoms resided in a space that alternated between the volumes created by these two radii, and collided with each other. Each *AINR* dynamics was evolved upto 15 ps, with a time step of 0.5 fs.

Newton's equations of motion were calculated using Langevin dynamics, with an equilibrium temperature of 1000.0 K (also the dynamics starting temperature). We have used this high temperature in order to increase the average kinetic energy of the reactants and for faster dynamics, as well as to try and mimic the early universe atmosphere. The nanoreactor simulations employ a piston to accelerate the reaction rate. We have employed the augmented direct inversion in the iterative subspace (ADIIS) algorithm[40] available in TeraChem as an alternative tool for self-consistent field calculations at each AIMD step in which the default DIIS algorithm[41] failed to converge. Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event.

The mechanistic pathways obtained from the *AINR* simulations were then analyzed with full quantum mechanical (QM) calculations. All the structures were optimized with coupled cluster singles doubles (CCSD)[42] and using the 6-311++g(d,p)[43] basis set. Gaussian09 software[44] was employed for the thermodynamic calculations.

3.3 Results and Discussions

In this section, we will briefly describe the formation of H_3^+ in the *AINR via* different short lived intermediates. We have taken a fixed composition of the He and H mixture and varied the overall positive charge density of the system (as shown in the Tables 1 and 2). During the simulations in each case, it was seen that HeH^+ formed at the very beginning of the dynamics as the first molecular species. In our first set of simulations, we have taken a homogeneous mixture of 30 atoms each of H and He. The *AINR* makes them collide with each other at a temperature of 1000.0 K. The simulation with no positive charge in the system does not produce any intermediates and H_3^+ at all throughout the dynamics. This led us to consider the possibility that a more appropriate set-up would include a positively charged system, which would mimic the collisions between the ionized state of the helium and hydrogen atoms present at the beginning of universe.[8] A positively charged environment for the formation of H_3^+ had also been considered by many previous reports, while investigating its origin from different organic molecules.[49] Therefore, we have varied the positive charge of the system by even numbers (Table 1) during the *AINR* dynamics. As the dynamics progressed, various short lived species such as He_2^{2+}, He_3^{2+} and He_2H^+ (snapshots shown in the Figures 3.1, 3.2, 3.3, 3.4) were seen to have formed in almost every simulation, though their time of appearance was different in each case. It was also observed that with the increase of the positive charge of the system, the formation of H_3^+ ions also increased, up to a point. The number of H_3^+ ions generated was equal to the positive charge in the system, up to a charge of +6 (see Table 3.1 below).

Table 3.1: *AINR* simulations with 30 He atoms and 30 H atoms: different entries represent the variation of the total positive charge of the system – by even numbers.

Total Charge	First Molecule	Intermediate Species	Dominant End Molecule	No. of H_3^+
0	-	-	-	-
2	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+, H_2	H_3^+	2
4	HeH^+	He_2^{2+}, He_2H^+,H^+, H_2	H_3^+	4
6	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+, H_2	H_3^+	6
8	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+, H_2	H_3^+	7
10	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+, H_2	H_3^+	7
20	HeH^+	He_2^{2+}, He_2H^+, H^+, H_2	H_3^+	5

However, upon further increase in the positive charge of the system beyond six – to eight or ten, the number of H_3^+ ions formed was not seen to be equal to the total positive charge of the system. Instead of H_3^+, the remaining positive charge of the system was balanced by H^+ or, in some cases, HeH^+. As shown in Table 3.1, in case of a positive charge of 10 and after 250 fs, we observed only seven H_3^+ ions remaining with three H^+, which balanced the total charge of the system. Natural population analysis (NPA) or the formal charge analysis has performed for all the atoms in a snapshot shown in table 3.2

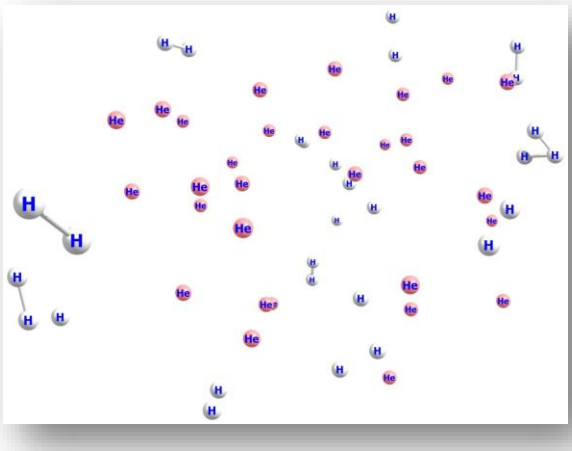


Figure 3.1. Snapshot of H_3^+ formation taken at 10 fs during the dynamics with 30 H and 30 He atoms with the aid of *AINR* approach.

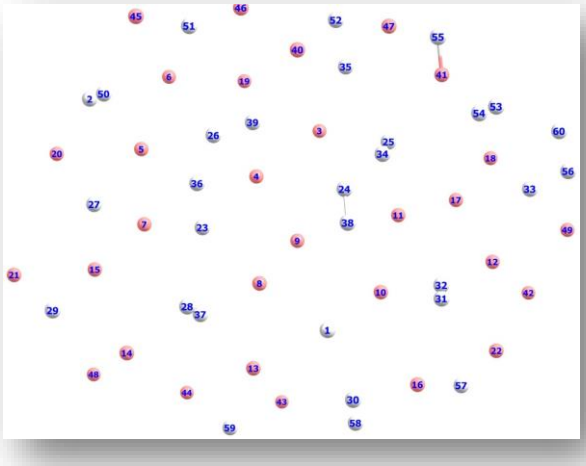


Figure 3.2. Snapshot of HeH^+ formation taken at 6 fs during the dynamics with 30 H and 30 He atoms with the aid of *AINR* approach.

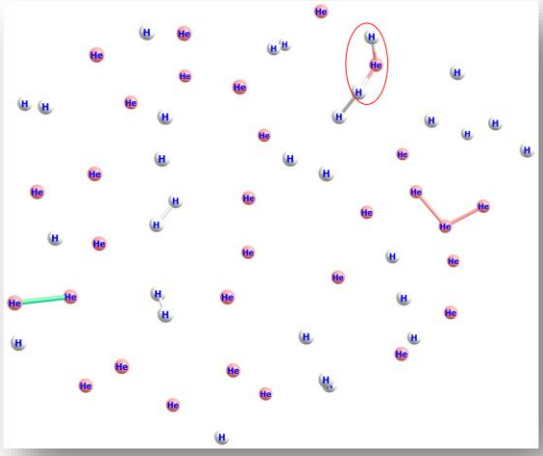


Figure 3.3. Snapshot of HeH_2^+ and He_3^{2+} formation taken at 52 fs during the dynamics with 30 H and 30 He atoms with the aid of *AINR* approach.

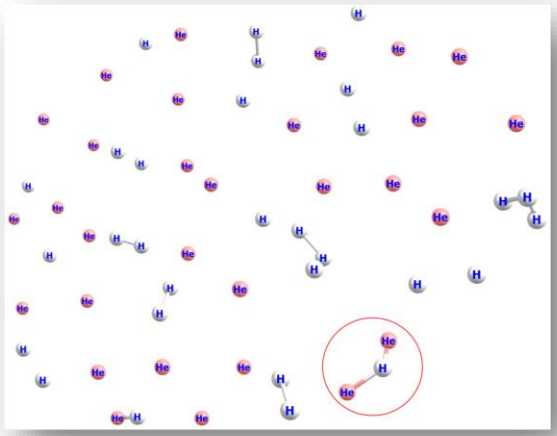


Figure 3.4 Snapshot of formation of short lived species He_2H^+ taken at 13 fs during the dynamics with 30 H and 30 He atoms with the aid of *AINR* approach.

Table 3.2. NPA charge analysis during the formation of H_3^+ by m06-2x/6-311++g(d,p) level of theory for the *AINR* dynamics with 30 atoms of H and 30 atoms of He taking six overall positive charge of the system.

1	H	-0.104183	31	H	0.363576
2	H	-0.104825	32	H	0.196063
3	He	0.008448	33	H	0.175834
4	He	0.012348	34	H	0.147991
5	He	0.003741	35	H	0.226533
6	He	0.000646	36	H	0.130613
7	He	-0.001848	37	H	0.248657
8	He	-0.003058	38	H	-0.008940
9	He	0.024006	39	H	0.531945
10	He	0.004825	40	He	0.005419

11	He	0.027614	41	He	-0.008979
12	He	-0.003834	42	He	0.033640
13	He	0.002035	43	He	0.005884
14	He	0.005294	44	He	0.049965
15	He	0.004756	45	He	0.005650
16	He	0.009581	46	He	-0.003980
17	He	0.019013	47	He	0.000942
18	He	0.000175	48	He	0.030665
19	He	0.006637	49	He	-0.005486
20	He	0.060766	50	H	0.286105
21	He	-0.000642	51	H	0.038690
22	He	0.007568	52	H	0.381589
23	H	-0.095962	53	H	0.087634
24	H	0.379667	54	H	0.094407
25	H	0.336298	55	H	0.387090
26	H	0.206802	56	H	0.709515
27	H	0.200995	57	H	-0.115415
28	H	-0.042724	58	H	0.105122
29	H	0.902932	59	H	0.087735
30	H	0.124714	60	H	-0.180251

Similarly, in another set of MD simulations, we have taken 29 H with 30 He atoms and varied the overall charge of the system by an odd number: 1, 3, 5 and so on. These observations

have been shown in Table 3.3 below. We have observed a similar trend for the formation of H_3^+ as the only end product up to a certain limit (here, the value is 5) of positive charge and beyond that, the total charge of the system was seen to be balanced by the sum of H_3^+, H^+ and HeH^+, as seen in the previous section when the positive charge was varied by even numbers.

In short, we can say that in all the cases of AINR dynamics studied, the formation of HeH^+ as the first molecule was observed. However, upon varying the total positive charge of the whole system, several short-lived species (He_2H^+, He_3^{2+}, He_2^{2+}) were observed (Table 3.1 and 3.2) after HeH^+ formation. At the end of the simulation, H_3^+ and H_2 were found to be the only stable species left in the reaction mixture.

Table 3.3. *AINR* simulations with 30 He atoms and 30 H atoms: different entries represent the variation of the total positive charge of the system – by odd numbers.

Total Charge	First Molecule	Intermediate Species	Dominant End Molecule	No. of H_3^+
0	-	-	-	-
1	HeH^+	He_2H^+, H^+,H_2	H_3^+	1
3	HeH^+	He_2^{2+}, He_2H^+, H^+, H_2	H_3^+	3
5	HeH^+	He_2^{2+}, He_2H^+, H^+, H_2	H_3^+	5
7	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+, H_2	H_3^+	6
9	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+, H_2	H_3^+	7
11	HeH^+	He_2^{2+}, He_2H^+, H^+, H_2	H_3^+	7
21	HeH^+	He_2^{2+}, He_2H^+, H^+, H_2	H_3^+, HeH^+	4

H_3^+ and other short lived molecules formation timescale. The formation timescale of different short lived species, along with the stable H_3^+, has been observed from femtosecond *AINR* simulations. In each and every simulation, HeH^+, which has been proposed to be the first formed molecule, was formed soon after the beginning of the dynamics. The time of appearance of HeH^+ is within 15 fs timesteps. Subsequently, other short lived species (He_2H^+, He_3^{2+}, He_2^{2+}) were

formed within the timescale of 0.1 ps (shown in Table 3.4). The observed timescale for the existence of such transient species is around 5-10 fs. Once these molecules are formed, they quickly dissociate and this ultimately leads to the formation of H_3^+ which also be observed throughout the dynamics. From the AINR dynamics, we have analyzed the data and found two pathways for the formation of H_3^+ starting from He and H in atomic states within the positively charge atmosphere. Both of the pathways involved the well known roaming hydrogen mechanism.[11,16,19] The most feasible pathway for H_3^+ formation is the abstraction of a proton from the first molecule HeH^+ by the roaming dihydrogen (shown in Figure 3.5).

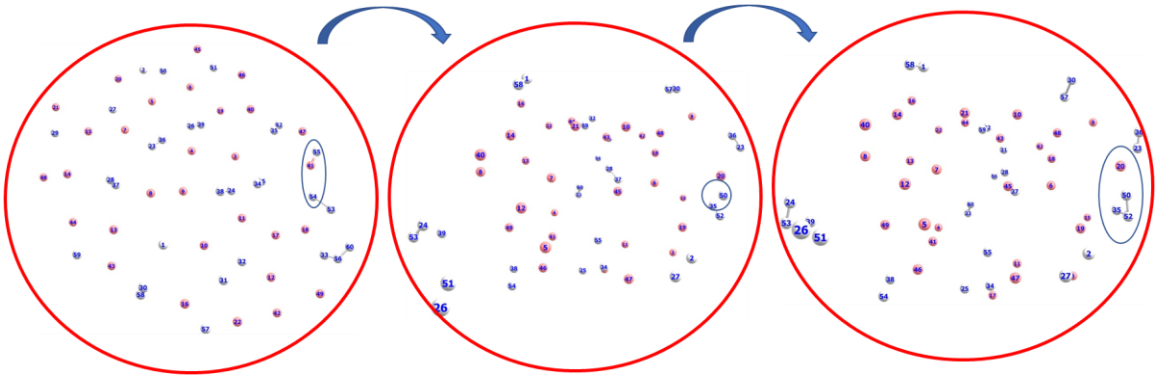


Figure 3.5 Snapshots of *AINR* simulations towards the formation of H_3^+ from HeH^+ and dihydrogen. HeH^+ has formed at the very beginning of the dynamics starting from atomic He and H. The snapshot has taken after 6 fs timescale.

The reaction free energy (ΔG) for this step has been calculated to be -32.2 kcal/mol (shown in Scheme 3.1). In another mechanistic pathway, there is no involvement of HeH^+. Instead of HeH^+, the proton abstraction occurs from a monocationic dihydrogen molecule by the roaming dihydrogen. This process is thermodynamically favourable by 27.5 kcal/mol.

$$HeH^+ + H_2 \xrightarrow{\Delta G = -32.2 \text{ kcal/mol}} H_3^+ + He$$

$$H_2^+ + H_2 \xrightarrow{\Delta G = -27.5 \text{ kcal/mol}} H_3^+ + H$$

Scheme 3.1. Pathways for H_3^+ formation. Values have been calculated at the CCSD/6-311++g(d,p) levels of theory in kcal/mol.

The total number of H_3^+ molecules formed is also found to be directly correlated with the total charge of the system, as well as the number of He and H atoms taken. Greater charge in the system yielded more short-lived species during the simulations. Most of the intermediate species were found to be formed within 100 fs (Table 3.4) and they were found to exist for only about 5-10 fs during the *AINR* simulations.

Table 3.4 Time (in fs) of first appearance of different species:

Total Charge	HeH^+	He_2H^+	He_2^{2+}	He_3^{2+}	H_3^+
0	-	-	-	-	-
2	4.5	11.0	10.0	25.5	27.0
4	9.0	25.5	19.5	80.0	45.0
6	5.0	12.5	16.0	49.5	9.0
8	4.0	12.5	22.5	53.0	8.0
10	5.5	18.0	14.0	19.0	7.0
20	4.0	26.5	14.5	22.0	9.0

Since it has been postulated that different ratios of helium to hydrogen atoms could have existed in the early universe,[1] we have further performed *AINR* dynamics with a 1:3 ratio of helium to hydrogen atoms and varied the total positive charge of the system (see Table 3.5 below). In such simulations, we have observed trends similar to those discussed in the previous sections, like the formation of HeH^+ as the first molecule and the subsequent formation of transient species (He_2H^+, He_3^{2+}, He_2^{2+}), leading eventually to H_3^+ formation.

Table 3.5: Time of occurrence (in fs) of different species from the *AINR* simulation of 1: 3 ratio of helium to hydrogen while varying total positive charge of the system.

Total Charge	HeH^+	He_2H^+	He_2^{2+}	He_3^{2+}	H_3^+
0	-	-	-	-	-
4	5.0	74.5	14.0	24.0	12.5
6	9.5	15.5	12.0	27.0	17.0

8	6.0	54.0	15.0	19.5	21.5
10	10.0	16.0	14.5	30.0	19.0
12	6.5	70.5	16.0	23.0	23.0

Since there has been speculation on the exact nature of the formed ion He_2H^+: whether it was formed as $[He\text{-}H\text{-}He]^+$ or as $[He\text{-}He\text{-}H]^+$,[48] we have addressed this issue *via AINR* dynamics followed by static DFT calculations. As shown in Figure 3.6 (a,b), two different routes leading to the formation of $[He\text{-}H\text{-}He]^+$ and $[He\text{-}He\text{-}H]^+$ were observed during the simulations, generated from the collision of HeH^+ and He. The thermodynamics was evaluated and it was found that the formation of the $[He\text{-}H\text{-}He]^+$ species was exergonic by 32.3 kcal/mol whereas the formation of $[He\text{-}He\text{-}H]^+$ was only favourable by 4.5 kcal/mol. In other words, our calculations indicate that He_2H^+ would have formed predominantly as $[He\text{-}H\text{-}He]^+$ rather than $[He\text{-}He\text{-}H]^+$.

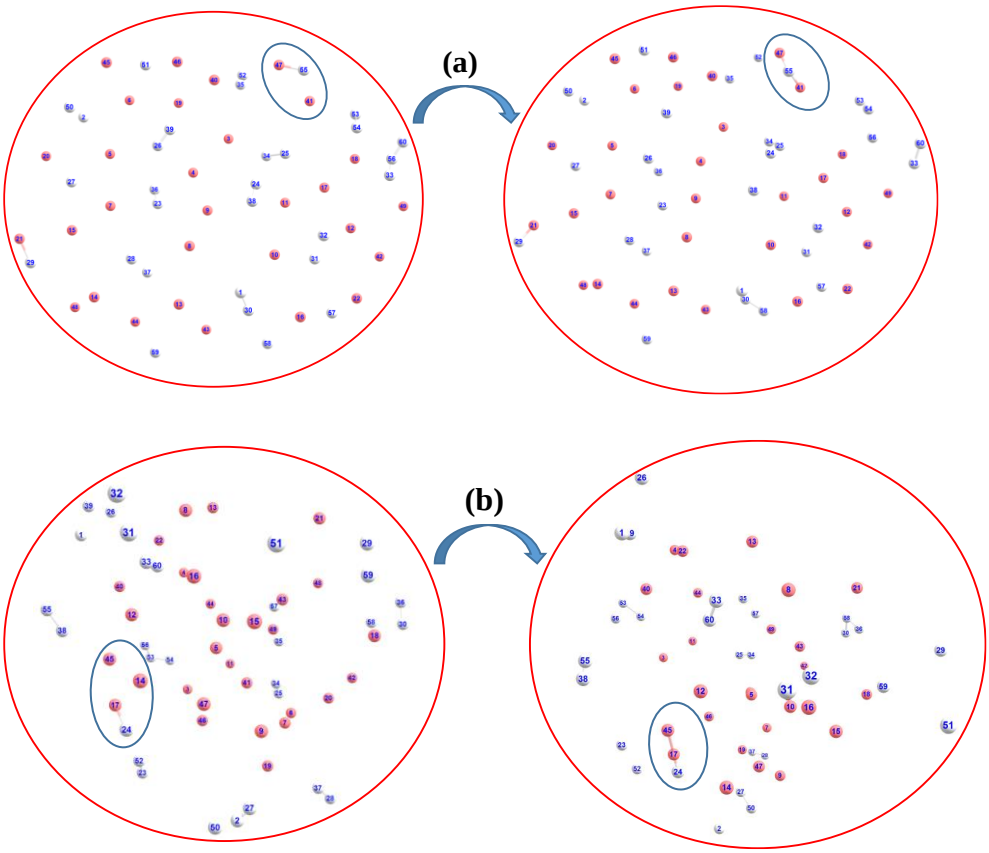


Figure 3.6. Snapshots of *AINR* simulations revealing the pathway towards the making of He_2H^+, in the form of **(a)** $[He\text{-}H\text{-}He]^+$ and **(b)** $[He\text{-}He\text{-}H]^+$ after 15 fs timescale.

In another set of simulations, we have taken a different ratio of helium to dihydrogen and simultaneously varied the total charge of the system. In these cases, due to the high charge density, the dihydrogen quickly dissociates into a proton and atomic hydrogen. Here too, we have observed similar trends: (i) HeH^+ is the first molecule to be formed, followed by (ii) the formation of other short lived species, leading to H_3^+, which remains at the end, along with one or two molecules of HeH^+ (shown in Table 3.6). For the case of 20 He and 5 H_2 having a total of 8 positive charge in the system, for instance, we observed that after a few collisions there was still one HeH^+ molecule roaming about with one H_3^+ and that they were in equilibrium with each other, due to the instantaneous proton transfer between HeH^+ and H_2. Similar trends were observed for other simulations where the total positive charge of the system was high (in our simulation conditions, the values were ≥ 16). It is worth mentioning that in this high positive charge atmosphere with comparatively low H atom density, the number of H_3^+ that survive after the collisions is either one or two depending upon the ratio of He to H (shown in Table 3.6). Also, due to very high positive charge density and high temperature (1000.0 K) the movement of the light H^+ ions was seen to be extremely fast and they repelled each other, going far away. This reduces the propensity towards the formation of H_3^+ in such simulations.

Table 3.6 Different Ratio of He to H while varying total positive charge of the system.

No. of He	No. of H	Total charge	First Molecule	Intermediate Species	Dominating end Molecule	No. of H_3^+
20	10	8	HeH^+	He_2^{2+}, H^+	HeH^+, H_3^+	1
30	20	16	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+	HeH^+, H_3^+	1
30	30	24	HeH^+	He_2^{2+}, He_2H^+, H^+	HeH^+, H_3^+	1
30	30	26	HeH^+	He_2^{2+}, He_2H^+, H^+	HeH^+, H_3^+	1
15	20	8	HeH^+	He_2^{2+}, He_2H^+, H^+	H_3^+	4
30	30	20	HeH^+	He_2^{2+}, He_3^{2+}, He_2H^+, H^+	HeH^+, H_3^+	4

Formation of dicationic He chain. Previously, there have been some reports[45-46] with regard to the formation of an He ion cluster, where He was present as a monocationic species. Our current *AINR* based dynamics study reveals that there is a possibility of formation of a dicationic He

chain of up to five He atoms: He_3^{2+}, He_4^{2+} and He_5^{2+}. In one of the simulations, we have taken a homogeneous mixture of H and He (15 atoms each) with overall positive charge of 20 for the system. After a certain amount of time had elapsed (1ps), we observed that a chain like structure had formed comprising of up to five helium atoms (shown in Figure 3.7).

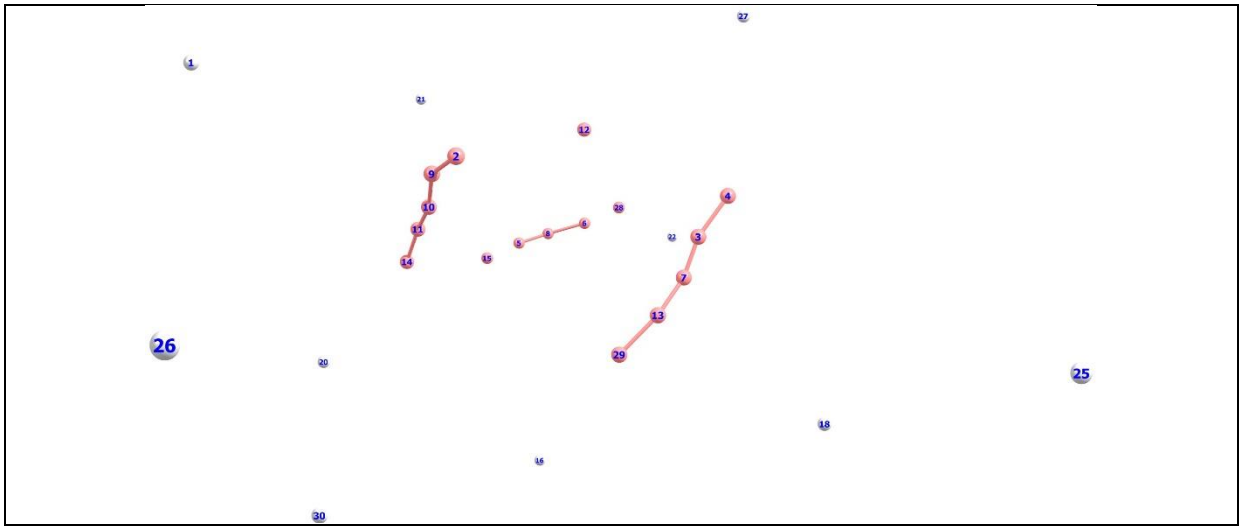
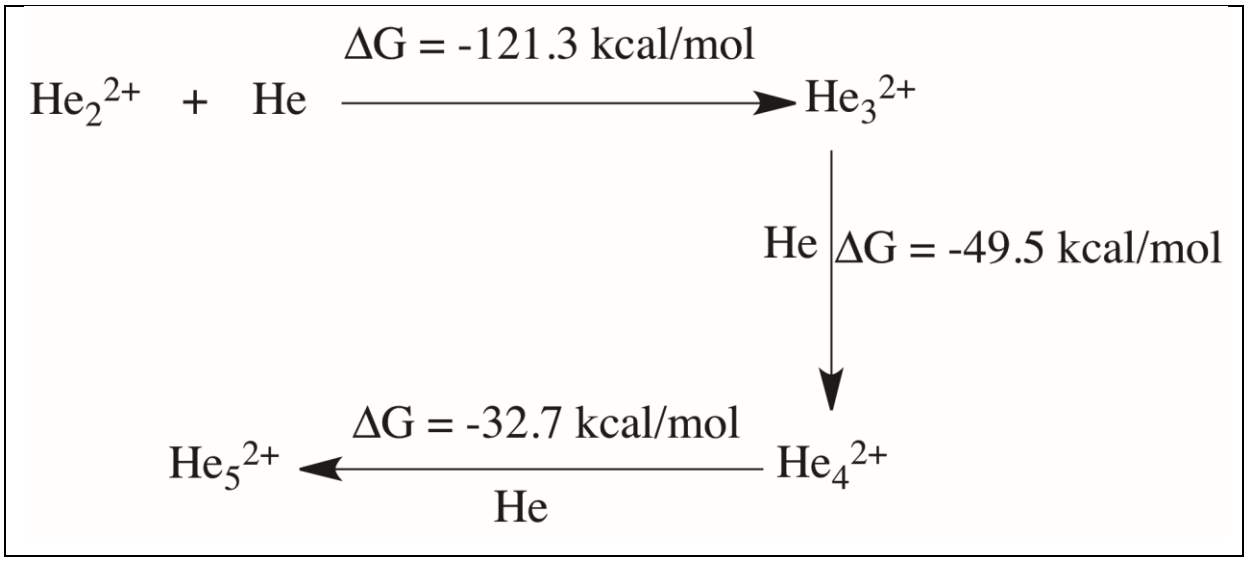


Figure 3.7 Dicationic He chain formation during the *AINR* simulation of 15 H and 15 He with an overall positive charge of 20. The snapshot has taken after 70 fs timrscale.

We have taken the snapshot of the dynamics frame and carried out natural population analysis (NPA) to calculate the charge on the He atoms in the formed linear chain. From the NPA charge analysis, it has been confirmed that all the He chains (He_3^{2+}, He_4^{2+}, He_5^{2+}) are dicationic in nature. For further confirmation of the stability of these dicationic He chains we have done thermodynamics calculations for the formation of the He chain starting from He_2^{2+} (shown in Scheme 3.2). The Gibbs free energy values suggest that the formation of the dicationic He chain up to He_5^{2+} is favourable, but further formation of He_6^{2+} is thermodynamically not feasible. For this reason, we did not observe any He chain beyond five He atoms in our *AINR* simulations.



Scheme 3.2: Pathways of He chain formation upto He_5^{2+}. Values have been calculated at the CCSD/6-311++g(d,p) levels of theory in kcal/mol.

3.4 Conclusions

In our current work, we have tried to shed light on the chemical reactions that might have taken place at the beginning of the universe. We have focused on how, in the very beginning, simple molecules came into being after the Big Bang. We have investigated how He and H atoms, which were the first atoms formed, collided with each other in a positively charged atmosphere. This has been done by using a fresh computational approach – by employing the *ab initio* nanoreactor (*AINR*). The simulations reveal the presence of unique dicationic helium chains of up to 5 atoms, which should act as a fillip for investigating the possibility of the presence of such species in helium clusters, which have recently received attention both from experimental and theoretical studies.[46,50] Our studies also confirm that HeH^+ was indeed the first molecule to be formed and that it played a vital role in the origin of H_3^+. The preservation of H_3^+, as a relatively stable species, in each of the simulations after every collision cycle, also explains the high abundance of H_3^+ in the early universe. As such, our work provides interesting computational insights into the origin of unique and interesting molecules at the dawn of the universe.

3.5 References:

1. Meyer, B. S. *ACS Symp. Series.* 2008, *981* (3), 39-60.
2. Gusten, R.; Wiesemeyer, H.; Neufeld, D.; Menten, K. M.; Garf, U. U.; Jacobs, K.; Klein, B.; Ricken, O.; Risacher, C.; Stutzki, J. *Nature* 2019, *568*, 357–359.
3. Hogness, T. R.; Lunn, E. G. *Phys. Rev*. 1925, *26*, 44–55.
4. (a) Sir Thomson, J. J. *Philos. Mag.* 1911, 6 (21), 225; (b) Sir Thomson, J. J. *Philos. Mag.* 1912, 6 (24), 209.
5. Oka, T. *Phsy. Rev. Lett.* 1980, *45 (7)*, 531.
6. Drossart, P.; Maillard, J. P.; Caldwell, J.; Kim, S. J.; Watson, J. K. G.; Majewski, W. A.; Tennyson, J.; Miller, S.; Atreya, S. K.; Clarke, J.T.; Waite Jr., J.H.; Wagener, R. *Nature* 1989, *340*, 539-541.
7. Watson, W. D. *The Astrophys. J.* 1973, *183*, L17-L20.
8. Oka, T. *Chem. Rev*. 2013, *113*, 8738–8761.
9. Olah, G. A.; Mathew, T.; Surya Prakash G.K. *J. Am. Chem. Soc.* 2016, *138*, 22, 6905-6911.
10. Pelley, J. *ACS Cent. Sci.* 2019, *5*, 5, 741–744.
11. Townsend, D.; Lahankar, S. A.; Lee, S. K.; Chambreau, S. D.; Suits, A. G.; Zang, X.; Rheinecker, J.; Hardings, L. B.; Bowman, J. M. *Science* 2004, *306*, 1158-1161.
12. Okino, T.; Furukawa, Y.; Liu, P.; Ichikawa, T.; Itakura, R., Hoshina, K.; Yamanouchi, K.; Nakano, H. *Chem. Phys. Lett.* 2006, *419*, 223-227.
13. De, S.; Rajput, J.; Roy, A.; Ghosh, P. N.; Safvan, C.P. *Phys. Rev. Lett.* 2006, *97*, 1–4.
14. Mebel, A. M.; Bandrauk, A. D. *J. Chem. Phys.* 2008, *129*, 224311.
15. Kushawaha, R. K.; Bapat, B. 2008, *463*, 42–46.
16. Nakai, K.; Kato, T.; Kono, H.; Yamanouchi, K. Communication: *J. Chem. Phys.* 2013, *139*, 1–5.
17. Ando, T.; Shimamoto, A.; Miura, S.; Iwasaki, A.; Nakai, K.; Yamanouchi, K. *Comm. Chem.* 2018, *1*, 7.
18. Ekanayake, N.; Severt, T.; Nairat, M.; Weingartz, N. P.; Farris, B. M.; Kaderiya, B.; Feizollah, P.; Jochim, B.; Ziaee, F.; Borne, K.; Kanaka Raju, P.; Carnes, K. D.; Rolles, D.; Rudenko, A.; Levine, B. G.; Jackson, J. E.; Ben-Itzhak, I.; Dantus, M. *Nature Comm.* 2018, *9*, 5186.

19. Palaudoux, J.; Hochlaf, M. *ACS Earth Space Chem.* 2019, *3*, 6, 980–985.
20. Fortenberry, R. C. *Chem* **2019**, *5*, 1012-1030.
21. Bovino, S.; Gianturco, F. A.; Tacconi, M. *Chem. Phys. Lett.* 2011, *554*, 47-52.
22. Razio, D. D. *Phys. Chem. Chem. Phys.* 2014, *16*, 11662-11672.
23. Esposito, F.; Coppola, C. M.; Fazio, D. D. *J. Phys. Chem.A* 2015, *119*, 51, 12615–12626.
24. González-Lezana, T.; Bossion, D.; Scribano, Y.; Bhowmick, S.; Suleimanov, Y. V. *J. Phys. Chem. A* 2019, *123*, 49, 10480–10489.
25. McLaughlin, D. R.; Thompson, D. L. *J. Chem. Phys.* 1973, *59*, 4393-4405.
26. Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. *Nat. Chem.* **2014**, *6*, 1044-1048.
27. Zimmerman, P. M. *J. Comput. Chem.* **2013**, *34*, 1385– 1392.
28. Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, *J. Chem. Theory Comput.* **2014**, *10*, 897– 907.
29. Ufimtsev, I. S.; Martinez, T. J., *J. Chem. Theory Comput*. 2009, *5*, 10, 2619−2628.
30. Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.* 2011, *2*, 14, 1789−1793.
31. Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*. 2011, *7*, 6, 1814−1823.
32. Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Chem. Theory Comput*. **2013**, *9*, 1, 213−221.
33. Ufimtsev, I. S.; Martinez, T. J. *Comput. Sci. Eng.***2008**, *10*, 26−34.
34. Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*. **2008**, *4*, 2, 222−231.
35. Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*. **2009**, *5*, 4, 1004−1015.
36. Fischer, C. F. *Comput. Phys. Comm.***1987**, *43*, 355−365.
37. Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.***1980**, *102*, 3, 939−947.
38. Feller, D.; Peterson, K. A. *J. Chem. Phys.***1998**, *108*, 154−176.
39. Das, T.; Ghule, S.; Vanka. K. *ACS Central Science***2019**, *5*, 9, 1532-1540.
40. Hu, X.; Yang, W. *J. Chem. Phys*. **2010**, *132*, 054109.
41. Pulay, P. *Chem. Phys. Lett.***1980**, *73*, 393−398.
42. Grotendorst, J.; Blugel, S.; Marx, D. *Computational Nanoscience*, **2006**, *31*, 245-278.
43. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.***1980**, *72*, 5639-5648.

44. Gaussian 09, Revision E.01; Frisch , M. J.; Trucks, G. W.; Schlegel, H. B.;Scuseria,G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.;Jr.,Peralta, J. E.; Ogliaro, F.,Bearpark, M.; Heyd, J. J. E.; Brothers,Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Raghavachari, J.K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

45. Oleksy, K.; Karlicky, F.; Kalus, R. *J. Chem. Phys.* **2010**, *133*, 164314.

46. Marinetti, F.; Bodo, E.; Gianturco, F. A.; Yurtzever, E. *ChemPhysChem* **2008**, *9*, 2618-2624.

47. Novotný, Oldřich, et al. *Science* 365.6454 (2019): 676-679.

48. Kim, Soon Tai, and Jae Shin Lee. *The Journal of chemical physics* 110.9 (1999): 4413-4418.

49. Pilling, S.; Andrade, D. P. P.; Neves, R.; Ferreira-Rodrigues, A. M.; Santos, A. C. F.; Boechat-Roberty, H. M. Mon. Not. R. Astron. Soc. 2007, 375 (4), 1488−1494.

50. Bieske, Evan J., and Otto Dopfer. *Chemical Reviews* 100.11 (2000): 3963-3998.

Chapter 4

Insights Into the Origin of Life: Did life begin from HCN and H_2O?

–An *ab initio* Nanoreactor Dynamics Approach

Chapter 4

Insights Into the Origin of Life: Did life begin from HCN and H_2O? –An *ab initio* Nanoreactor Dynamics Approach

Abstract

The seminal Urey−Miller experiments showed that molecules crucial to life such as HCN could have formed in the reducing atmosphere of the Hadean Earth and then dissolved in the oceans. Subsequent proponents of the "RNA World" hypothesis have shown aqueous HCN to be the starting point for the formation of the precursors of RNA and proteins. However, the conditions of early Earth suggest that aqueous HCN would have had to react under a significant number of constraints. Therefore, given the limiting conditions, could RNA and protein precursors still have formed from aqueous HCN? If so, what mechanistic routes would have been followed? The current computational study, with the aid of the *ab initio* nanoreactor (AINR), a powerful new tool in computational chemistry, addresses these crucial questions.

4.1 Introduction

How life originated[1–13] on Earth is one of the most fundamental questions of science, and has generated considerable interest. Research and discussion has resulted in two principal positions that are held today: the "RNA World" hypothesis[14–19] and the "metabolism-first" principle.[20–23] According to the RNA World hypothesis, life on Earth originated from the self-replicating molecules of ribonucleic acid (RNA),[24–26] which is the polymeric form of activated ribonucleotides.[27–29] The metabolism-first principle argues, on the other hand, that simple metal catalysts were present in the water in early Earth and aided in creating a soup of organic building blocks that subsequently formed the biomolecules necessary for life. The RNA World hypothesis has gained increased acceptance in recent times, with several experimental studies[26–28,30–35] indicating how hydrogen cyanide (HCN), known to exist on prebiotic Earth, could have been the starting point of many synthetic routes leading to the formation of RNA and protein precursors (see Figure 4.1A).

However, questions remain as to how HCN could have actually functioned in prebiotic conditions. As the famous Urey−Miller experiments have shown, HCN would have formed in the reducing atmosphere that existed during prebiotic times,[36] after which it would have condensed into the oceans.[13,37] HCN has a low boiling point, but at high pH (8−10), it is possible for it to exist in aqueous solution, even if the temperature of the water is 80.0−100.0 °C. However, since the hazy atmosphere[38] of the Hadean Earth would have made it difficult for high-energy photons to reach the Earth's surface (much like the red surface of Titan today, because of a similar haziness in the atmosphere), a lot of the reactions shown in Figure 4.1 A, which depend upon photochemistry or an electric spark, may not have been possible for aqueous HCN. Hence, the more plausible alternative would have been thermochemistry. It is possible that temperatures at the surface of the water bodies of early Earth (3.5−4.0 billion years ago) would have been about 80.0−100.0 °C,[39] which suggest favorable conditions for thermochemistry, but if thermochemistry predominated in the oceans of early Earth, it could be argued that hydrolysis would have taken precedence over the polymerization of HCN. This is because HCN polymerization would have had to begin with HCN dimerization and the subsequent reaction of the product with more HCN molecules. In other words, the polymerization of HCN would have required a series of second-order reactions in HCN, while the competing hydrolysis of HCN

would have simply required the HCN collision with the surrounding solvent water molecules. Indeed, previous studies[34,40] have shown that in dilute aqueous concentrations of HCN, hydrolysis is favored over oligomerization.

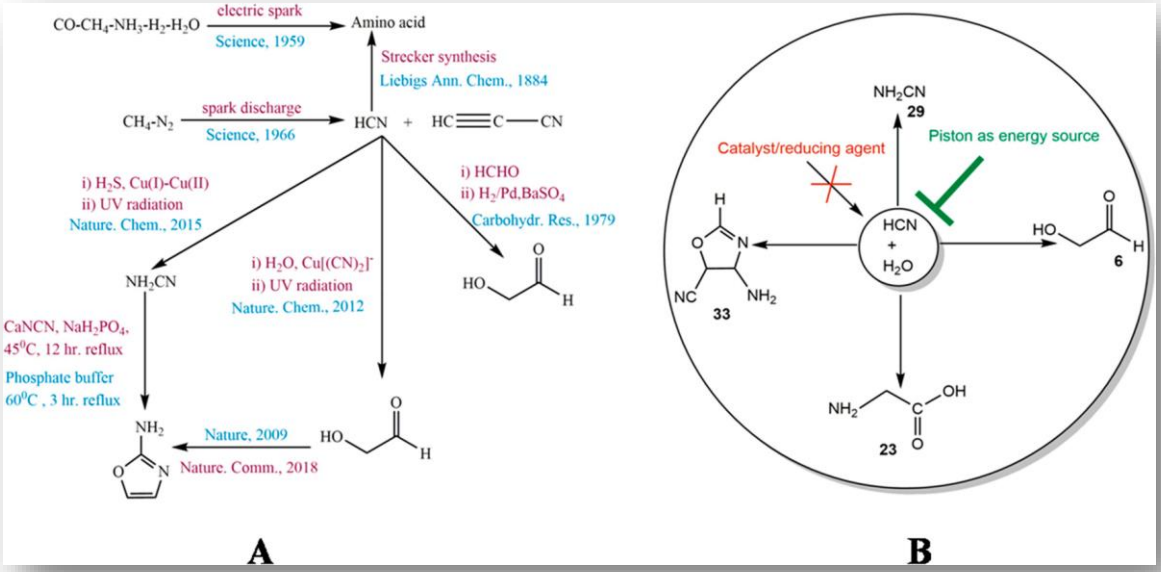


Figure 4.1 (A) Previously synthesized RNA and protein precursors (amino acids, cyanoacetylene, cyanamide, glycoaldehyde, and 2-amino-oxazole). (B) The *ab initio* nanoreactor (AINR) approach, yielding RNA and protein precursors, beginning from only two different reacting molecules, HCN and H_2O, obtained in "one-pot", under the same reaction conditions.

Then, there is also the issue of too-high temperatures: experiments have shown[41,42] that at temperatures above 100.0 °C, decomposition of the formed RNA and protein precursors would occur. Therefore, the reactions would have had to happen around 100.0 °C.[43,44]

Hence, for the RNA World hypothesis to be true, there are several constraints that have to be kept in mind: (i) thermal, not photochemical conditions, (ii) reactions where monomeric and not polymeric HCN would predominate, (iii) without mediation from metal catalysts, (iv) at temperatures not exceeding 100.0 °C, and (v) having reactions with free energy barriers not exceeding 40.0 kcal/mol. To this list, one could add (vi) the need to avoid chemical processes involving the protonation of substrates, since HCN, with a pKa of 9.31, is a weak acid and would have largely remained in undissociated form in solution. The protonation of water would also

have been suppressed since the pH of water has been estimated to be between 8.0 and 9.0 in prebiotic times at the surface of the ocean.

This list of conditions appears formidable and leads to the important question: could life have begun under these circumstances? This current work attempts to answer this question, through the agency of the *ab initio* nanoreactor (AINR).

The AINR method, recently developed,[45] allows one to obtain reaction pathways and products without controlling the chemical system in any way.[46,47] This represents a major shift in what one can do with computational chemistry, because, using the AINR, one can now discover new reactions, completely independent of experimental input. This was demonstrated by Martinez and co-workers[45] when they found entirely plausible, new pathways for the formation of amino acids, from a computational re-enactment of the Urey−Miller experiment.[48] In the current work, we have conducted full quantum mechanical molecular dynamics (MD) simulations on systems employing the AINR approach on systems containing a mixture of molecules of HCN and H_2O. The goal has been to follow the chemical reactions that can occur through collisions between the molecules and observe what new species are formed as a result. In short, our objective has been to perform the equivalent of an experimental study while satisfying the conditions outlined in (i−vi) above. Remarkably, we have found, as will be shown in the Results and Discussion, that just the interaction of HCN and H_2O was sufficient to eventually lead to the formation of the experimentally reported precursor molecules to RNA and proteins: cyanamide,[27,30−32,49,50] glycolaldehyde,[27,30,31,51−56] an oxazole derivative,[27,57-58] and the amino acid glycine[36,53,59-63] (as shown in Figure 4.1B).

4.2 Computational Methods

4.2.1 *Ab Initio* Molecular Dynamics (AIMD) Simulations. The nanoreactor AIMD simulations were performed with the TeraChem 1.9 software package[64-70] where Born−Oppenheimer potential energy surface by using the Hartree−Fock (HF)[71] electronic wave function and the 3-21g(d) Gaussian basis set[72]. This method has been implemented in TeraChem by Martinez and co-workers.[45] This approach was deemed acceptable because the HF method is well-known for predicting chemically reasonable structures.[73] Also, it should be noted HF was not employed to determine barrier heights and reaction rates: its only role was in the discovery process. This was

also the approach employed by Martinez and coworkers in their original AINR paper (employing HF/3-21g), where they replicated the results obtained from the Urey−Miller experiment, as well as from the interaction of acetylene molecules.[45] We note here that we did also attempt discovery in the AINR simulations with DFT using the B3LYP density functional and the 3-21g(d) basis set and did find the preliminary intermediates (formamide, formic acid, formaldehyde and others) through this approach as with HF/3-21g(d). However, this was at greater computational expense and did not appear to give different results from the HF/3-21g(d) approach. Hence, we have limited the discovery process to HF/3-21g(d) in the AINR simulations. The AINR simulation results that have been discussed here pertain to the case in which 16 H_2O and 15 HCN molecules were taken together in a spherical box of radii 10.0 and 3.5 Å (the system alternated between the two radii, in order for the collisions to take place - see original paper by Martinez and co-workers[45]). This system was allowed to evolve for 750 ps and generated the intermediates and reaction pathways that have been discussed. Additionally, we have also performed several AINR simulations where we varied different parameters, such as (i) the ratio of the reactant species, (ii) the total number of molecules taken in the simulation box, (iii) the spherical boundary conditions, (iv) the temperature, and (v) the total time of the AIMD simulations. The results obtained by changing the parameters (i−v) have been discussed in the Results and Discussions section. In general, they indicate that while most of the intermediates were discovered by varying (i) to (v), the most comprehensive results were obtained from the simulation case described here: taking 16 H_2O and 15 HCN molecules in a spherical box of radii 10.0 and 3.5 Å. Moreover, for this case, multiple simulations were also performed from the same initial configuration and were seen to give rise to all the desired intermediates and products, although the time of formation of these species during the simulations was seen to change from simulation to simulation. Langevin dynamics was employed to calculate Newton's equations of motion with an equilibrium temperature of 2000.0 K (also the starting temperature of the dynamics). We have used this high temperature in order to increase the average kinetic energy of the reactant molecules and for faster dynamics, allowing the overcoming of noncovalent interactions without the breaking of covalent bonds. This, too, follows the example of the work with the AINR done by Martinez and coworkers.[45] The nanoreactor simulations employ a piston to accelerate the reaction rate. We have employed the augmented direct inversion in the iterative subspace (ADIIS) algorithm[74] available in TeraChem as an alternative tool for self-consistent field calculations at each AIMD

step in which the default DIIS algorithm[75] failed to converge. Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event.

4.2.2 AINR Spherical Boundary Conditions: Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event. The spherical boundary conditions were applied in the form of a sum of two harmonic terms. The molecules were restricted to move inside a spherical volume by a boundary potential, with a time-dependent component :

$$V(r,t)= f(t)U(r,r_1,k_1) +(1- f(t)) U(r,r_2,k_2)$$

$$U(r,r_0,k) = mk/2\ (r - r_0)^2\theta(r - r_0);\ f(t) = \theta(\lfloor t/T \rfloor - t/T + \tau/T)$$

Where k_1=1.0 kcal mol^{-1} $Å^{-2}$, r_1 =10.0 Å, k_2=0.5 kcalmol^{-1} $Å^{-2}$, r_2=3.5Å, τ=1.7ps, T=2.0 ps, ⌊⌋ is representing the floor function and θ is the heaviside step function. For more details regarding the each mathematical terms please check the reference (45).The restraint potential forces the atoms with a radial position 10.0 Å to 3.5 Å towards the center of the sphere and allows them to collide. When the sphere is expanded again, the molecules present in the smaller volume diffuse rapidly (because of the high simulation temperature) to occupy the larger volume. Due to the repeating compression and expansion of spherical volume, the molecules collide and relax. Therefore, throughout the simulation, new molecules are formed and then break again to form other new molecules. We have run simulations with 93 atoms (16 H_2O + 15 HCN) for a total time of 750.0 ps with a timestep of 0.5 fs.

4.3 Results and Discussions

4.3.1 Optimizations of the *ab initio* Nanoreactor (AINR) Simulations: We have done several AINR simulations for optimizing the initial conditions of the simulations, based on the different parameters that can affect the results. The parameters are (i) the ratio of the reactant species, (ii) the number of molecules taken in the simulation box, (iii) spherical boundary conditions, (iv) temperature, and (v) the total time of the AIMD simulations. We now briefly discuss the results obtained from our several simulations based on the tested nanoreactor parameters:

(i) We have done the AINR simulation with a 1:2 mixture of HCN (9 molecules) and H_2O (18 molecules), while setting the other parameters to be k_1 = 1.0 kcal mol^{-1} $Å^{-2}$ (the force constant at the outer boundary), r_1 = 8.5 Å, k_2 = 0.5 kcal mol^{-1} $Å^{-2}$ (the force constant at the inner boundary), r_2 = 3.0 Å, τ = 1.7 ps, Total time between collisions = 2.0 ps. In this simulation, we have found the initial hydrolyzed products of HCN such as formamide, formic acid and formaldehyde, but not the other intermediates species or the RNA and protein precursors. The mechanistic pathways for the formation of the hydrolyzed products are similar to those that we have found in our main AINR simulation. The total time evolution for this simulation is ~700ps.

(ii) In another AINR simulation, we have taken a 2:1 mixture of HCN (18 molecules) and H_2O (9 molecules), keeping all other parameters the same as (i) and running for ~700ps. In this case, instead of hydrolyzed products, we have found more oligomeric products of HCN. Very few formamide molecules were formed, and due to the lack of water molecules, the system did not further lead to the formation of the desired RNA and protein precursor molecules.

(iii) In our simulations, the source of carbon and nitrogen is HCN, and the oxygen source is water. This has influenced our decision to take almost a 1:1 ratio of HCN and H_2O, so that it can maximize the interaction between the two different moieties. In order to check the validity of this approach, we have done two different simulations, where we have taken (11HCN +13H_2O) and (15 HCN +16 H_2O) mixtures and run the AINR simulations for ~750 ps. In both these cases, we have come up with the desired RNA and protein precursors. In our current manuscript, we have reported the results obtained from the (15 HCN +16 H_2O) mixture AINR simulation. Therefore, the results indicate that if one takes a similar number of HCN and H_2O molecules, i.e. in about a 1:1 ratio, it will maximize the probability of getting the desired final products.

(iv) Another important parameter in the AINR simulations is the spherical boundary condition, which we have discussed in the "**AINR Spherical Boundary Conditions**" subsection (4.2.2) in the Computational Details section. Optimizing the boundary in the AINR is a trial and error process. We had to fix the boundary in such a way so that the collision between the molecules in the inner sphere would be effective and the molecules would get enough space to relax in the

outer sphere. Also, the point to be noted is that the collisions should not be at such a high velocity that the molecules breaks into it elemental form.

(v) Temperature is another important parameter in the AINR simulations. The temperature that we have used in our simulations (2000 K) is not the actual reaction temperature at which the reactions would occur. The reason that we have provided such a high temperature is to avoid noncovalent interactions such as hydrogen bonding in our AINR simulation and also to provide enough kinetic energy to the molecules so that they could collide with each other and cross the activation barriers, leading to the products. We reiterate that the purpose of the AINR is to act as a tool for discovery of chemical reactions, the feasibility of which could then be determined with careful, high level QM (DFT, MP2 and CC2) studies of the thermodynamics and kinetics of the discovered reactions. Therefore the parameters (ratio of the reactants, spherical boundary conditions and temperature) that would lead to the best possibility of discovering new processes have to be employed, regardless of whether they necessarily represent the actual experimental conditions or not.

(vi) The goal of the AINR simulations is to find new reactions and mechanistic pathways, and not to equilibrate the systems. We have run most of the simulations at a ~1ns timescale, which sufficed to yield different interesting intermediates and products, as well as the corresponding mechanistic pathways.

The AINR approach makes use of collisions between the molecules of HCN and H_2O, and this provides the energy required to cross the activation barriers for each of the elementary steps of the reactions. The simulations have been done on systems having nearly homogeneously mixed HCN and H_2O molecules as the starting reactants. Sixteen H_2O and 15 HCN molecules were taken together, and the system was allowed to evolve for 750 ps. Collisions between the molecules gave rise to new species. It should be noted that homogeneous mixtures of HCN and H_2O do not represent the exact ratios of HCN and H_2O present in a localized region of the early Earth oceans or hydrothermal vents. The reason such mixtures were employed was to maximize the possibility of interactions between HCN and H_2O in the AINR. This would increase the probability of obtaining different products during the simulations. The goal of the AINR simulations was to obtain mechanistic pathways for the formation of different RNA and protein

precursors beginning from HCN and H_2O. Studying homogeneous mixtures of HCN and H_2O afforded the best possibility of realizing this goal. Figure 4.2 below illustrates how a system starting with a mixture of HCN and H_2O molecules evolves with time. The AINR approach thus leads to the discovery of new species from the starting compounds, next section discusses the results that have been obtained by this approach.

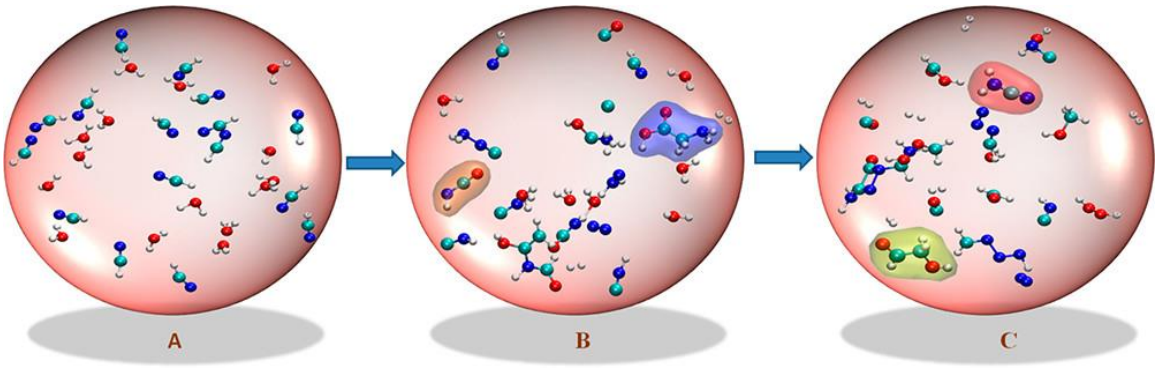


Figure 4.2. Snapshots of AINR simulations. (A) the beginning, 0.0 ps: only HCN and H_2O present. (B) after 100 ps, glycine (blue surface) has formed, along with molecules such as isocyanic acid (pale yellow surface). (C) after 250 ps: glycoaldehyde (green surface) and cyanamide (orange surface) have formed, along with other oligomeric species. Color scheme: oxygen: red, carbon: teal, hydrogen: gray and nitrogen: blue.

During the AINR simulation a selection of the acyclic and heterocyclic products that were discovered from the *ab initio* nanoreactor simulations, including ribonucleotide and amino acid precursors and other intermediate compounds discussed in previous section. We have shown that apart from the important precursors for RNA and protein, a lot of diverse acyclic organic compounds (shown in figure 4.3) and heterocyclic organic species (shown in figure 4.4) were also formed during the simulations.

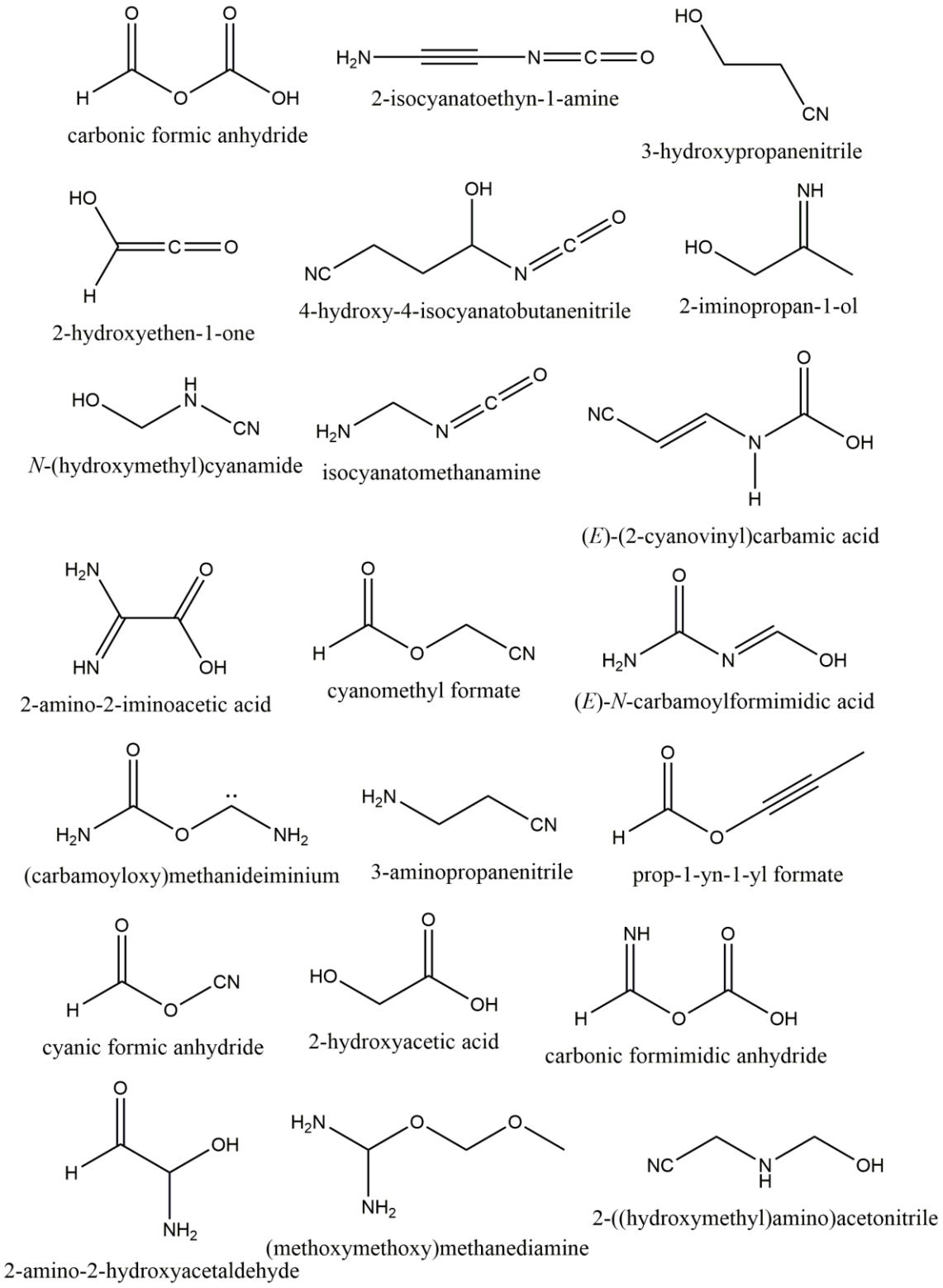
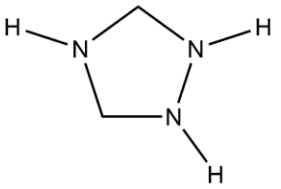
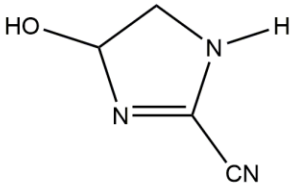


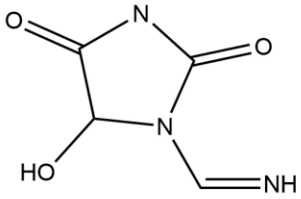
Figure 4.3. Apart from the important precursors for RNA and protein, a lot of diverse acyclic organic compounds were also formed during the simulations.



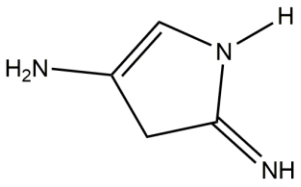
1,2,4-triazolidine



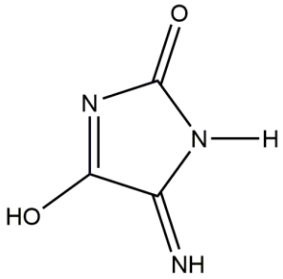
4-hydroxy-4,5-dihydro-1*H*-imidazole-2-carbonitrile



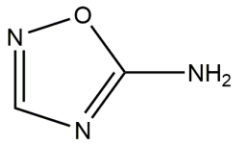
5-hydroxy-1-(iminomethyl)-3λ^2-imidazolidine-2,4-dione



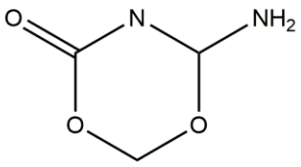
5-amino-4,5-dihydro-1*H*-pyrrol-3-amine



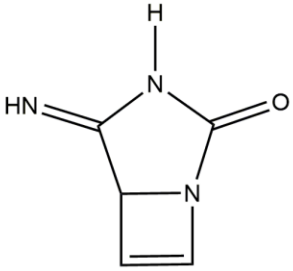
4-hydroxy-5-imino-1,5-dihydro-2*H*-imidazol-2-one



1,2,4-oxadiazol-5-amine



6-amino-1,3,5λ^2-dioxazinan-4-one



4-imino-1,3-diazabicyclo[3.2.0]hept-6-en-2-one

Figure 4.4. Apart from the oxazole (ribonucletide precursor), a lot of diverse cyclic organic compounds were also formed during the simulations.

4.4 Conclusions

The current work shows that interaction between only two different molecules- HCN and H_2O - would have been sufficient to give rise to most of the important precursors to RNA and proteins in prebiotic times. Taking advantage of the recently developed AINR method,[45] which has allowed us to discover new reaction pathways, we have shown that cyanamide, glycoladehyde, oxazole derivative, and glycine all could have been formed from only a single carbon and nitrogen source molecule: HCN and a single oxygen source molecule: H_2O. HCN, as has been noted in the literature, may have occupied a "unique position in terrestrial pre-biological chemistry". [40] The current work shows that just the interaction between HCN and water as the starting reactants would have been sufficient to eventually lead to the precursors of RNA and proteins. This is significant because it shows that the reactions could have happened ubiquitously in the water bodies all over the Earth. What is also important is that all the conditions specified as (i−vi) earlier were adhered to during the simulations.

4.5 References:

(1) Gańti,T. *Oxford Univsersity Press: Oxford UK*, **2003**.

(2) Dyson, F. *Cambridge University Press: Cambridge UK*, **1999**.

(3) Sutherland, D. *Nat. Rev*., **2017**, *1*, 1−7.

(4) Schopf, J. W. *University of California Press: Oakland*, CA, **2002**.

(5) Steel, M.; Penny, D., *Nature*, **2010**, *465*, 168−169.

(6) Oparin, A. I. *World Publishing: Cleveland*, **2003**.

(7) Courier, D.; Oro, J. *World Publishing: Cleveland*, **2002**.

(8) Bernal, J. D. *World Publishing: Cleveland*, **1967**.

(9) Sutherland, J. D. *Angew. Chem., Int. Ed*., **2016**, *55*, 104−121.

(10) Bracher, P. J. *Nat. Chem.*, **2015**, *7*, 273−274.

(11) Pascal, R.; Pross, A.; Sutherland, J. D. *Open Biol.*, **2013**, *3*, 130156−130164.

(12) Bada, J. L. *Earth Planet. Sci. Lett*., **2004**, *226*, 1−15.

(13) Bada, J. L. *Chem. Soc. Rev*., **2013**, *42*, 2186− 2196.

(14) Woese, C.; *Harper & Row*, **1967**, pp 179−195

(15) Crick, F. H. C. *J. Mol. Biol*., **1968**, *38*, 367−379.

(16) Orgel, L. E. *J. Mol. Biol*., **1968**, *38*, 381−393.

(17) Joyce, G. F. *Nature*., **2002**, *418*, 214−221.

(18) Joyce, G. F.; Orgel, L. E.; Gesteland, R. F. Cech; T. R.; Atkins, J. F., *Eds.; Cold Spring Harbor Laboratory Press: New York US, 2006*, pp 23−56.

(19) Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.,* **2004**, *39*, 99−123.

(20) Schmidt, S.; Sunyaev, S.; Bork, P.; Dandekar, T. *Trends Biochem. Sci.* **2003**, *28*, 336−341.

(21) Caetano-Anolles, G.; Kim, H. S.; Mittenthal, J. A. *Proc. Natl. Acad. Sci., U. S. A***. 2007**, *104*, 9358− 9363.

(22) Copley, R. R.; Bork, P. *J. Mol. Biol*., **2000**, *303*, 627−640.

(23) Danchin, A. Beilstein *J. Org. Chem*., **2017**, *13*, 1119−1135.

(24) Ferris, J. P.; Hill, A. R.; Liu, R., Jr; Orgel, L. E. *Nature,* **1996**, *381*, 59−61.

(25) Verlander, M. S.; Lohrmann, R.; Orgel, L. E. *J. Mol. Evol.,* **1973**, *2*, 303−316.

(26) Szostak, J. W. *Nature* 2009, 459, 171−172.

(27) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature,* **2009**, *459*, 239−242.

(28) Ritson, D. J.; Sutherland, J. D. *Angew. Chem., Int. Ed.*, **2013**, *52*, 5845−5847.

(29) Bowler, F. R.; Chan, C. K.; Duffy, C. D.; Gerland, B.; Islam, S.; Powner, M. W.; Sutherland, J. D.; Xu, J. *Nat.Chem*., **2013**, *5*, 383−389.

(30) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. *Nat. Chem*., **2015**, *7*, 301−307.

(31) Ritson, D. J.; Battilocchio, C.; Ley, S. V.; Sutherland, J. D. *Nat. Commun.,* **2018**, *9*, 1821−1830.

(32) Anastasi, C.; Buchet, F. F.; Crowe, M. A.; Helliwell, M.; Raftery, J.; Sutherland, J. D. *Chem. - Eur. J.*, **2008**, *14*, 2375−2388.

(33) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *Science*, **1966**, *154*, 784−785.

(34) Ferris, J. P.; Hagan, W. J., Jr. *Tetrahedron*, **1984**, *40*, 1093−1120.

(35) Schlesinger, G.; Miller, S. L. *J. Mol. Evol.*, **1983**, *19*, 383−390.

(36) Miller, S. L. *Science*, **1953**, *117*, 528−529.

(37) Bada, J. L.; Cleaves, H. J. *Proc.Natl.Acad.Sci.U.S.A.*, **2015**, *112*, E342.

(38) Arney, N. G; Domagal-Goldman, D. S; Meadows, S. V. Earth and Planetary *Astrophysics*, **2017**, *arXiv:1711.01675.*

(39) Kasting, J. F.; Ackerman, T. P. *Science*, **1986**, *234*, 1383−1385.

(40) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *J. Mol. Biol.*, **1967**, *30*, 223−253.

(41) Aubrey, A. D.; Cleaves, H. J.; Bada, J. L. *Origins Life Evol. Biospheres*, **2009**, *39*, 91−108.

(42)White,R.H. *Nature*, **1984**, *310*, 430−432.

(43) Miller, S. L.; Bada, J. L., *Nature*, **1988**, *334*, 609−611.

(44) Szori, M.; Jojart, B.; Izsak, R.; Szori, K.; Csizmadia, I. G.; Viskolcz, B. *Phys. Chem. Chem. Phys.*, **2011**, *13*, 7449−7458.

45. Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J., *Nat. Chem.*, **2014,** *6*, 1044−1048.

46. Zimmerman, P. M., *J. Comput. Chem.*, **2013**, *34*, 1385−1392.

47. Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. *J. Chem. Theory Comput.*, **2014**, *10*, 897−907.

48. Miller, S. L.; Urey, H. C. *Science*, **1959**, *130*, 245−251.

49. Turner, B. E.; Liszt, H. S.; Kaifu, N.; Kisliakov, A. G. *Astrophys. J.* **1975**, *201*, L149−L152.

50. Lohrmann, R. *J. Mol. Evol.*, **1972**, *1*, 263−269.

51. Decker, P.; Schweer, H.; Pohlamnn, R. Bioids: X. J. *Chromatogr. A*, **1982**, *244*, 281−291.

52. Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A., *Science*, **2004**, *303*, 196.

53. Serianni, A. S.; Clark, E. L.; Barker, R. *Carbohydr. Res.*, **1979**, *72*, 79−91.

54. Fischer, E. *Ber. Dtsch. Chem. Ges.*, **1889**, *22*, 2204−2205.

55. Ritson, D.; Sutherland, J. D. *Nat. Chem.*, **2012**, *4*, 895−899.

56. Benner, S. A.; Kim, H. J.; Carrigan, M. A. *Acc. Chem. Res.*, **2012**, *45*, 2025−2034.

57. Cockerill, A. F.; Deacon, A.; Harrison, R. G.; Osborne, D. J.; Prime, D. M.; Ross, W. J.; Todd, A.; Verge, J. P. *Synthesis*, **1976**, *76*, 591−593.

58. Eschenmoser, A.; Loewenthal, E. *Chem. Soc. Rev*., **1992**, *21*, 1−16.

59. Strecker, A. *Liebigs Ann. Chem*., **1854**, *91*, 349− 351.

60. Bar-Nun, A.; Bar-Nun, N.; Bauer, S. H.; Sagan, C., *Science*, **1970**, *168*, 470−473.

61. Matthews, C. N.; Moser, R. E., *Nature*, **1967**, *215*, 1230−1234.

62. Goldman, N.; Reed, E. J.; Fried, L. E.; William Kuo, I.-F.; Maiti, A. *Nat. Chem.*, **2010**, *2* , 949−954.

63. Saitta, A. M.; Saija, F. *Proc. Natl. Acad. Sci. U. S. A.*, **2014**, *111*, 13768−13773.

(64) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.*, **2009**, *5*, 2619−2628.

(65) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.*, **2011**, *2*, 1789−1793.

(66) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.*, **2011**, *7*, 1814−1823.

(67) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Chem. Theory Comput.*, **2013**, *9*, 213−221.

(68) Ufimtsev, I. S.; Martinez, T. *J. Comput. Sci. Eng*., **2008**, *10*, 26−34.

(69) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*., **2008**, *4*, 222−231.

(70) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*., **2009**, *5*, 1004−1015.

(71) Froese, F. C. *Comput. Phys. Commun*., **1987**, *43*, 355−365.

(72) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc*., **1980**, *102*, 939−947.

(73) Feller, D.; Peterson, K. A. *J. Chem. Phys*. **1998**, *108*, 154−176.

(74) Hu, X.; Yang, W. *J. Chem. Phys*. **2010**, *132*, 054109.

(75) Pulay, P. *Chem. Phys. Lett*. **1980**, *73*, 393−398.

Chapter 5

From Messy Chemistry to the Origin of Life: Insights from Complete Quantum Mechanical (QM) Perspective

Chapter 5

From Messy Chemistry to the Origin of Life: Insights from Complete Quantum Mechanical (QM) Perspective

Abstract

Six decades ago the path breaking Miller-Urey experiment produced various amino acid *via* the interaction between small organic molecules and energy source. The chemistry behind the origin of life was not any single reaction rather it was complex event or messy which is not clear. Gratifyingly, not only do the results from the *ab initio* nanoreactor approach show that aqueous HCN could indeed have been the source of RNA and protein precursors, but they also indicate that just the interaction of HCN with water would have sufficed to begin a series of reactions and make a complex reaction network leading to the precursors towards RNA and protein. The current work therefore provides important missing links in the story of prebiotic chemistry and charts the road from aqueous HCN to the precursors of RNA and proteins through messy chemistry.

5.1 Introduction

In certain perspectives, prebiotic science is very much like an engineered science. Around six decades ago by knowing the significance of amino acids as an important life building unit, Stanley Miller[1] prevailing with regards to acquiring some of them, basically performed a chemical reaction by producing an electric release in a gas blend. At that point this combination utilized by Miller is the best possible way to get amino acids under the prebiotically plausible condition. After that people had synthesized several life building blocks by taking the Miller experiment as a benchmark.

In a simple way one can say that, prebiotic chemistry could be depicted as wasteful synthetic chemistry which performed under non ideal conditions. These non ideal conditions is resolved exclusively by their alleged likenesses with the normal conditions influencing the prebiotic time early earth condition early Earth condition. Tragically, these conditions are inadequately portrayed, and moreover, taking into account that the life on Earth was not likely a local event. Moreover, it's not possible for anyone to envision that all the occasions, which prompted the primary living cells, occurred in a one of a kind locus following a linear pathway. Prebiotic chemistry should imitate the pathways by which matter intricacy expanded on the young Earth yet these pathways were trailed by some coincidence; they were not planned and will remain everlastingly obscure. Undoubtedly, no fossils, no follows, nothing stays to validate one pathway concerning others. The prebiotic chemistry expert doesn't replicate something he knows rather he designs potential pathways, which can expand the intricacy of the molecules on which he is working. This is the place where prebiotic chemistry unmistakably varies from synthetic chemistry.

In the previous chapter we have shown that by using AINR interaction between HCN and H_2O leads to the production of RNA and protein precursor molecules. Though the precursors formed successfully through AINR but this is not sufficient to conclude without doing thermodynamics and kinetics studies.

But there is one major issue of too-high temperatures during the prebiotic condition: experiments have shown[2-3] that at temperatures above 100.0 °C, decomposition of the formed RNA and protein precursors would occur. Therefore, the reactions would have had to happen

around 100.0 °C,[4-5] which indicates that the barriers (ΔG values) of the reactions of monomeric HCN in water could have only been about 40.0 kcal/mol: previous computational studies[6-10] have shown that chemical reactions occurring at temperatures of around 100.0 °C have barriers in the region of 40.0 kcal/mol. Barriers higher than 40.0 kcal/ mol would have led to much slower reactions (or no reactions) at 100.0 °C. Slower reactions may be possible at higher barriers, but this would lead to the possibility of other side reactions also becoming competitive and causing significant reduction in the formation of desired products.

The AINR approach shown in the previous chapter leads to the discovery of new species especially the formation of life building blocks from the starting compounds HCN and H_2O. In this current chapter we have shown the analysis of the data through the connectivity graph (shown in Figure 5.1) allows the exploration of new mechanistic pathways. In the results and discussions section discusses the results that have been obtained by this approach. Furthermore, analysis of the data allowed us to determine the mechanistic pathways by which HCN and H_2O reacted together to yield intermediates and, eventually, the RNA and protein precursors. We subsequently subjected these pathways to a full static quantum chemical study with RI-CC2, RI-MP2 as well as density functional theory (DFT) and thus obtained all the barriers ($\Delta G^{\#}$) for the reactions involved in these processes, as well as the energies (ΔG) of the reactions. As will be discussed in the Results and Discussion, this has led to results that not only reveal interesting pathways for the formation of the precursor molecules beginning from aqueous HCN but also indicate that these mechanistic routes would have been thermodynamically and kinetically feasible.

5.2 Computational Methods

5.2.1 Analysis of the AINR trajectory from the *ab initio* Nanoreactor: After the AIMD run, we have analyzed the simulation trajectories. What this involves is the identification of new molecules and the pathways of their formation. This was done by using data analysis and visualization with the Python libraries NetworkX,[11] Numpy[12] and Graphviz.[13] The two-state hidden Markov model[14] (HMM) was employed for this purpose. The description of the HMM model is provided below. We would like to point out, however, that we did not observe any significant improvement in the trajectory analysis tree by applying the HMM model. The most

critical simulation steps to uncover the reaction mechanism are collision steps because new bonds are formed and old ones are broken. Molecules come very close during collision steps. In general, when collision steps are viewed through the visualization software, one would observe many unrealistic bonds between the atoms, making analysis quite difficult. However, in our data, we observed only a slight increase in the number of unrealistic bonds in some collision steps (average increase of ~2 to ~5 bonds), showing that HMM analysis may not be necessary. After analyzing collision steps with HMM, we observed a reduction in unrealistic bonds for most of the steps (average decrease of ~40% to ~50%). It can be seen that 50% of a small number (i.e., ~2 to ~5) is still insignificant to provide us with any real advantage. HMM only added an extra overhead to the analysis as it was one of the slowest steps. Therefore, we did not observe any significant improvement in the trajectory analysis tree by applying the HMM model.Therefore, we have relied more on our Python code as well as on the manual visualization of the simulation trajectories in the Molden[15] and VMD[16] softwares, for determining the best possible routes to the formation of the intermediate molecules observed in the nanoreactor.

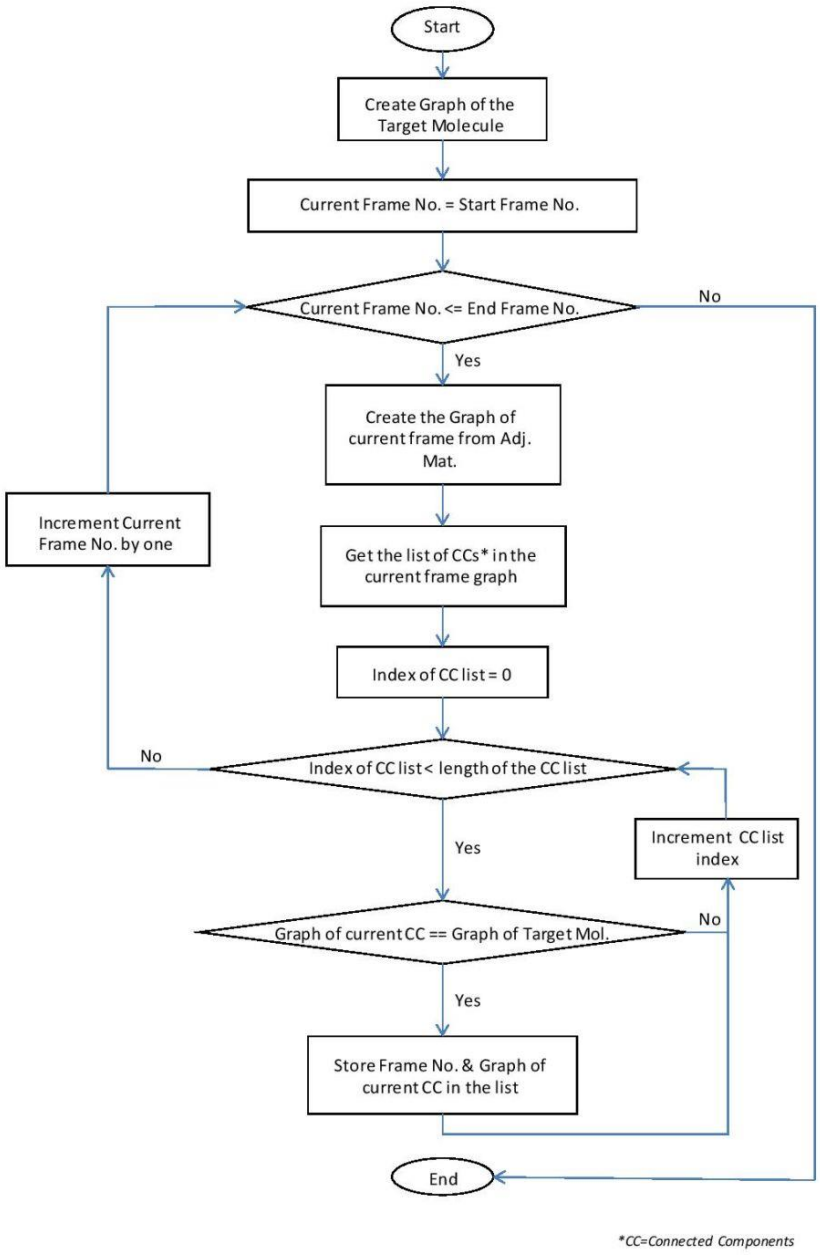
5.2.2 The Hidden Markov Model (HMM):

The hidden Markov model (HMM) is a tool for representing the probability distribution over a sequence of observations. The HMM has two important properties. First, the observations are generated by a process whose states are hidden from the observer. Second, it assumes that the state of this hidden process satisfies the Markov property, i.e., when predicting the future, the past does not matter, only the present. The HMM allows us to predict the most probable sequence of hidden states for the given sequence of observations. It is specified by transition and emission probabilities.

For the current data, produced from the *ab initio* nanoreactor simulations, the hidden and observed states are the same, i.e. 0 and 1 (1 = bond and 0 = no bond between a pair of atoms). These states are elements of a connectivity matrix. We obtain the sequence of observed states from the simulations and predict the hidden states from the HMM. We have employed the *Viterbi* algorithm[17] to find the most probable sequence of the hidden states. The connectivity matrix constructed from these hidden states is supposed to give us improved connectivity between the atoms. The observed sequence from the whole cycle (collision + non-collision steps) was

provided as input to the model. As the collision steps are most complicated to analyze, we have employed the HMM to construct the connectivity matrix of only the collision steps. The transition and emission probabilities were obtained from the original *ab initio* nanoreactor paper by Martinez and co-workers.[18]

Flowchart of the Implementation of the Hidden Markov Model (HMM)



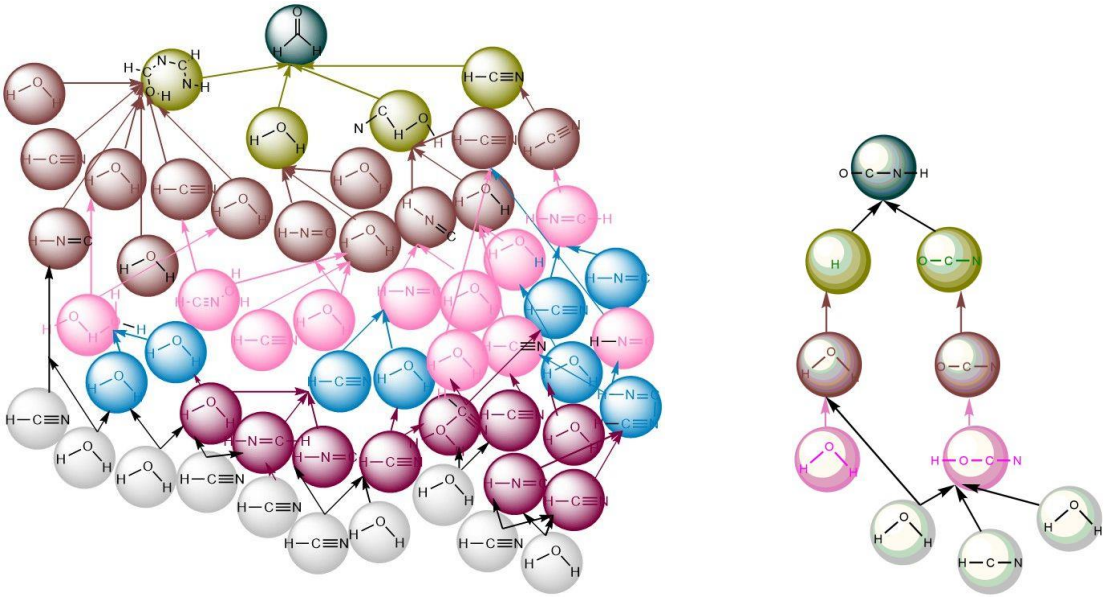


Figure 5.1. The representation of the complex connectivity graph for the formation of formaldehyde (A), and the simplest connectivity graph for the formation of isocyanic acid (B), starting from HCN and H_2O. How the connectivity graph changes after each collision is represented by different colours of the sphere: deep blue represents the target molecule; green, brown, pink are the intermediate species and at the bottom, white represents the starting molecules. For further understanding, the reader is also encouraged to look at the methodology outlined in the paper by Martinez and co-workers.

5.2.3 Quantum Mechanical Calculations: In order to determine the reaction free energy (ΔG) and energy barriers ($\Delta G^{\#}$), we have done the minimum energy pathway (MEP) search by full quantum mechanical calculations, including zero point energy, internal energy, and entropic contributions, with the temperature taken to be 298.15 K. Density functional theory (DFT) have been employed for all the structures calculations. Geometry optimizations and transition state search calculations were carried out with the Turbomole 7.0 software package[19] using the TZVP basis set[20] and the B3LYP three parameter hybrid density functional.[21] Dispersion corrections (D3) were included in all the calculations. Solvent corrections were included with the dielectric continuum solvent model COSMO[22] with $\varepsilon = 80.0$. Furthermore, in order to make our data more reliable and also to refine the energies, the single point energy calculation of all the transition states and connecting reactants, intermediates and products were then done at the RI-

CC2/TZVP+COSMO (ε=80.0) and RI-MP2[23]/TZVP+COSMO(ε=80.0) level of theory and the corresponding values have been reported in the Figures 5.2, 5.4 and 5.8. The trends for the ΔG and the $\Delta G^{\#}$ values were seen to match for the calculations done at both the levels of theory. In addition to the calculations done with Turbomole 7.0, we have also optimized all the transition states, corresponding reactants and products with the Gaussian09 software package[24] at the B3LYP/6-311++g(d,p)[25] level of theory (with DFT-D2,[26-30] a general, empirical dispersion correction proposed by Stefan Grimme for DFT calculations). Furthermore, we have also done all the calculations with the M06-2X hybrid fuctional[31-33] and the 6-311++g(d,p) basis set. In all the Gaussian09 calculations, we have modeled the solvent with the polarizable continuum model (PCM),[34] with water (ε = 80.0) as the solvent. These values are shown in Figures 5.3, 5.5, 5.9. Therefore, we have employed four different levels of theory for the QM calculations: B3LYP-D3/TZVP+COSMO(ε=80.0)//RI-CC2/TZVP+COSMO(ε=80.0) and B3LYP-D3/TZVP+COSMO(ε=80.0)//RI-MP2/TZVP+COSMO(ε=80.0) in Turbomole 7.0, as well as B3LYP- D2/6-311++g(d,p)+PCM (ε=80.0) and M06-2X/6-311++g(d,p)+PCM (ε=80.0) in Gaussian09 for refining the reaction free energies and barrier heights. To confirm the local minima or transition state structures frequency calculations were performed for all the stationary points. We have further done intrinsic reaction coordinate (IRC) calculations to confirm that the obtained transition states connect with the correct reactants and products.

The translational entropy can be calculated by using the Sackur-Tetrode equation, which can be represented as;

$$S_{trans} = Rln[(\frac{10^{-15/2}}{N_A^4[X]})(\frac{2\pi MRTe^{5/3}}{h^2})^{3/2}] \qquad (5.1)$$

Where [X] is the concentration of the molecule, M is the mass of the particle, T is the temperature N_A is the Avogadro number, k is the Boltzmann constant and h is the Planck constant. However, in the solution phase, the Sackur-Tetrode equation belittles the translational entropy, as it avoid the molecular volume (V_{mol}). Mammen *et al.* introduced the *free volume* correction to conquer this problem regarding the translational entropy that calculates the exact translational entropy.[8] According to this method, it has been assumed that the volume of molecules in solution is smaller than the total volume.

$$V_{free} = C_{free}\left(\left(\sqrt[3]{\frac{10^{27}}{[X]N_0}} - \sqrt[3]{V_{mol}}\right)\right)^3 \quad (5.2)$$

Where C_{free} is 8 (for cube), N_0 is the Avogadro's number and [X]=55.5 mol/lt. After free volume correction to Equation 5.1, the translational entropy is represented by

$$S_{trans}^{analyte} = Rln\left[\left(\frac{10^{-15/2}V_{free}^{solvent}}{N_0^4[analyte]}\right)\left(\frac{2\pi MRTe^{5/3}}{h^2}\right)^{3/2}\right] = 11.1 + 12.5\ln(T) + 12.5\ln(M) + 8.3lnV_{free}^{solvent} \quad (5.3)$$

Where, $V^{solvent}$ is the free volume, T is the temperature. Therefore, this equation contributes the free translational entropy in solution.

5.3 Results and Discussions

5.3.1 Analysis of the Reaction Pathways Leading to the Formation of Specific Compounds. Formation of the Protein Precursor: Glycine. By using the AINR we have obtained different mechanistic pathways towards the formation of several important biomolecules starting from the aqueous HCN as reactant. In most of the pathways, formaldehyde 3, urea 26, formaldimine 2, and glycolonitrile 4 were seen to be formed as intermediates. This suggests that these species were the key intermediates en route to the formation of the target molecules, as has also been noted by experimentalists.[36-42] Moreover, small molecules such as CO 13, CO_2 11, and H_2 were produced (Figure 5.2), and these were seen to take part in the synthesis of comparatively larger organic molecules. Figure 3 below describes how the relevant intermediates formic acid, CO, and CO_2 are formed from HCN and H_2O, leading from HCN, 1, through the intermediate species 7, formamide 8, formic acid 9, to carbon monoxide, CO, 13. The complete free energy profile is shown in Figure 5.3.

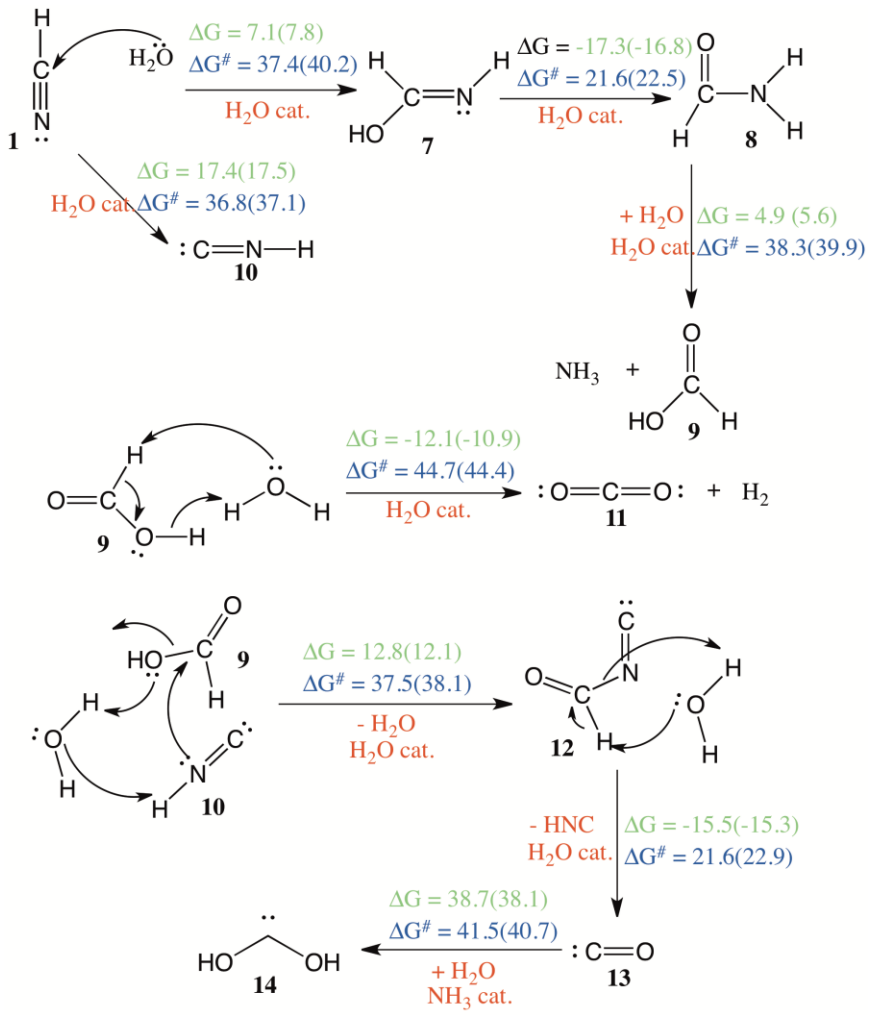


Figure 5.2 Sequence of elementary reaction steps derived from the AINR: the formation of HCOOH, CO_2, and CO starting from HCN and H_2O. Molecules labeled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO- (ε = 80.0) and the B3LYP-D3/TZVP+ COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses) levels of theory in kcal/mol.

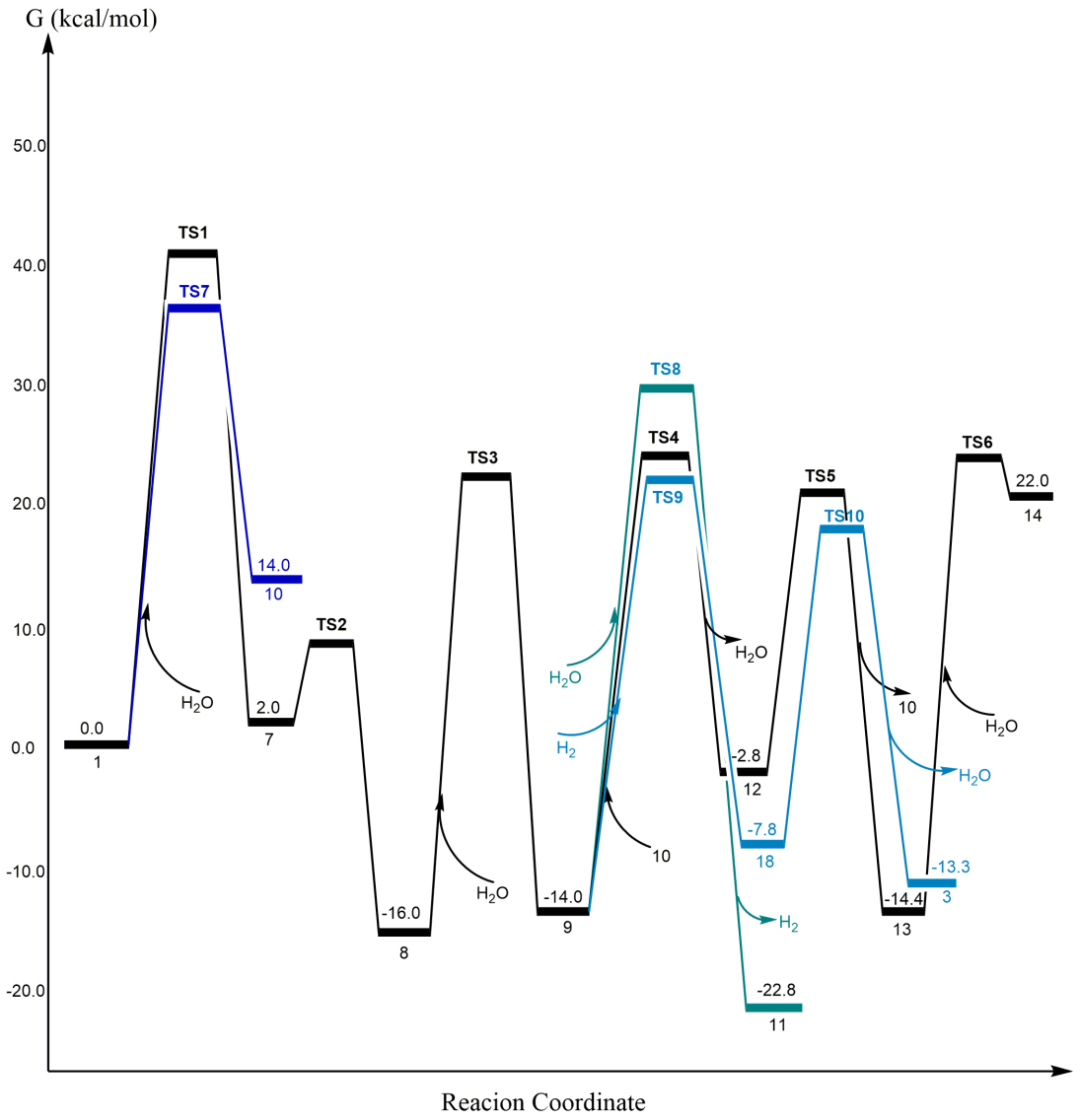
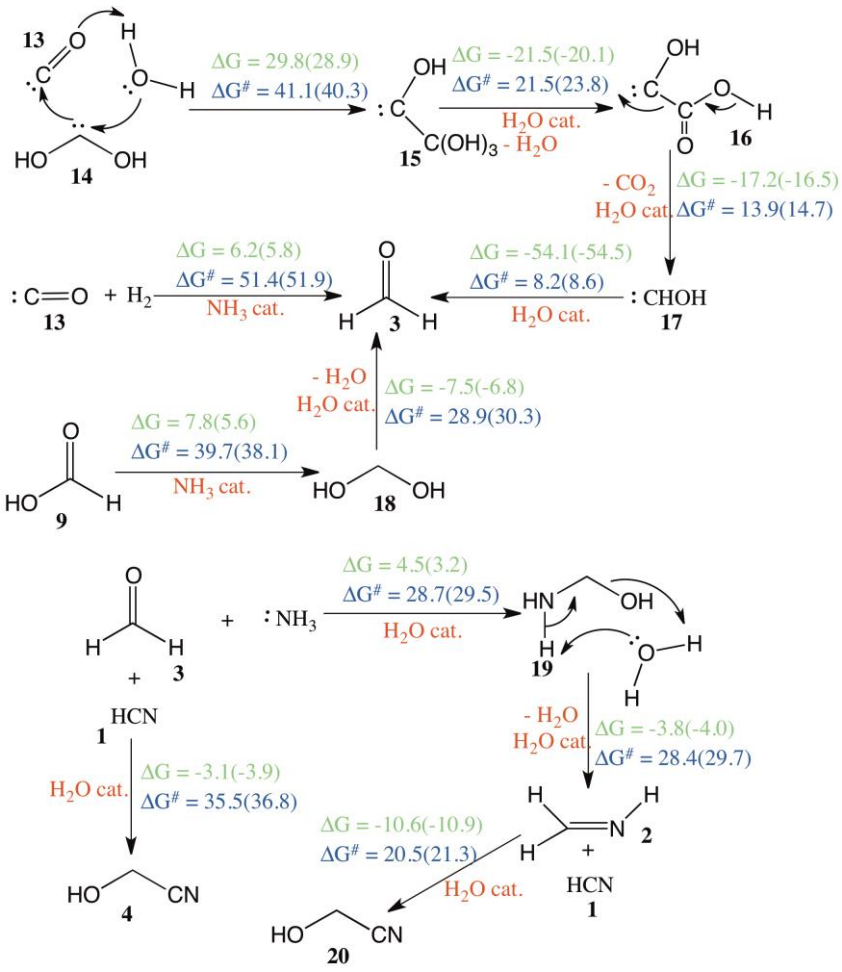


Figure 5.3 The free energy profile for the formation of intermediate molecules of the RNA and protein precursors starting from HCN and water molecules. The relative energies of the reactants and products of each elementary step have been represented with respect to HCN and the barrier has been calculated from the reactant species for each elementary step reaction. The values (in kcal/mol) have been obtained at the B3LYP-D2/6-311++g(d,p)+PCM(Ɛ=80.0) level of theory by Gaussian09 software package.

Figure 5.2 shows that species such as ammonia and dihydrogen were created in the AINR from the interaction between the HCN and water, and they turned out to be important substrates in subsequent reactions. This is interesting because it suggests that HCN and water would have created all the necessary reactants in subsequent steps.

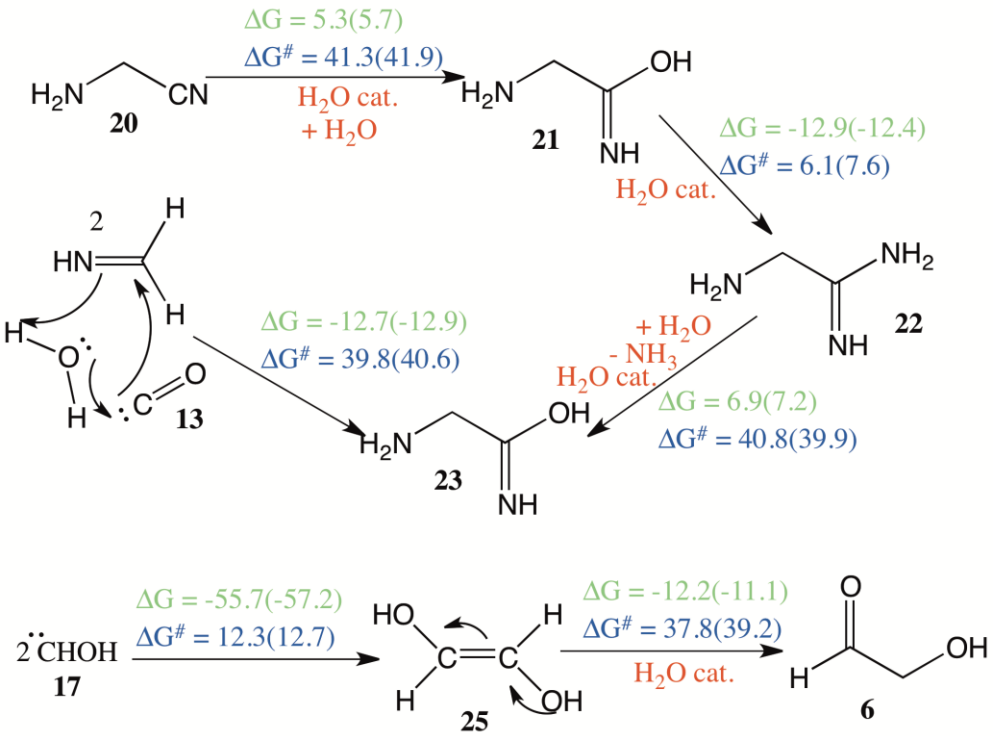
However, it could be argued that since the concentration of HCN would have been low such species would have been formed in low concentrations as well, which would have further reduced the yield of the subsequent products. The counter argument to this is that ammonia and other reactant species were also present separately in the oceans at that time because the protective haze of the Titan-like atmosphere would have prevented the photochemical degradation through UV of molecules such as ammonia in the atmosphere, and such molecules could have dissolved separately in the oceans as well and could be thus available for the reactions shown later in Figure 5.4.

Now, as we continue along this path, we find that important intermediate species: formaldehyde 3 and formaldimine 2 are formed during glycine 23 synthesis *via* several elementary steps (see Figure 5.4 A and energy profile in Figure 5.5). These intermediates lead to the formation of glycine 23, the precursor to proteins. A perusal of the two most feasible pathways found for glycine formation (shown in Figure 5.4 B) shows that 2 is present as an intermediate in both the cases. All the pathways are seen to proceed *via* stepwise elementary reaction steps. In one of the pathways, 2 reacts with hydrogen cyanide 1 to produce 20, which, through further stepwise hydrolysis, leads to 23. This is the well-known Strecker synthesis pathway.[43] That the AINR finds the same is gratifying and can serve as a validation for the computational approach. What is also finding is that the pathway does not involve protonation. We do note, though, that the AINR has also found another pathway, involving a trimolecular reaction between 13, 2, and water, with a five-membered transition state, which also leads to 23 (encircled in violet in Figure 5.4 B).



(A)

Figure 5.4 (A) The sequence of elementary reaction steps derived from the AINR: the formation of formaldehyde, formaldimine, glycolonitrile, and aminoacetonitrile. Values have been calculated at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO- (ε = 80.0) and the B3LYP-D3/TZVP+ COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses) levels of theory in kcal/mol.



(B)

Figure 5.4 (B) The formation of the target species: glycine and sugar. Values have been calculated at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO- (ε = 80.0) and the B3LYP-D3/TZVP+ COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses) levels of theory in kcal/mol.

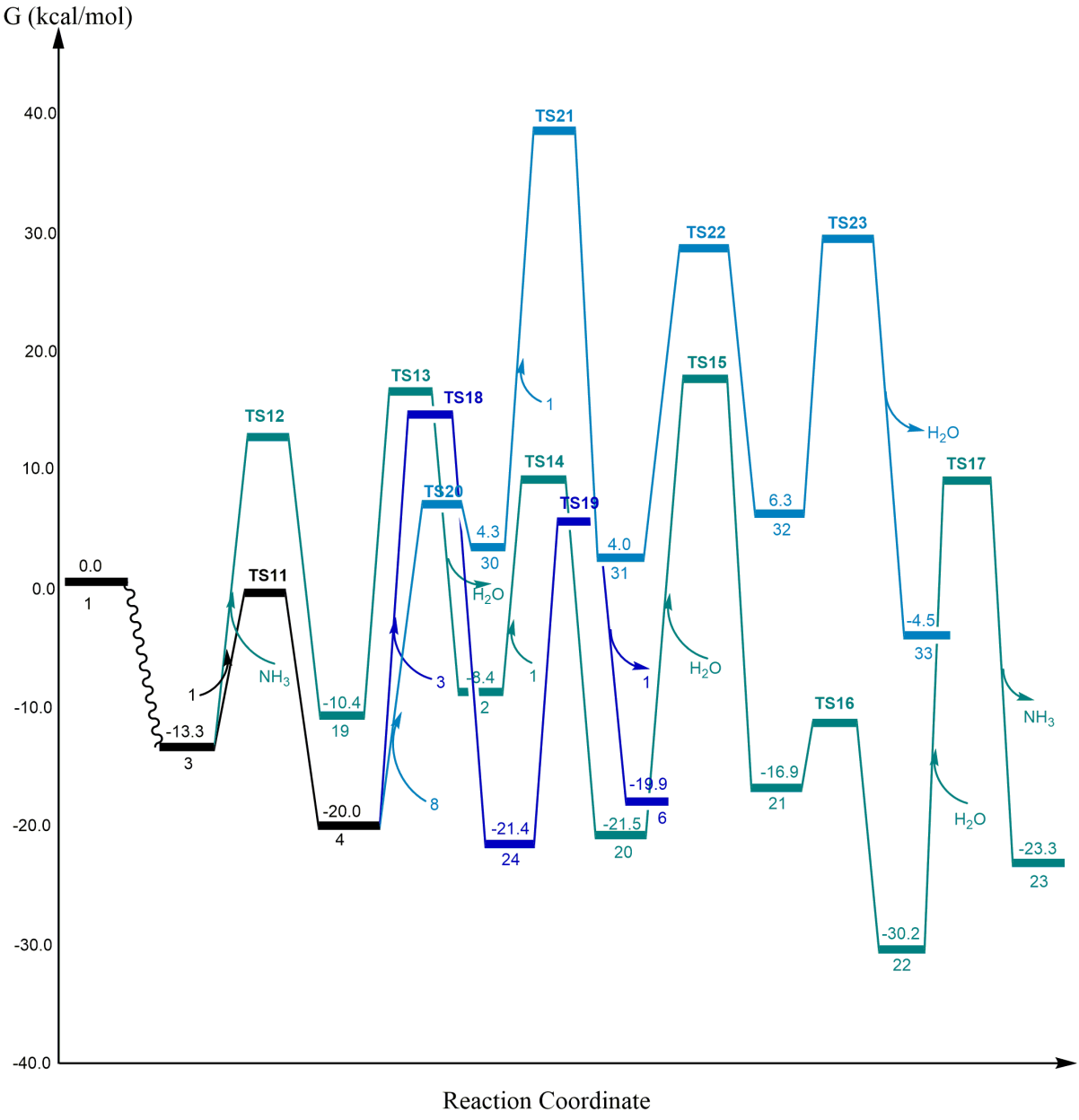


Figure 5.5 The reaction free energy profile diagram for the formation of the RNA precursors: glycoaldehyde and oxazole; and protein precursors: the glycine molecules *via* intermediate species formaldehyde, formaldimine and glycolonitrile beginning from HCN and H_2O. The relative free energies of the reactants and products for each elementary step are represented with respect to the beginning reactants and the barrier has been calculated from the reactant species of each elementary step. The values (in kcal/mol) have been obtained at the B3LYP- D2/6-311++g(d,p)+PCM (Ɛ=80.0) level of theory by Gaussian09 software package.

5.3.2 Pathways for RNA Building Units. The previous section discussed feasible pathways for the formation of the important protein precursor: glycine **23**. In this section, we discuss how the AINR also yields feasible pathways for the formation of important RNA precursors such as cyanamide **29**, glycoaldehyde (sugar) **6**, and the oxazole derivative **33**. The experimental synthesis of the sugar **6**, one of the most important RNA precursors, is challenging, but several reports have emerged recently where this has been achieved. Recently, Sutherland and co-workers[39] synthesized sugar from HCN and H_2O in the presence of a copper cyanide catalyst, through a photoredox cycle. They proposed a mechanistic pathway for the formation of the sugar **6** in the absence of the copper cyanide catalyst (Figure 5.6 A), involving two reduction steps, through which HCN **1** would be reduced to **2** and **4** to imine **5** in the presence of an H_2O catalyst molecule. We have calculated the barriers for these two processes (at the same level of theory as the mechanistic pathways investigated in the current work) and found them to be 83.2 and 85.8 kcal/mol respectively (see Figure 5.6 B, Supporting Information). Interestingly, the results obtained from the current computational studies reveal a completely different pathway for sugar formation, avoiding the reduction steps. The feasible pathway derived from the nanoreactor for the formation of **6**, where two molecules of hydroxymethylene carbene **17**, the formation of which was discussed in the previous section (see Figure 5.4 A), dimerize with a low barrier to form **24**, which further tautomerizes and leads to the formation of **6** (encircled in violet in Figure 5B). This pathway is found to be facile, with reduction not involved in any of the steps, with the slowest step seen to be ~37.0 kcal/mol, which is considerably lower (by more than 40.0 kcal/mol) than the barriers for the pathways (83.2 and 85.8 kcal/mol) that have been proposed in the literature.[39] The implication of this is that the process of forming sugar **6** would not have needed the presence and intervention of metal catalysts but would have been possible under thermal conditions, at temperatures of 80.0−100.0 °C.

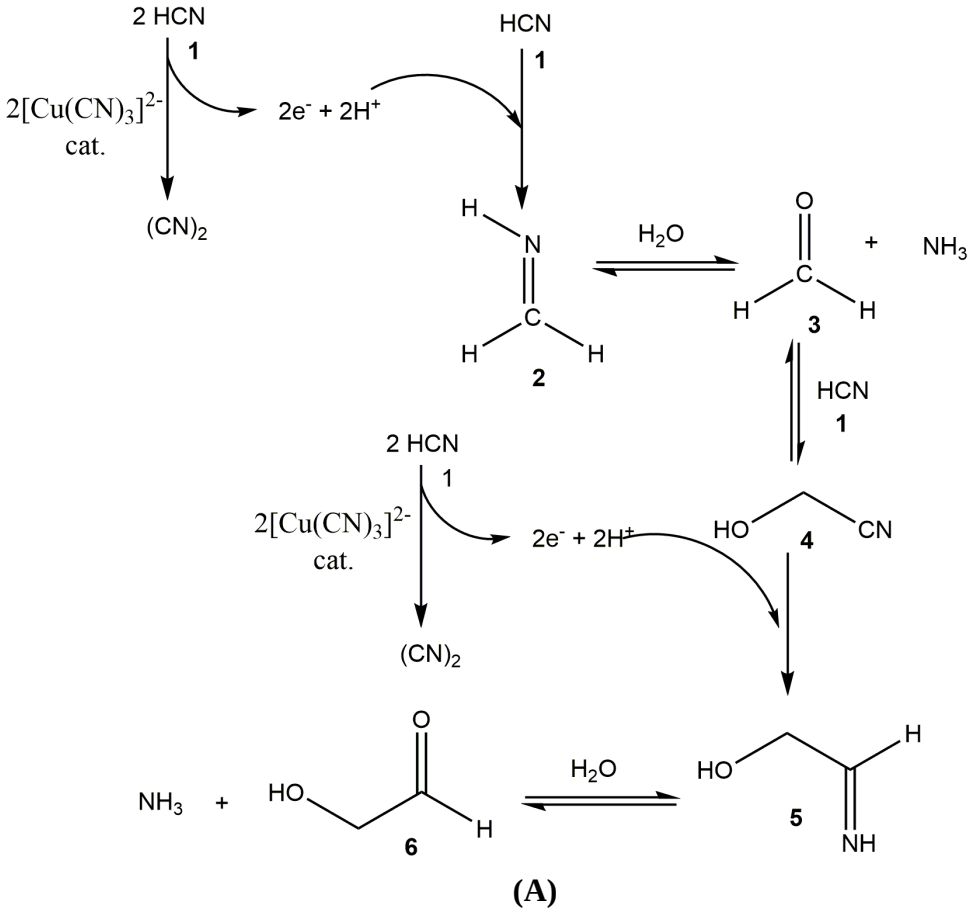
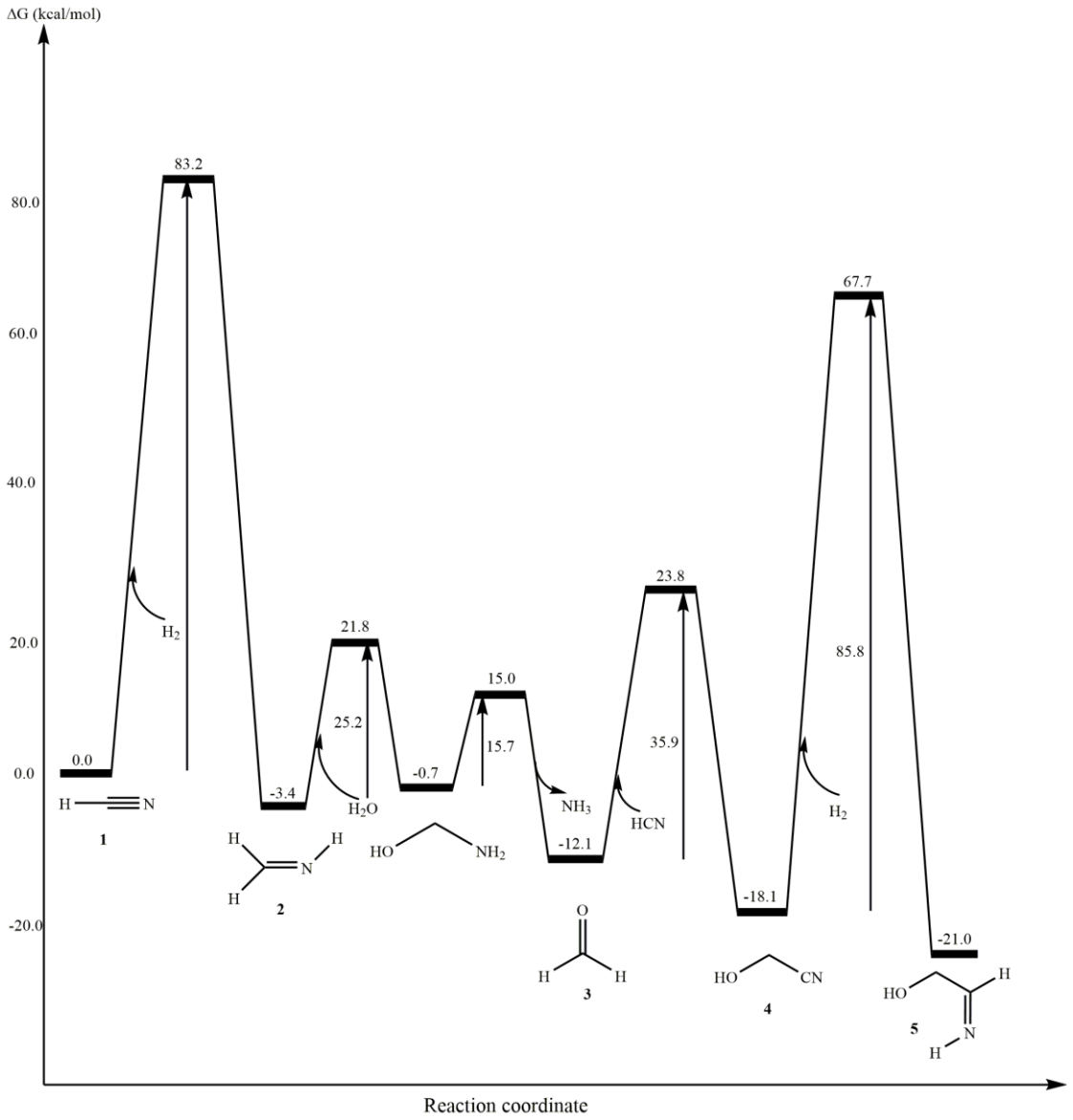


Figure 5.6 (A) The sugar synthesis pathway proposed by Sutherland and co-workers from HCN and water. It involves a photoredox cycling of the copper cyanide complex catalyst, producing two protons and two hydrated electrons from HCN, which further reduce another HCN molecule **1** to aldimine **2** in one step and glycolonitrile **4** to imine **5** in another step.



(B)

Figure 5.6 (B) The reaction free energy profile diagram for the sugar formation *via* the pathway proposed by Sutherland and co-workers. The relative free energy of the reactants and the products for each elementary step are represented with respect to the beginning reactants and the barrier has been calculated from the reactant species, for each elementary step. The values (in kcal/mol) have been obtained at the B3LYP-D3/TZVP+COSMO(Ɛ=80.0)//RI-CC2/TZVP+COSMO(Ɛ=80.0) level of theory by the use of the Turbomole 7.0 software package.

It is also interesting to note that Schreiner's group has recently reported[42] that carbene **17** is the intermediate en route to the formation of **6**. The fact that the AINR discovers the same intermediate, carbene **17**, to be important for the formation of **6** as this recent, independent, experimental study by Schreiner and co-workers is quite remarkable. We note, however, that the pathway that has been proposed for the

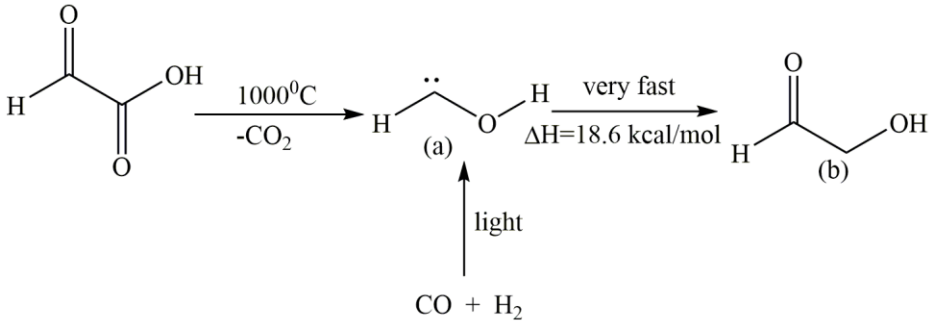


Figure 5.7 Schreiner and co-workers have proposed a new reaction pathway for sugar formation *via* hydroxyl methylene in the gas phase or on surfaces in the absence of a base.

transformation of **17** to **6** in the absence of solvent and base by Schreiner and co-workers (shown in Figure 5.7) is different from the one that we have discussed here. Since both the solvent and the base are relevant in our calculations (and in early Earth), it is quite likely that the facile pathways reported here transforming **17** to **6** represent avenues by which this important RNA precursor was formed in prebiotic Earth.

Cyanamide **29**, another precursor of RNA, is formed *via* two important intermediate species isocyanic acid **28**, or urea **26**. **28** reacts with ammonia in the presence of H_2O as a proton shuttling catalyst to produce **27**, which is further dehydrated in the presence of an NH_3 molecule acting as a catalyst, leading to the formation of **29** (shown in Figure 5.8 and the energy profile in Figure 5.9).

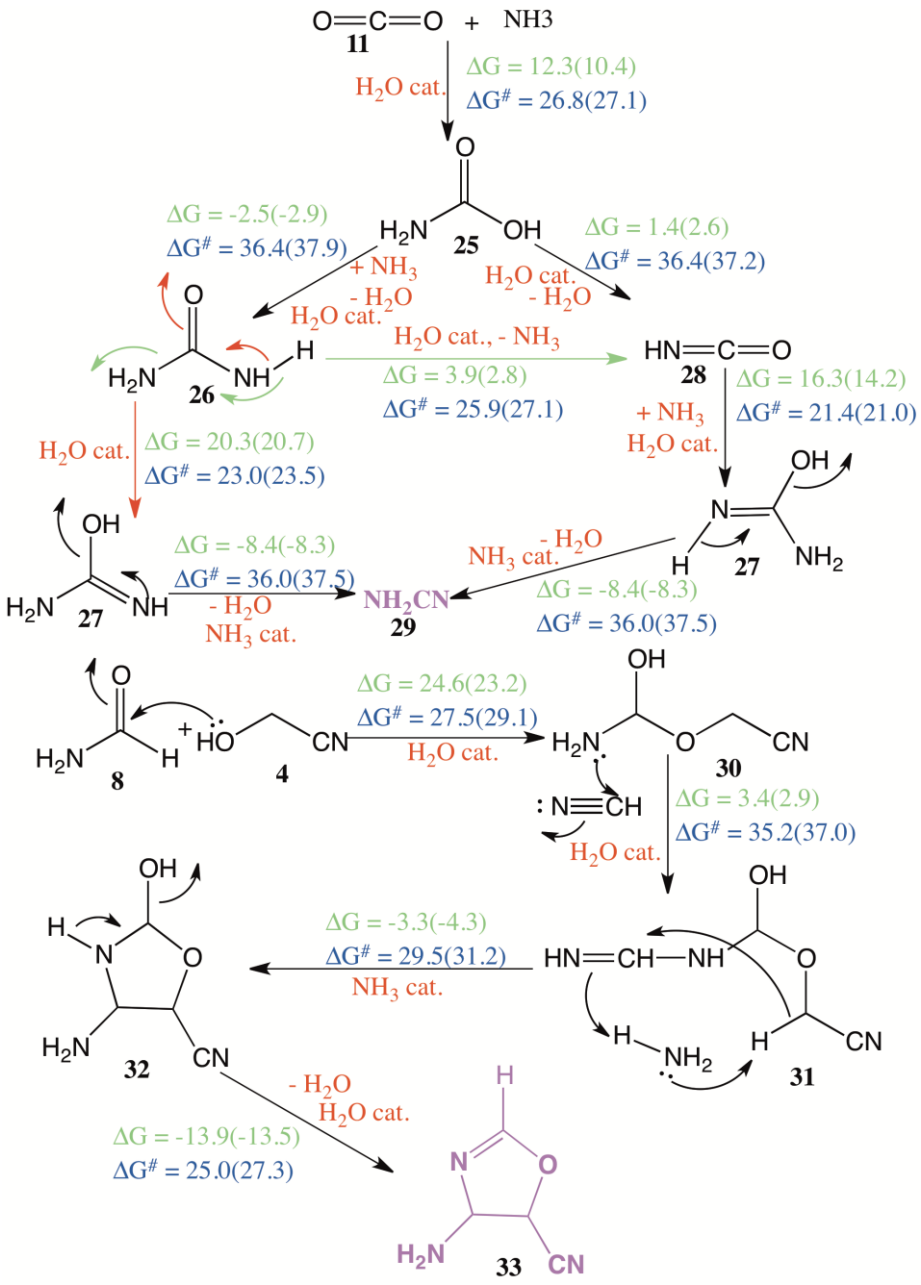


Figure 5.8 Formation of the target species: cyanamide and the oxazole derivative. Values have been calculated at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO- (ε = 80.0) and the B3LYP-D3/TZVP+ COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses) levels of theory in kcal/mol.

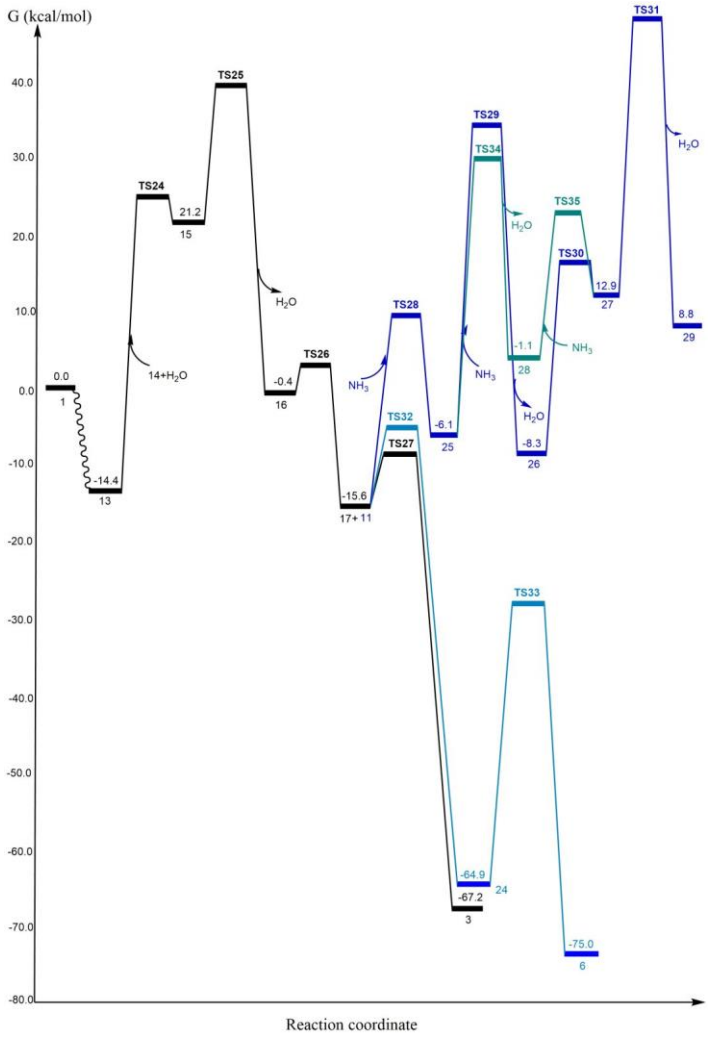


Figure 5.9 The reaction free energy profile diagram for the formation of the RNA precursors: cyanamide and sugar, starting from HCN and H_2O and with CO_2, urea, formaldehyde, glycolonitrile and other intermediates formed along the route. The relative free energy values of the reactants and products for each elementary step are represented with respect to the beginning reactants and the barrier has been calculated from the reactant species for each elementary step reaction. The values (in kcal/mol) have been represented at the B3LYP-D2/6-311++g(d,p)+PCM(Ɛ=80.0) level of theory with DFT calculated with the Gaussian09 software package.

During the nanoreactor simulations, apart from simple and complex acyclic organic compounds, numerous cyclic compounds such as oxazole, imidazole, as well as isoxazole derivatives were also seen to have formed (see Figure 4.4, Chapter 4). Among

these heterocyclic compounds, oxazole derivatives are among the more important, because 2-amino-oxazole is known to be an important precursor for RNA synthesis.[40,44-45] In our current work, we have shown a very feasible pathway for the formation of one of the oxazole derivatives **33** (encircled in violet in Figure 5.8) which is formed during the reaction between **8** with **4** (shown in Figure 5.8, with the energy profile in Figure 5.5).

5.4 Implications of the Current Work

There are, however, important questions that need to be addressed. First, there is the issue of low concentrations of HCN in water in early Earth conditions, which would have reduced the concentration of the subsequently formed RNA and protein precursors. A resolution to this problem is suggested by a recent molecular dynamics report which indicates that in dilute systems the HCN concentration is an order of magnitude larger in the surface layer than in the bulk liquid phase.[46] Such HCN concentration effects at the surface of water bodies would have facilitated the chemistry described here. Furthermore, certain hydrothermal vents at the bottom of the ocean are in the vicinity of cold seawater, as well as ice. Recent reports suggest that HCN could be stabilized and concentrated at water−ice interfaces as well.[47-48] Water containing this more concentrated HCN could then have seeped into the hydrothermal vents in the vicinity of the cold water−ice and undergone hydrolysis at higher temperatures inside the vents. The other issue is with regard to the eventual products of the HCN hydrolysis. Our results show that the formation of precursor molecules of RNA and protein would have been feasible, but the question then is would these precursors have been formed in sufficient concentrations to then react with each other, in order to lead to greater complexity? One solution that can be provided to this problem is to invoke the idea of a "warm little pond" that had been suggested by Charles Darwin in 1871,[49] i.e., to consider shallow ponds, lakes, estuaries, or tidal lagoons in prebiotic times that would have had temperatures of about 100.0 °C. The reactions discussed here could have happened in such water bodies, and then evaporation of the water would have led to increased concentration of the products formed. Such a scenario would have led to greater interactions between the molecules formed, and thus, to more complex molecules. It is also possible that such precursors could have seeped out with water from hydrothermal vents and been concentrated at water−ice interfaces in the vicinity of the vents, which would then have allowed subsequent, more complex molecules to have emerged. Another clarification that should be made is with

regard to the specified condition (iii) in the Introduction, regarding the need to have chemical reactions occurring without the need for metal catalysts. This condition increases the probability of the chemical reactions taking place all over the Earth's oceans, and not just in the few, select regions of the Earth where metal based catalysts were available. However, in the regions where metal based catalysts were available in early Earth, their presence would have been beneficial and accelerated the formation of the precursors for RNA and proteins. The other salient points gleaned from the AINR studies are as follows:

(i) As the pathways found for sugar formation (see Figure 5.4 B) indicate, lower barriers have been obtained for chemical conversions when an pathway alternative to reduction was found (see the respective steps in the reduction pathways, shown in Figure 5.6 B). This corroborates experimental observations that indicate that reduction was generally avoided in prebiotic chemistry.[50]

(ii) The formation of low valent species such as carbenes (**17, 12, 14, 15, 16**, NH_2COH) is an important reason why most of the barriers for the mechanistic pathways discovered through the AINR approach were seen to be reasonable to low. This interesting fact echoes previous hypotheses that low valent main group compounds are important intermediates in mechanistic cycles.[42,51-52]

(iii) The AINR was seen to exploit H_2O or NH_3 molecules as proton shuttling catalysts in most of the elementary reaction steps. This, again, has relevance in the context of recent reports,[18,53-54] suggesting that a lot of biology occurs with the mediation of H_2O as a proton shuttling catalyst. Moreover, the role of NH_3 as a proton shuttling agent has also been explored in the literature.[55-58]

5.5 Conclusions

Pathways that were found to be feasible were seen to avoid the reduction step, corroborating previous experimental reports.[50] Most of the steps of the discovered mechanistic routes have barriers that are low to moderate, with only a few higher barriers of ~40.0 kcal/mol, which suggests that the reactions could have occurred without the mediation of metal catalysts and through the aid of thermochemistry alone. This insight is valuable because it helps to explain how the reactions could have taken place in the absence of photochemical activity on the surface

of Earth's oceans. These findings make it possible to imagine that the molecules necessary for building larger, more complex entities such as RNA and proteins could have existed and interacted together in at least some of the water bodies present in early Earth. The current work thus indicates that HCN and H_2O could have been the Adam and Eve of chemical evolution- the source of the precursor molecules that formed the basis of life on Earth.

5.6 References

1. Miller, S. L. *Science*, **1953**, *117*, 528−529.
2. Aubrey, A. D.; Cleaves, H. J.; Bada, J. L. *Origins Life Evol. Biospheres*, **2009**, *39*, 91−108
3. White, R. H., *Nature*, **1984**, *310*, 430−432.
4. Miller, S. L.; Bada, J. L. *Nature*, **1988**, *334*, 609−611.
5. Szori, M.; Jojart, B.; Izsak, R.; Szori, K.; Csizmadia, I. G.; Viskolcz, B. *Phys. Chem. Chem. Phys.*, **2011**, *13*, 7449−7458.
6. Florian, J.; Warshel, A. *J. Phys. Chem. B*, **1998**, *102*, 719−734.
7. Song, J. L.; Wang, T.; Zhang, X.; Chung, W. L.; Wu, D. Y. A *ACS Catal.*, **2017**, *7*, 1361−1368.
8. Lee, B. T.; Mckee, L. M. *Inorg. Chem*., **2009**, *48*, 7564−7575.
9. Nguyen, T. M.; Nguyen, S. V.; Matus, H. M.; Gopakumar, G.; Dixon, A. D. *J. Phys. Chem. A*, **2007**, *111*, 679−690.
10. Yamakawa, M.; Ito, H.; Noyori, R., *J. Am. Chem. Soc*., **2000**, *122*, 1466−1478.
11. Hagberg, A. A.; Schult, D. A.; Swart, P. J. *Proceedings of the 7th Python in Science Conference* (eds Varoquaux, G., Vaught. T & Millman, J.) **2008**, 11–15 (SciPy).
12. Travis, E.; Oliphant, E. *USA: Trelgol Publishing*., **2006.**
13. Gansner, E. R.; North, S. C. *Pract. Exper*. **2000**, *30*, 1203– 1233.
14. Jurafsky, D.; Martin, J. H. *Prentice Hall, Englewood Cliffs*, New Jersey, Draft of September 11 **2018**.
15. Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol.* Design **2000**, *14*, 123-134.
16. Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14.1*, 33-38.
17. Viterbi, A. J. *Transactions on Information Theory*. **1967**, *13*, 260–269.

18. Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. *Nat. Chem*, **2014**, *6*, 1044-1048.

19. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett*., **1989**, *162*, 165−169.

20. Ansgar, S.; Christian, H.; Reinhart, A. *J. Chem. Phys*., **1994**, *100*, 5829−5835.

21. Becke, A. D. *J. Chem. Phys.*, **1993**, *98*, 5648–5652.

22. Klamt, A.; Schuurmann, G. J. *Chem. Soc., Perkin Trans*. **1993**, *2*, 799−805.

23. Grotendorst, J.; Bl¨ugel, S.; Marx, D *Computational Nanoscience,* **2006,** 31, 245-278.

24. Gaussian 09, Revision **E.01**; Frisch , M. J.; Trucks, G. W.; Schlegel, H. B.;Scuseria,G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.;Jr.,Peralta, J. E.; Ogliaro, F.,Bearpark, M.; Heyd, J. J. E.; Brothers,Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Raghavachari, J.K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., *Wallingford CT*, **2009.**

25. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.*, **1980**, *72*, 5639-5648.

26. Hepburn, J.; Scoles, G.; Penco, R. *Chem. Phys. Lett*. **1975**, *36*, 451–456.

27. Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys*.**1977**, *19*, 119–130.

28. Grimme, S. *J. Comput. Chem*. **2004**, *25*, 1463–1473.

29. Grimme, S. *J. Comput. Chem*. **2006**, *27*, 1787–1799.

30. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys*. **2010**, *132*, 154104-154119.

31. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.*, **2008**, *120*, 215–241.

32. Zhao, Y.; Truhlar, D. G. *J. Phys. Chem.* A, **2006**, *110*, 13126– 13130.

33. Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
34. Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.*, **2005**, *105*, 2999-3094.
35. Mammen, M.; Shakhnovich, E. I.; Deutch, J. M.; Whitesides, G. M. *J. Org. Chem.*, **1998**, *63*, 3821−3830.
36. Bowler, F. R.; Chan, C. K.; Duffy, C. D.; Gerland, B.; Islam, S.; Powner, M. W.; Sutherland, J. D.; Xu, J. *Nat. Chem.*, **2013**, *5*, 383−389.
37. Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D., *Nat. Chem*., **2015**, *7*, 301−307.
38. Schlesinger, G.; Miller, S. L. *J. Mol. Evol.*, **1983**, *19*, 383−390.
39. Ritson, D.; Sutherland, J. D. *Nat. Chem*., **2012**, *4*, 895−899.
40. Eschenmoser, A.; Loewenthal, E. *Chem. Soc. Rev*., **1992**, *21*, 1−16.
41. Thaddeus, P. Philos. *Trans. R. Soc., B* **2006**, *361*, 1681−1687.
42. Eckhardt, A. K.; Linden, M. M.; Wende, R. C.; Bernhardt, B.; Schreiner, P. R. *Nat. Chem*., **2018**, *10*, 1141− 1147.
43. Strecker, A. *Liebigs Ann. Chem*. **1854**, *91*, 349− 351.
44. Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature*, **2009**, *459*, 239−242.
45. Cockerill, A. F.; Deacon, A.; Harrison, R. G.; Osborne, D. J.; Prime, D. M.; Ross, W. J.; Todd, A.; Verge, J. P. *Synthesis*, **1976**, *76*, 591−593.
46. Fabian, B.; Szori, M.; Jedlovszky, P. *J. Phys. Chem. C*, **2014**, *118*, 21469−21482.
47. Szori, M.; Jedlovszky, P. *J. Phys. Chem. C*, **2014**, *118*, 3599−3609
48. Menor-Salvan, C.; Marín-Yaseli, R. M. *Chem. Soc. Rev*., **2012**, *41*, 5404−5415.
49. Ball, P. *The Royal Society of Chemistry: Cambridge UK*, **2005**; pp , 1−205.
50. Danger, G.; Plasson, R.; Pascal, R. *Chem. Soc. Rev*., **2012**, *41*, 5416−5429.
51. Mandal, S. K.; Roesky, H. W. *Acc. Chem. Res.* **2012**, *45*, 298−307.
52. Nesterov, V.; Reiter, D.; Bag, P.; Frisch, P.; Holzner, R.; Porzelt, A.; Inoue, S. *Chem. Rev*., **2018**, *118*, 9678−9842.
53. Mikulski, R.; West, D.; Sippel, K. H.; Avvaru, B. S.; Aggarwal, M.; Tu, C.; McKenna, R.; Silverman, D. N. *Biochemistry*, **2013**, *52*, 125−131.
54. De Vivo, M.; Ensing, B.; Klein, M. L. *J. Am. Chem. Soc*., **2005**, *127*, 11226−11227.

55. Pal, A.; Vanka, K. *Inorg. Chem*., **2016**, *55*, 558−565.
56. Cord-Ruwisch, R.; Law, Y.; Cheng, K. Y. *Bioresour. Technol*. **2011**, *102*, 9691−9696.
57. Meuwly, M.; Karplus, M. *J. Chem. Phys*., **2002**, *116*, 2572−2585.
58. Jaroszewski, L.; Lesyng, B.; Tanner, J. J.; McCammon, J.A. *Chem. Phys. Lett.*, **1990**, *175*, 282−288.

Chapter 6

The Effect of Oxidizing Atmosphere on the Origin of Life in Prebiotic Earth & Interstellar Space

Chapter 6

The Effect of Oxidizing Atmosphere on the Origin of Life in Prebiotic Earth & Interstellar Space

Abstract

Around six decades ago the crucial Urey-Miller experiment demonstrated that life building blocks such as amino acids could have formed from the reducing environment that was supposed to have existed in the primitive earth. Similarly, the composition of interstellar ice mixtures could have also produced several RNA and protein precursors. Though hypotheses and experimental studies have proposed that life building units might not have evolved in an oxidizing atmosphere, theoretical studies have not been done to provide insights into why the presence of oxygen would have been deleterious for the development of the building blocks of life. In our present study, we have tried to mimic the conditions in the primitive earth and interstellar space to show how the various life building blocks emerged into the earth and also attempted to further understand the potential negative effect of an oxygen rich atmosphere by introducing dioxygen to the Urey-Miller mixture as well as to an interstellar ice composition. With the help of the *ab initio* nanoreactor (AINR), a recently developed computational tool, the current computational study provides some interesting insights into such chemistry.

6.1 Introduction

How life on Earth began stays an unsolved scientific problem. Researchers in this area have endeavored to address this issue computationally[1-4] and experimentally.[5-18] Under conceivable prebiotic conditions, molecules such as methane, carbon dioxide and ammonia, which would have contributed to a reducing environment, were available in the Hadean atmopsphere. Furthermore, incoming interplanetary residue particles from comets and meteorites may have added to the prebiotic atmosphere. Likewise, hydrothermal vents could have also possessed a combination of different reducing gas mixtures. The primitive organic soup may have been complicated, but it did not probably incorporate the entirety of the mixtures that are possible in present-day atmospheres.

Of the multitude of theories with respect to cause of life that were proposed,[19-20] few were as accapted as that of Oparin,[21-22] who proposed a significant timescale for the synthesis of organic species abiotically. Comparable thoughts were proposed also by Haldane.[23] The cutting edge center around the atmosphere as the wellspring of prebiotic science dates to the celebrated Urey-Miller experiments of the 1950s.[12,24-26] These investigations were planned to demonstrate the sorts of disequilibrium chemistry that would have come about because of electrical releases in, or UV radiation being consumed in, profoundly diminished environments in which methane, ammonia, and water were all significant constituents. These investigations were driven by Harold C. Urey's hypothesis that Earth accumulated as a cool body and that its environment was overwhelmed by hydrogen and the hydrides of normal volatiles. Miller and coworkers, and numerous different experimentalists who have since performed experiments, have reliably discovered that a wide scope of amino acids and other prebiotically intriguing molecules in such conditions.[27] These experiments were exceptionally powerful in guiding the consideration of prebiotic chemists to a highly reduced early stage environment. From that point forward, a wide assortment of organic species of biochemical importance have been tentatively synthesized from basic molecules like water, ammonia, HCN[28] and methane. These experimental results have been featured by the disclosure of an enormous assortment of organic molecules in the interstellar space of the Milky Way and in comets.[29]

In any case, photochemical investigations showed that any methane[30] or ammonia[31-32] in the environment would rapidly be obliterated. There were geology based arguments that the

Earth's initial atmosphere was made, for the most part, out of H_2O, CO, and N_2, with just modest quantities of CO_2 and H_2, and basically no CH_4 or NH_3.[33-34] Urey-Miller type experiments considered the more oxidized combinations of present day volcanic gases, particularly where CO_2 was plentiful.[12,28,35] Cleaves *et al.* suggested that spark discharge products of HCN, ammonia and amino acids from the mixture of CO_2-N_2-H_2O could be less baffling if the water was permitted to get acidic. Therefore, scientists have synthesized different life building units in the last five six decades by taking the gaseous composition of either reducing or neutral atmospheric conditions of primitive earth, suggested from several geochemical scenario. Hence, the question certainly arises as to why there were no such experiments where oxidizing atmospheric compositions had been taken. There might be some hidden role of the oxidizing atmosphere towards the inhibition of the formation of life building blocks.

In our current study, we have shown how the oxidizing atmosphere inhibits the formation of sugar and amino acids which are the precursors to the RNA and proteins in the prebiotic earth as well as interstellar space. This issue been computationally unexplored till date. We have done *ab initio* nanoreactor dynamics (AINR) simulations by taking Urey-Miller gaseous compositions, representing the prebiotic earth atmosphere, containing NH_3, CO, CH_4, H_2O, H_2 species. Further, we have added CO_2 and O_2 molecules: both together, and separately in new considered mixtures, and observed the effects towards the formation of life building block precursors. Furthermore, we have performed similar reactive dynamics with taking the bare interstellar ice composition (CH_3OH, NH_3, CO, H_2O) and similarly added the CO_2 and O_2 molecules and checked the outcomes from the dynamics by using high level full static quantum chemical studies with density functional theory (DFT) and thus obtained all the barriers ($\Delta G^{\#}$) for the reactions involved in these processes, as well as the energies (ΔG) of the reactions.

6.2 Computational Methods:

6.2.1 *Ab Initio* Molecular Dynamics (AIMD) Simulations. The TeraChem 1.9[36-42] software package has been employed for performing AIMD simulations. In order to calculate the Born−Oppenheimer potential energy surface Hartree−Fock (HF)[43], the electronic wave function and the 3-21g(d)[44] Gaussian basis set has been employed. This method has been implemented in TeraChem by Todd Martinez and co-workers. This approach was deemed acceptable because the

HF method is well-known for predicting chemically reasonable structures.[45] Also, it should be noted that HF was not employed to determine the thermodynamics and reaction rates of the reactions: its only role was in the discovery process. This was also the approach employed by Martinez and co-workers in their original *AINR* paper (employing HF/3-21g), where they replicated the results obtained from the Urey−Miller experiment, as well as from the interaction of acetylene molecules. The similar method (HF/3-21g(d)) was also employed by us in our previous report on reaction pathways leading to the formation of precursors of RNA and sugars.[46]

The results were obtained from the *AINR* simulations by varying the composition of the gaseous mixture of Urey-Miller, interstellar ice analogues, as well as hydrothermal vent compositions. The systems were constrained in a spherical boundary of r_1 and r_2 radii (please see Chapter 4, Section 4.2.2 for details regarding AINR spherical boundary conditions) depending upon the number of molecules of various gaseous mixture taken within the spherical volume of AINR, so that the atoms resided in a space that alternated between the volumes created by these two radii, and collided with each other. Each *AINR* dynamics was evolved upto ~1ns, with a time step of 0.5 fs.

For Newton's equations of motion, Langevin dynamics has been employed with an equilibrium temperature of 2000.0 K (also the simulations starting temperature). We have used this high temperature in order to increase the average kinetic energy of the reactants and for faster dynamics, allowing the overcoming of noncovalent interactions without the breaking of covalent bonds. The nanoreactor simulations employ a piston to accelerate the reaction rate. We have employed the augmented direct inversion in the iterative subspace (ADIIS) algorithm[47] available in TeraChem as an alternative tool for self-consistent field calculations at each AIMD step in which the default DIIS algorithm[48] failed to converge. Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event.

The mechanistic pathways obtained from the AINR simulations were then analyzed to determine the reaction free energy (ΔG) and energy barriers ($\Delta G^{\#}$). We have done the minimum energy pathway (MEP) search by full quantum mechanical calculations, including zero point energy, internal energy, and entropic contributions, with the temperature taken to be 298.15 K. All the calculations for the structures reported have been done using density functional theory

(DFT). Geometry optimizations and transition state search calculations were carried out with the Turbomole 7.0 software package[49] using the TZVP basis set[50] and the B3LYP three parameter hybrid density functional.[51] Dispersion corrections (D3) were included in all the calculations. Solvent corrections were included with the dielectric continuum solvent model COSMO[52] with ε = 80.0.

6.3 Results and Discussions

The approach of AINR makes use of collisions between the molecules of the Urey-Miller gaseous mixture (CH_4, CO, H_2, NH_3, H_2O), as well as the interstellar ice composition (NH_3, CO, CH_3OH, H_2O and small amounts of CO_2) which affords the energy required to overcome the activation barriers for each of the elementary steps of the complete reaction network. The simulations have been done on systems having nearly homogeneously mixed 7CH4, 7CO, $7H_2$, $7NH_3$ and $7H_2O$ molecules as the starting reactants in case of Urey-Miller simulation, previously done by Wang *et al.*[2] $12NH_3$, 7CO, $7CH_3OH$ and $7H_2O$ molecules were taken together for the interstellar ice matrix simulation, and the systems were followed in the AINR for ~1.0 ns. Collisions between the molecules gave rise to the formation of new molecules. It should be pointed out that the homogeneous mixtures that were taken for the Urey-Miller and interstellar ice simulations do not represent the exact ratios of the gaseous mixtures present in the primitive Earth atmosphere or interstellar ice. The reason for taking such mixtures was to maximize the possibility of collisions between the molecules in the AINR. This would increase the probability of obtaining different products during the simulations. The aim of the AINR simulations was to discovering new species or mechanistic pathways for the formation of important life building units. In the next sections we have discussed the results that have been obtained by the AINR approach.

6.3.1 Effect of oxygen rich atmosphere towards Urey-Miller gaseous mixture: Our goal is to check the formation of important precursors of RNA and proteins through the series of AINR dynamics. Here we have especially focused on the formation of gycine (protein precursor), glycoaldehyde (two carbon sugar) and cyanamide (RNA precursors). The reason behind the choice of these three precursor molecules is to distinguish between nitrogenous and oxygenous chemistry. There is a clear structural difference among these molecules, where glycoaldehyde and cyanamide contain only oxygens and nitrogens as heteroatoms respectively. On the other

hand, glycine contains both the heteroatoms. During the progression of the simulation with the bare Urey-Miller mixture, we have observed that all the above mentioned precursor molecules have formed in different periods of time (in ps). Furthermore, we have performed a series of simulations where we have included molecular oxygen and carbon dioxide individually, as well as together, into the Urey-Miller mixture (shown in Table 6.1). Interestingly, it has been observed that there was no formation of glycoaldehyde and glycine in the presence of molecular oxygen but only nitrogen heteroatom containing cyanamides were formed. It is also to be noted that in the AINR dynamics (shown in Table 6.1) we have kept the Urey-Miller composition fixed and varied the molecular oxygen amount, as well as considered cases where we have varied both at the same time.

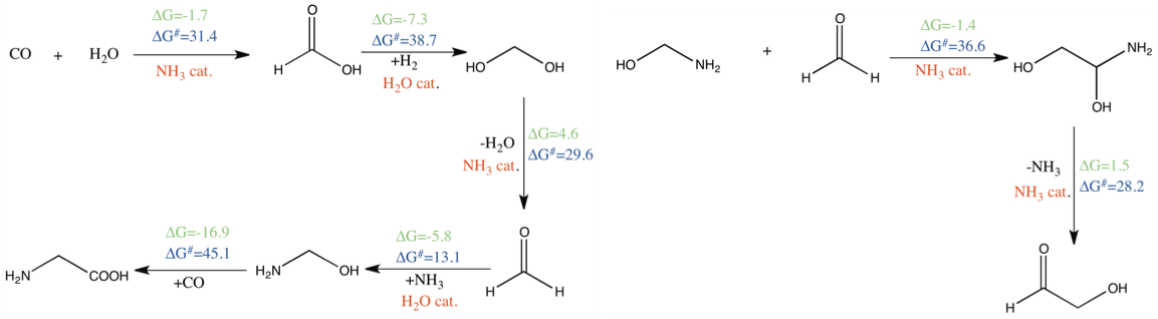
Table 6.1: Addition of O_2 & CO_2 in the Urey-Miller Gaseous Mixture

Composition	glycoaldehyde	glycine	cyanamide
NH_3+CO+H_2O+H_2+CH_4	✓	✓	✓
NH_3+CO+H_2O+H_2+CH_4+O_2	✗	✗	✓
NH_3+CO_2+H_2O+H_2+CH_4+O_2	✗	✗	✓
NH_3+CO_2+H_2O+H_2+CH_4	✓	✗	✓
NH_3+CO_2+H_2O+H_2+CH_4+O_2+CO	✗	✗	✓
NH_3+CO_2+H_2O+H_2+CH_4+CO	✓	✗	✓

6.3.2 Following a specific reaction formation of glycine and glycoaldehyde. The utilization of the nanoreactor produces as a yield a wide range of pathways to new species from the beginning reactants. In a large portion of the reaction network, formaldehyde, formaldimine and formic acid were formed as intermediates. This indicates that these species were the key intermediates in the

transition to the formation of desired life building units, as also has been pointed out by experimentalists.[53-59] Apart from this, a lot of diverse organic species were also observed to have formed during the simulations. The pathways described here (shown in figure 6.1) are the feasible ones towards the formation of glycine and glycoaldehyde - the major precursors of protein and RNA respectively. Here, glycine was formed (Figure 6.1 A) starting from the reaction between CO and H_2O, leading to the formation of formic acid, which further hydrogenated to methanediol. Dehydration of methanediol led to the formation of the very important prebiotic intermediate formaldehyde. During this pathway, formaldehyde reacts with ammonia, leading to the formation of aminomethanol, which is another key intermediate and also a precursor to formaldimine, reacting directly with CO to yield glycine. This pathway reveals that CO is currently considered to be toxic for human life, but played a key role towards the formation of important life building units.

Here, we have shown a new mechanistic pathway for sugar formation (Figure 6.1 B) through an AINR study where aminomethanol reacted with formaldehyde, leading to the formation of an intermediate species. Further deammoniation yielded glycoaldehyde.



(A) **(B)**

Figure 6.1 (A) Pathway for the formation of glycine during the *ab initio* nanoreactor (AINR) dynamics. **(B)** Pathway for the formation of glycoaldehyde during the *AINR* dynamics. The reaction energies (ΔG) are shown in green and the activation barriers ($\Delta G^{\#}$) are shown in blue, calculated at the B3LYP-D3/TZVP+COSMO (ε=80.0) level of theory with DFT. All the energies are in kcal/mol.

6.3.3 Effect of an Oxygen Rich Mixture in Interstellar Ice

In the previous section, we have discussed about the effect of the inclusion of oxygen into the Urey-Miller mixture. Now, we will discuss the effect of an oxidizing atmosphere in the interstellar space ice mixture. The interstellar space ice analogues consist of very simple and small molecules including NH_3, H_2O, CO, CH_3OH and a little CO_2. Previous experiments[60-62] have shown that by taking different ratios of interstellar ice forming molecules and irradiating them with ultraviolet photons, scientists have successfully synthesized different life building block precursors such as amino acids, sugars and nucleic acid bases. In our current computational study with the aid of AINR, we have not only shown the effect of an oxygen rich atmosphere towards the formation of RNA and protein precursors, but also shown some feasible mechanistic pathways for the formation of glycoaldehyde, which has not been explored yet. During the AINR dynamics with a bare interstellar ice analogues composition (NH_3, CO, CH_3OH, H_2O or NH_3, CO, CH_3OH, H_2O, CO_2), we have observed the formation of three of the desired precursors (glycine , glycoaldehyde and cyanamide) to RNA and proteins (shown in Table 6.2).

Table 6.2: Addition of O_2 & CO_2 Towards Interstellar Gaseous Mixture

Composition	glycoaldehyde	glycine	cyanamide
NH_3 +CH_3OH+H_2O+CO+CO_2	✓	✓	✓
NH_3+CH_3OH+H_2O+CO	✓	✓	✓
NH_3+CH_3OH+H_2O+CO_2	✓	✗	✓
NH_3+CH_3OH+H_2O+CO+O_2	✗	✗	✓
NH_3+CH_3OH+H_2O+CO_2+O_2	✗	✗	✓

Now, when we have added the molecular oxygen into the bare interstellar ice mixture composition, similar results were observed like in the Urey-Miller case where there was no formation of glycine and glycoaldehyde but only cynanamide was formed throughout the AINR

trajectories. In all the cases, we have taken an almost homogeneous mixture of bare interstellar ice components and varied the oxygen amounts in the initial reactants mixture and allow them for AINR dynamics. It has been clear from the dynamics that oxygen inhibits the formation of precursors molecules where both oxygen and nitrogen heteroatoms are present or only oxygen is present. If one looks at the structure of the target species, in the case of glycine, glycoaldehyde and cyanamide, the source of oxygen is from water or methanol or carbon monoxide or carbon dioxide but on the other hand, the sole source of nitrogen is from NH_3. Now, due to the inclusion of oxygen molecules into the interstellar ice mixture, the concentration of oxygen percentage increases, which inhibits the formation of the actual target species (glycine and glycoaldehyde) and forms, instead, some unwanted side products. Due to the very complex reaction network in the presence of oxygen molecules in our current scope, we could not properly track the reaction pathways for the inhibition of the product formation.

During the AINR simulation with a bare interstellar ice composition, we have found a feasible pathway for the formation of two carbon sugars: glycoaldehyde (shown in Figure 6.2). The reaction between methanol and CO_2 leads to the formation of glycolic acid, which further hydrogenates. Subsequent dehydration yields glocoaldehyde. The calculated activation barriers are thermally accessible under prebiotic conditions.

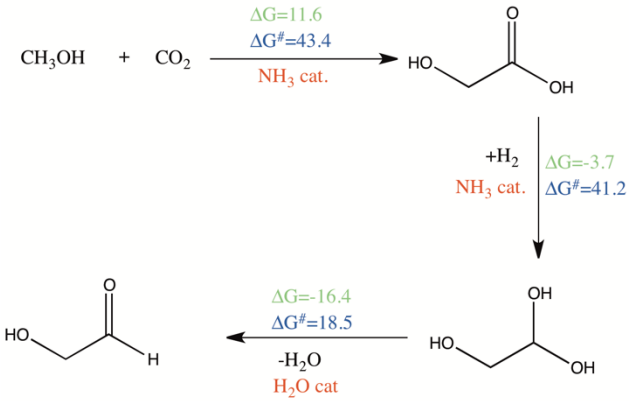


Figure 6.2 Pathway for the formation of sugar during the *ab initio* nanoreactor dynamics. The reaction energies (ΔG) are shown in green and the activation barriers ($\Delta G^{\#}$) are shown in blue, calculated at the B3LYP-D3/TZVP+COSMO (ε=80.0) level of theory with DFT. All the energies are in kcal/mol.

6.4 Conclusions

The current work shows the effect of oxygen rich atmosphere towards the formation of some important precursor molecules of RNA and proteins in the primitive earth atmosphere, as well as in interstellar ice analogues.. Taking advantage of the recently developed AINR method,[45] which has allowed us to discover new reaction pathways without any predefined reaction coordinate, we have found that the oxidizing atmosphere hindered the formation of glycoladehyde and glycine, the RNA and protein precursors in the case of prebiotic earth and interstellar space. But there is no such effect in the case of cyanamide formation. These results strongly indicate that the biomolecules had formed in the comparatively toxic atmosphere when there was no oxygen; rather the atmosphere was more CO rich. Apart from this, we have also shown the thermally feasible pathways for the formation of glycine and glycoaldehyde. We can extend our study in the future to check the direct involvement of molecular oxygen towards the inhibition of sugar and glycine formation by a complete and accurate mechanistic study.

6.5 References:

1. Goldman, N., Reed, E. J., Fried, L. E., Kuo, I. F. W. & Maiti, A. *Nature Chem.*, **2010**, *2*, 949–954.
2. Todd J. Martinez , *Nature chem.*, **2014**, *6*,1044-1048.
3. Saitta, A. M.; Saija, F. *Proc. Natl. Acad. Sci. U. S. A.*, **2014**, *111*, 13768−13773.
4. Ayyappan A., Nandi S. , Bhattacharyya D *Chemistry - A European Journal*, **2018**, 24, 4885-4894**.**
5. Hogness, T. R.; Lunn, E. G. *Phys. Rev.*, **1925**, *26*, 44–55.
6. Ganti, T.́ *Oxford Univsersity Press: Oxford UK*, **2003**.
7. Dyson, F. *Cambridge University Press: Cambridge UK*, **1999**.
8. Sutherland, D. *Nat. Rev.*, **2017**, *1*, 1−7
9. Oparin, A. I. *World Publishing: Cleveland*, **2003.**
10. Bernal, J. D. *World Publishing: Cleveland*, **1967**.
11. Sutherland, J. D. *Angew. Chem., Int. Ed.* **2016**, *55*, 104−121
12. Miller, S. L. & Urey, H. C. *Science*, **1959**, *130*, 245–251.
13. Orgel, L. E. *Science*, **1966**, *154*, 784–785.
14. Miller S. L, *Nature*, **1995**, *375*, 772–774.

15. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol*., **2004**, *39*, 99–123.
16. Powner, M.W., Gerland, B. & Sutherland, J. D. *Nature*, **2009**, 459, 239–242.
17. Ritson D., Sutherland J.D. *Nature Chem.*; **2012**, *4*; 895–899.
18. B. H. Patel, C. P., D. J. Ritson, C. D. Duffy, and J. D. Sutherland , *Nature Chem.*, **2015**, *7*, 301–307.
19. Farley, J., *Baltimore: Johns Hopkins Univ. Press*., **1977**, 225pp
20. Kamminga, H.. *Ori· gins Life*, **1988,** 1 8: I- II
21. Oparin, A. I. **1924**. Proiskhodenie Zhizni. Moscow: Moscoksky Rabotichii. 71 pp. Transl., **1967,** as appendix in Bernal, J. D. The Or (qin of Lile. *Cleveland: World*. *345*
22. Oparin, A. l., *New York: DoveL*, **1936**, 270 pp.
23. Haldane, J. S. *Ration. Ann.*, **1929**, *148*, 3-10
24. Miller, S. L., *Biochim. Biophys. Acta*, **1957**, 23, 480-87
25. Miller, S. L., *Science*, **1953**, *117*, 528-529
26. Oro, J., *Nature*, **1961**, *190*, 389-90
27. M., Gariglio, P., Oro, J. **1990**. College Park Colloq. Chern. Evol., 8th, ed. C. Ponnamperuma, F. Eirich. Hampton, Va: Deepak-Sci. Techno!. *Corp. In press*
28. Miller, S. L., Quant. Bioi., **1987**, *52*, 17-27
29. Oro, J., Mills, T., *Adv. Space Res*., **1989**, *9*, 105-120
30. Kvenvolden, K. A., Lawless, J. G., Pon· namperuma, C., *Proc. Natl. Acad. Sci. USA*, **1971**, *68*, 486-490
31. Kuhn, H., Waser, J., *Angew. Chem. Int. Ed. Engl*., **1981**, 20, 500-520.
32. Pollack, J. P., Kasting, J. F., *Icarus*, **1987**, *71*, 203-224
33. Holland, H. D., *Princeton: Princeton Univ. Press*., **1984**, 583 pp
34. Abelson, P. H., *Proc. Natl. Acad. Sci. USA*, **1966**, 55, 1365-1372.
35. Lazcano, A., Oro, J., Miller, S. L., *Precambrian Res.*, **1983**, 20, 259-282
36. Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*., **2009**, *5*, 10, 2619−2628.
37. Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.*, **2011**, *2*, 14, 1789−1793.
38. Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*. **2011**, *7*, 6, 1814−1823.
39. Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Chem. Theory Comput*. **2013**, *9*, 1, 213−221.

40. Ufimtsev, I. S.; Martinez, T. J. *Comput. Sci. Eng.*,**2008**, *10*, 26−34.
41. Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*. **2008**, *4*, 2, 222−231.
42. Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput*. **2009**, *5*, 4, 1004−1015.
43. Fischer, C. F. *Comput. Phys. Comm.***1987**, *43*, 355−365.
44. Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.*, **1980**, *102*, 3, 939−947.
45. Feller, D.; Peterson, K. A. *J. Chem. Phys.*, **1998**, *108*, 154−176.
46. Das, T.; Ghule, S.; Vanka. K. *ACS Central Science.*, **2019**, *5*, 9, 1532-1540.
47. Hu, X.; Yang, W. *J. Chem. Phys*. **2010**, *132*, 054109.
48. Pulay, P. *Chem. Phys. Lett.*,**1980**, *73*, 393−398.
49. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett*. **1989**, *162*, 165−169.
50. Ansgar, S.; Christian, H.; Reinhart, A. *J. Chem. Phys*.**1994**, *100*, 5829−5835.
51. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
52. Klamt, A.; Schuurmann, G. J. *Chem. Soc., Perkin Trans*. **1993**, *2*, 799−805.
53. Bowler, F. R.; Chan, C. K.; Duffy, C. D.; Gerland, B.; Islam, S.; Powner, M. W.; Sutherland, J. D.; Xu, J. *Nature. Chem*., **2013**, *5*, 383−389.
54. Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. *Nat. Chem*. **2015**, *7*, 301−307.
55. Schlesinger, G.; Miller, S. L. *J. Mol. Evol*., **1983**, *19*, 383−390.
56. Ritson, D.; Sutherland, J. D. *Nature. Chem*., **2012**, *4*, 895−899.
57. Eschenmoser, A.; Loewenthal, E. *Chem. Soc. Rev*., **1992**, *21*, 1−16.
58. Thaddeus, P. Philos. *Trans. R. Soc., B* **2006**, *361*, 1681−1687.
59. Eckhardt, A. K.; Linden, M. M.; Wende, R. C.; Bernhardt, B.; Schreiner, P. R. *Nature. Chem*., **2018**, *10*, 1141− 1147.
60. Oba, Y; Takano, Y; Naraoka, H; Watanabe, N.; Kouchi, A. *Nat. commun*.; **2019**, *10*, 4413.
61. Meinert, C; Myrgorodska, I; Marcellus, P. D.; Buhse, T; *Science*, **2016**, *352*, 208-212.
62. Hollis, J; Lovas, F. J.; Jewell, P. R.; *The Astrophys. Journal*; **2000**, *540*, 107-112.

Chapter 7

Summary and Future Outlook

Chapter 7

Summary and Future Outlook

7.1 Focus of this Thesis

Studies in the areas of cosmology,[1-3] prebiotic chemistry[4-10] and interstellar space chemistry[11-13] have been carried out to understand the chemical evolution of life on the primitive Earth, as well as the evolution of small molecules after the Big Bang.[14] In this thesis work, by employing state-of–the-art computational methods, especially *ab initio* molecular dynamics (AIMD), with the nanoreactor approach and quantum chemical calculations with RICC2, RIMP2 and DFT, attempts have been made to explore some unsolved questions in the area of cosmology and prebiotic chemistry. We have concentrated on this area of research not only because of the general interest but also to explore or discover new chemistry through computational means, because there are several questions that have not been completely answered. Hence, in this thesis, three critical areas where messy chemistry could play an important role: a) cosmology, b) prebiotic chemistry and c) interstellar space chemistry, have been demonstrated., The results can be summarized as follows:

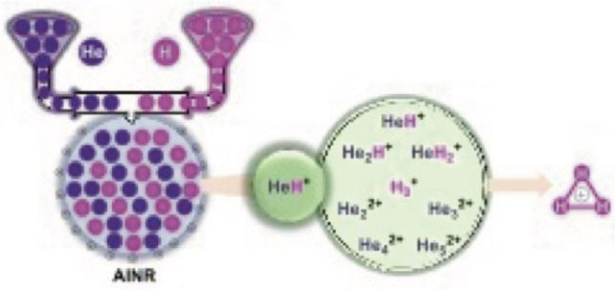
(i) *Ab initio* nanoreactor dynamics followed by QM calculations have been performed to shed light on the cosmological appearance of HeH^+ and H_3^+. To study the effect of the ionized atmosphere in the early universe, we have performed simulations on systems containing atoms/ions of helium and hydrogen, and have obtained reaction profiles by varying their mixture ratio and the total charge of the system. Our calculations show that HeH^+ is certainly the initial molecule to be formed, and thereafter, some fleeting intermediates such as He_2H^+, HeH_2^+, H_2^+, and He_n^{2+} (n=2-5) are formed, with the moderately stable H_3^+ and H_2 finally being created in the AINR spherical vessel.

(ii) Later, we have studied the formation of life building blocks through thermal collisions between HCN and H_2O. Quantum chemical calculations with reactive AIMD simulations have been performed to explore the chemistry behind the origin of life on

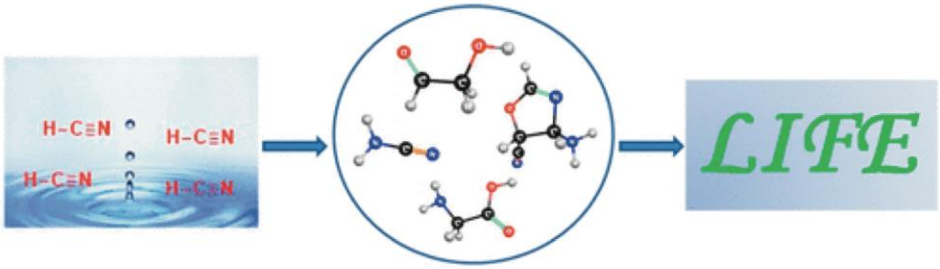
the early earth. Our studies reveal that, even under a certain number of constraints present on the early earth, the thermal interaction between two small molecules: HCN, the source of carbon and nitrogen, and H_2O, the source of oxygen, was enough to create the precursor molecules to RNA and proteins. Through the aid of *ab initio* nanoreactor (AINR) dynamics, we have shown that cyanamide, glocoaldehyde, glycine, an oxazole derivative and many more important small organic molecules could have been formed in a "single pot" reaction.

(iii) Furthermore, the AINR has also been instrumental in finding the thermodynamics and kinetics of each elementary step in the entire reaction network when HCN and H_2O are interacted together. In order to study the mechanistic pathways, we have done full QM calculations with the aid of RI-CC2, RI-MP2 and DFT methods. The calculated pathways have been found to be feasible under the prebiotic conditions. The activation energy barrier for each of the elementary steps has been calculated to be ~40.0 kcal/mol, which would have been achievable around the near boiling temperature of water in the prebiotic oceans and also would have been possible without needing metal catalysts.

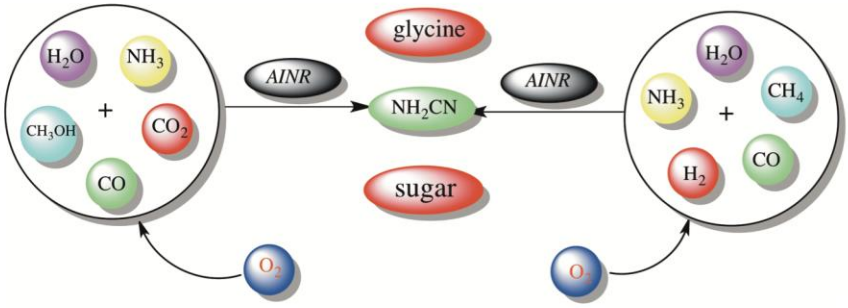
(iv) Finally, we have studied the effect of an oxidizing atmosphere towards the formation of different precursor molecules of life's building units in the prebiotic earth atmosphere, as well as in interstellar space. For studying such chemistry we have performed a series of AINR dynamics by taking the Urey-Miller gaseous composition (NH_3, CO, CH_4, H_2, H_2O) and the interstellar ice matrix compostion (NH_3, CH_3OH, H_2O, CO, very little CO_2); in order to mimic the early earth and interstellar space atmospheric conditions respectively. During the simulations, several of life's building units were seen to have formed. Furthermore, we have introduced molecular oxygen and carbon dioxide into the mixture of both Urey-Miller and interstellar ice compositions, in order to do the AINR dynamics and have analyzed the simulation trajectories to check the formation of the precursors towards RNA and proteins. Furthermore, full QM calculation have been performed to check the mechanistic pathways. Interestingly our results suggest that an oxygen rich atmosphere would have inhibted the formation of the precursors molecules (glycoaldehyde, glycine) of life's building blocks.



Chemical reactions at the beginning of the universe



Chemical reactions during the origin of life



Effect of O_2 during chemical evolution

Figure 7.1 Representation of the research work presented in the thesis.

7.2 Computational Methods

All the nanoreactor AIMD simulations were performed with the TeraChem 1.9 software package.[15-16] The Born−Oppenheimer potential energy surface was calculated using the Hartree−Fock (HF) electronic wave function and the 3-21g(d) Gaussian basis set, for solving cosmology, prebiotic chemistry and interstellar chemistry problems. The AINR boundary conditions, the temperature for the dynamics, the ratio of the reaction mixtures taken and other parameters in the molecular dynamics simulations have been discussed in the Computational Details section of each chapter. The calculations with DFT have been performed using the Turbomole 7.0 suites of programs.[17] The geometry optimizations have been done with the B3LYP three parameter hybrid density functional and TZVP basis set in the case of prebiotic chemistry and interstellar chemistry. Dispersion corrections (D3) and solvent corrections (COSMO) have been included. Furthermore, in order to make our data more reliable and also to refine the energies, the single point energy calculation of all the transition states and connecting reactants, intermediates and products for HCN and water chemistry have been then done further at the RI-CC2/TZVP+COSMO (ε=80.0) and RI-MP2/TZVP+COSMO(ε=80.0) levels of theory. Furthermore, we have also done the calculations for prebiotic chemistry with the M06-2X hybrid functional and the 6-311++g(d,p) basis set by employing the Gaussian09 software package.[18]

7.3 Future Aspects

Insights achieved from the work represented in this thesis shed light on critical areas of research and are also likely to boost experimentalists for designing or modeling systems based on prebiotic or interstellar environments. The proposed mechanistic pathways towards the formation of different life building units and HeH^+ chemistry can lead experimentalists into studies of catalyst free reactions, as well as simplify the reaction optimization conditions. Therefore, the work introduced in this thesis is a significant step forward in the area of catalyst free prebiotic chemistry, cosmology and interstellar space chemistry. Furthermore, we can extend our research areas to solve several interesting problems in the near future. Some future aspects of this thesis work are summarized below.

7.3.1 Insights into dissociative electron recombination study on HeH^+ in presence of electric and magnetic field

The way the universe, and all the elements, came into being is one of the fascinating questions of science. Attempts to answer this question has led to the Big Bang theory. The low mass particles recombined in reverse order of their ionization potential during the Big Bang nucleosynthesis. Before the recombination of hydrogen, He^{2+} and He^{+} joined first with free electrons due to their higher ionization potential. Around that point, in this low-density climate, the universe's first molecular bond formed between neutral helium particles with protons by the radiative association process in the HeH^{+},

$(He + H^{+} \rightarrow HeH^{+} + h\nu)$.

Recently in space, HeH^{+} has been detected and it is thought to be the first formed molecule in the early universe. By using a cryogenic ion storage ring combined with an electron beam, Novotny *et al.* have proposed[19] rotational state-specific rate coefficients for the dissociation of HeH^{+}. They have found a significant decrease in dissociation rate at the lowest rotation level of HeH^{+}.

As a computational chemist, we can tackle this experimental finding from a quantum chemical perspective. To do this, we have to model an electron (not a simple task) and perform AIMD simulations with HeH^{+}. Also, we can see the effect towards the dissociation of HeH^{+} molecule upon application of an external electric and magnetic field. The outcome of the simulation trajectory might provide some insights in this area.

7.3.2 Methane or hydrocarbon based origin of life on Titan

The atmosphere of Titan is the layer of gases encompassing this biggest moon of Saturn. Titan's lower environment is fundamentally made up of nitrogen (94.2%), methane (5.65%), and hydrogen (0.099%). Apart from these gases, different hydrocarbons, such as ethane, acetylene, diacetylene, propane, methylacetylene, PAHs are also present. Moreover different gases such as carbon dioxide, hydrogen cyanide, cyanoacetylene, carbon monoxide, acetonitrile, argon, and helium are also present. However, there is no water on Titan. Very recently NASA's researchers could identify a molecule, known to be cyclopropenylidene or C_3H_2, in Titan's climate that has never been identified in some other atmosphere. Researchers think that this basic carbon-based

molecule might be an antecedent to more unpredictable mixtures that could be the basis of conceivable life on Titan.

Therefore, unlike the prebiotic earth, we can assume that in Titan there is a sea of hydrocarbon. However, the hazy atmosphere of the Titan (red surface) would have made it difficult for high-energy photons to reach the Titan surface. So photochemistry is not possible under this condition. High-pressure conditions (surface pressure of Titan is 50% higher than Earth) is important under this scenario.

It might be possible the molecules present on Titan's surface could be the same ones that formed the building blocks of life on Earth. Also, scientists suspect that around 3.8 billion years ago, earth's atmosphere could have been similar to that on Titan today. A computational strategy involving using reactive molecular dynamics (high pressure molecular dynamics, shockwave MD, nanoreactor dynamics) can shed light on what potential building blocks towards an extraterrestrial form of life could exist on Titan. One can then speculate on hydrocarbon based forms of life on Titan and Titan-like bodies in the universe

7.4 References

1. Ganti, T.́ *Oxford Univsersity Press: Oxford UK*, **2003**.
2. Dyson, F.; *Cambridge University Press: Cambridge UK*, **1999**.
3. Sutherland, D. *Nature. Rev*. **2017**, *1*, 1−7
4. Steel, M.; Penny, D. *Nature*, **2010**, *465*, 168−169.
5. Oparin, A. I.; *World Publishing: Cleveland*, **2003.**
6. Bernal, J. D.; *World Publishing: Cleveland*, **1967**.
7. Sutherland, J. D. *Angew. Chem., Int. Ed.*, **2016**, *55*, 104−121
8. Saslaw WC, Zipoy D., *Nature*, **1967**, *216*, 976
9. Schmeltekopf A.L, Fehsenfeld F. F, Ferguson E. E., *Ap. J.* 1967, 148:L155
10. Puy D, Alecian G, Le Bourlot J, Leorat J, Pineau Des Forˆets G. *Astron Astrophys*., **1993**, *267*, 337
11. Oba, Y; Takano, Y; Naraoka, H; Watanabe, N.; Kouchi, A., *Nat. commun*.; **2019**, *10*, 4413.
12. Meinert, C; Myrgorodska, I; Marcellus, P. D.; Buhse, T; *Science*, **2016**, *352*, 208-212.

13. Hollis, J; Lovas, F. J.; Jewell, P. R.; *The Astrophys. Journal*; *2000*, *540*, 107.
14. Gusten, R.; Wiesemeyer, H.; Neufeld, D.; Menten, K. M.; Garf, U. U.; Jacobs, K.; Klein, B.; Ricken, O.; Risacher, C.; Stutzki, J. *Nature* **2019**, *568*, 357–359
15. Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Chem. Theory Comput.*, **2013**, *9*, 213−221.
16. Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J., *J. Chem. Theory Comput.*, **2013**, *9*, 213−221.
17. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
18. Gaussian 09, Revision **E.01**; Frisch , M. J.; Trucks, G. W.; Schlegel, H. B.;Scuseria,G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;Nakatsuji, H.; Caricato, M.; Li, X.; Hratchians, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.;Jr.,Peralta, J. E.; Ogliaro, F.,Bearpark, M.; Heyd, J. J. E.; Brothers,Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Raghavachari, J.K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., *Wallingford CT*, **2009.**
19. Becke, A; Saurabh, S; Kalosi, A; Paul, D; Wilhelm, P; Novotny, O; *Science*, **2019**, *365*, 676-679.

ABSTRACT

Name of the Student: Tamal Das **Registration No.:**10CC16J26001

Faculty of Study: Chemical Science **Year of Submission:** 2021

AcSIR Academic Centre/CSIR Lab: CSIR-National Chemical Laboratory

Name of the Supervisor: Dr. Kumar Vanka

Title of the Thesis: Insights into Messy Chemistry Related to Cosmology and Origin of Life Obtained by Employing State-of-the-art Computational Methods

How life began on the early earth, as well as how small molecules and ions were first formed at the beginning of the universe are some the biggest unsolved questions in science that have intrigued researchers over time. However, no clear answers have yet been received for researchers in the fields of cosmology, prebiotic chemistry, interstellar chemistry as well as astrochemistry. From previous studies, it has become clear that the creation of the universe and the origin of life are not the result of any single event. Rather, they are likely the product of highly complex or "messy" chemical processes. In this chapter, an overview has been provided on how messy chemistry plays a role in cosmology, as well as in prebiotic chemistry.

In this thesis work, by employing state-of–the-art computational methods especially *ab initio* molecular dynamics (AIMD) with nanoreactor approach and quantum chemical calculations with RICC2, RIMP2 and DFT have been carried out to explore some unsolved questions in the area of cosmology and prebiotic chemistry. We have concentrated into this area of research not only because of the general interest and scientists have done hundreds of chemical experiments into this fields but also explore or discover new chemistry through computational perspective because there are several questions which has not been completely explored. Hence, in this thesis, three critical areas of the Universe: a) cosmology, b) prebiotic chemistry and c) interstellar space chemistry where messy chemistry could play an important role in the reaction network, have been demonstrated.

Details of the publications emanating from the thesis work

1) List of publication(s) in SCI Journal(s) (published & accepted) emanating from the thesis work

(i). **T. Das,** S. Ghule and K. Vanka, Insights into the Origin of Life: Dit it Begin from HCN and H_2O? *ACS Cent. Sci.* **2019**, *5*, 1532-1540.

2) List of Papers with abstract presented (oral/poster) at national/international conferences/seminars with complete details.

(i) Oral presentation entitled " How Life Begins from HCN and H_2O- An *ab initio* Nanoreactor Dynamics Approach" Research Foundation day CSIR NCL Pune, 2018

Abstract

The ab initio nanoreactor, developed by the group of Todd Martinez (Nature Chemistry 2014, 6, 1044), is a powerful new tool in *ab initio* molecular dynamics (AIMD), because it can be employed to discover new chemistry, something that was not possible earlier in computational chemistry. Taking advantage of this method, we have explored an interesting and important question: whether the ingredients of life - the precursor molecules for RNA and proteins - could have formed from only one source for carbon and nitrogen, HCN, and only one source for oxygen, H2O.The results that we have obtained indicate that this indeed is a feasible possibility. Beginning from a mixture of only HCN and H2O molecules, important precursors for RNA and proteins

(ii) Presented a poster entitled "Insights into the Origin of Life: Did it Begin from HCN and H_2O?" at Science day, CSIR NCL Pune, 26-27 Feb 2019 **(Recieved a Best Poster Award)**.

Abstract

The seminal Urey−Miller experiments showed that molecules crucial to life such as HCN could have formed in the reducing atmosphere of the Hadean Earth and then dissolved in the oceans. Subsequent proponents of the "RNA World" hypothesis have

shown aqueous HCN to be the starting point for the formation of the precursors of RNA and proteins. However, the conditions of early Earth suggest that aqueous HCN would have had to react under a significant number of constraints. Therefore, given the limiting conditions, could RNA and protein precursors still have formed from aqueous HCN? If so, what mechanistic routes would have been followed? The current computational study, with the aid of the *ab initio* nanoreactor (AINR), a powerful new tool in computational chemistry, addresses these crucial questions.

(iii) Presented a poster entitled, "Insights into the Origin of Life: Did it Begin from HCN and H_2O?" at the Conference on Recent Trends in Catalysis (RTC 2020), NIT Calicut, 26-29 Feb 2020.

Abstract

The seminal Urey−Miller experiments showed that molecules crucial to life such as HCN could have formed in the reducing atmosphere of the Hadean Earth and then dissolved in the oceans. Subsequent proponents of the "RNA World" hypothesis have shown aqueous HCN to be the starting point for the formation of the precursors of RNA and proteins. However, the conditions of early Earth suggest that aqueous HCN would have had to react under a significant number of constraints. Therefore, given the limiting conditions, could RNA and protein precursors still have formed from aqueous HCN? If so, what mechanistic routes would have been followed? The current computational study, with the aid of the *ab initio* nanoreactor (AINR), a powerful new tool in computational chemistry, addresses these crucial questions.

3) A copy of all SCI publication(s), emanating from the thesis, to be bound at the end of the thesis.

ACS central science

Research Article

Cite This: ACS Cent. Sci. 2019, 5, 1532−1540

http://pubs.acs.org/journal/acscii

Insights Into the Origin of Life: Did It Begin from HCN and H_2O?

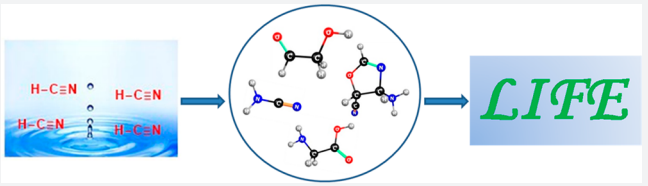
Tamal Das,[†,‡] Siddharth Ghule,[†,‡] and Kumar Vanka*[,†,‡]

[†]Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Dr. Homi Bhabha Road, Pashan, Pune 411008, India

[‡]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

Supporting Information

ABSTRACT: The seminal Urey−Miller experiments showed that molecules crucial to life such as HCN could have formed in the reducing atmosphere of the Hadean Earth and then dissolved in the oceans. Subsequent proponents of the "RNA World" hypothesis have shown aqueous HCN to be the starting point for the formation of the precursors of RNA and proteins. However, the conditions of early Earth suggest that aqueous HCN would have had to react under a significant number of constraints. Therefore, given the limiting conditions, could RNA and protein precursors still have formed from aqueous HCN? If so, what mechanistic routes would have been followed? The current computational study, with the aid of the ab initio nanoreactor (AINR), a powerful new tool in computational chemistry, addresses these crucial questions. Gratifyingly, not only do the results from the AINR approach show that aqueous HCN could indeed have been the source of RNA and protein precursors, but they also indicate that just the interaction of HCN with water would have sufficed to begin a series of reactions leading to the precursors. The current work therefore provides important missing links in the story of prebiotic chemistry and charts the road from aqueous HCN to the precursors of RNA and proteins.



INTRODUCTION

How life originated[1−13] on Earth is one of the most fundamental questions of science, and has generated considerable interest. Research and discussion has resulted in two principal positions that are held today: the "RNA World" hypothesis[14−19] and the "metabolism-first" principle.[20−23] According to the RNA World hypothesis, life on Earth originated from the self-replicating molecules of ribonucleic acid (RNA),[24−26] which is the polymeric form of activated ribonucleotides.[27−29] The metabolism-first principle argues, on the other hand, that simple metal catalysts were present in the water in early Earth and aided in creating a soup of organic building blocks that subsequently formed the biomolecules necessary for life. The RNA World hypothesis has gained increased acceptance in recent times, with several experimental studies[26−28,30−35] indicating how hydrogen cyanide (HCN), known to exist on prebiotic Earth, could have been the starting point of many synthetic routes leading to the formation of RNA and protein precursors (see Figure 1A).

However, questions remain as to how HCN could have actually functioned in prebiotic conditions. As the famous Urey−Miller experiments have shown, HCN would have formed in the reducing atmosphere that existed during prebiotic times,[36] after which it would have condensed into the oceans.[13,37] HCN has a low boiling point, but at high pH (8−10), it is possible for it to exist in aqueous solution, even if the temperature of the water is 80.0−100.0 °C. However, since the hazy atmosphere[38] of the Hadean Earth would have made it difficult for high-energy photons to reach the Earth's surface (much like the red surface of Titan today, because of a similar haziness in the atmosphere), a lot of the reactions shown in Figure 1, which depend upon photochemistry or an electric spark, may not have been possible for aqueous HCN. Hence, the more plausible alternative would have been thermochemistry. It is possible that temperatures at the surface of the water bodies of early Earth (3.5−4.0 billion years ago) would have been about 80.0−100.0 °C,[39] which suggest favorable conditions for thermochemistry, but if thermochemistry predominated in the oceans of early Earth, it could be argued that hydrolysis would have taken precedence over the polymerization of HCN. This is because HCN polymerization would have had to begin with HCN dimerization and the subsequent reaction of the product with more HCN molecules. In other words, the polymerization of HCN would have required a series of second-order reactions in HCN, while the competing hydrolysis of HCN would have simply required the HCN collision with the surrounding solvent water molecules. Indeed, previous studies[34,40] have shown that in dilute aqueous concentrations of HCN, hydrolysis is favored over oligomerization.

Then, there is also the issue of too-high temperatures: experiments have shown[41,42] that at temperatures above 100.0 °C, decomposition of the formed RNA and protein precursors would occur. Therefore, the reactions would have had to happen around 100.0 °C,[43,44] which indicates that the barriers (ΔG values) of the reactions of monomeric HCN in water could have only been about 40.0 kcal/mol: previous computa-

Received: May 28, 2019
Published: August 7, 2019

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

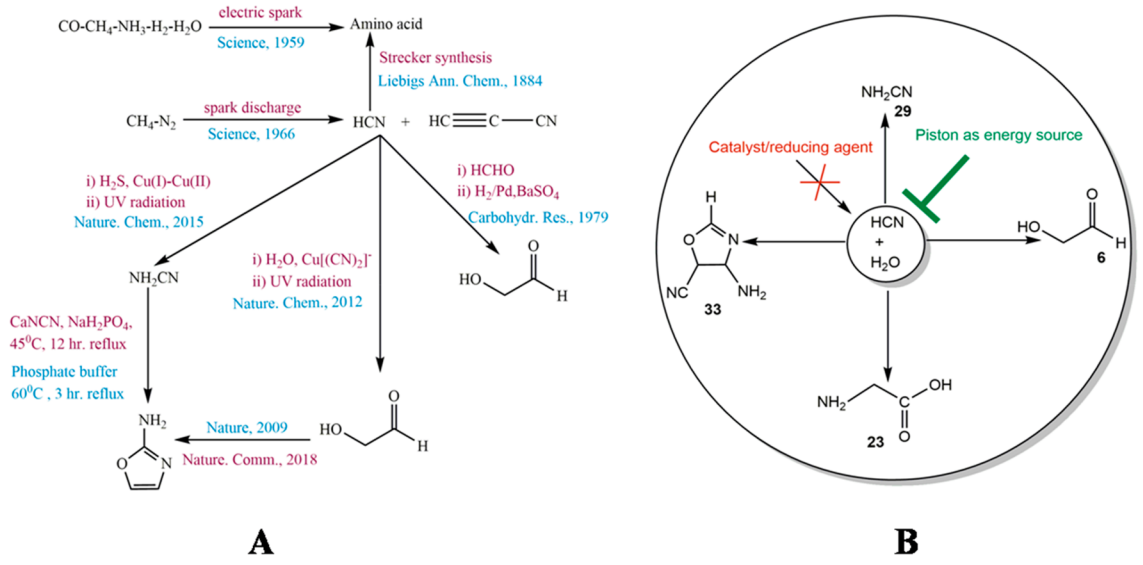


Figure 1. (A) Previously synthesized RNA and protein precursors (amino acids, cyanoacetylene, cyanamide, glycoaldehyde, and 2-amino-oxazole). (B) The ab initio nanoreactor (AINR) approach, yielding RNA and protein precursors, beginning from only two different reacting molecules, HCN and H_2O, obtained in "one-pot", under the same reaction conditions.

tional studies[45−49] have shown that chemical reactions occurring at temperatures of around 100.0 °C have barriers in the region of 40.0 kcal/mol. Barriers higher than 40.0 kcal/mol would have led to much slower reactions (or no reactions) at 100.0 °C. Slower reactions may be possible at higher barriers, but this would lead to the possibility of other side reactions also becoming competitive and causing significant reduction in the formation of desired products.

Hence, for the RNA World hypothesis to be true, there are several constraints that have to be kept in mind: (i) thermal, not photochemical conditions, (ii) reactions where monomeric and not polymeric HCN would predominate, (iii) without mediation from metal catalysts, (iv) at temperatures not exceeding 100.0 °C, and (v) having reactions with free energy barriers not exceeding 40.0 kcal/mol. To this list, one could add (vi) the need to avoid chemical processes involving the protonation of substrates, since HCN, with a pK_a of 9.31, is a weak acid and would have largely remained in undissociated form in solution. The protonation of water would also have been suppressed since the pH of water has been estimated to be between 8.0 and 9.0 in prebiotic times at the surface of the ocean.

This list of conditions appears formidable and leads to the important question: could life have begun under these circumstances? This current work attempts to answer this question, through the agency of the ab initio nanoreactor (AINR).

The AINR method, recently developed,[50] allows one to obtain reaction pathways and products without controlling the chemical system in any way.[51,52] This represents a major shift in what one can do with computational chemistry, because, using the AINR, one can now *discover* new reactions, completely independent of experimental input. This was demonstrated by Martinez and co-workers[50] when they found entirely plausible, new pathways for the formation of amino acids, from a computational re-enactment of the Urey−Miller experiment.[53] In the current work, we have conducted full quantum mechanical molecular dynamics (MD) simulations on systems employing the AINR approach on systems containing a mixture of molecules of HCN and H_2O. The goal has been to follow the chemical reactions that can occur through collisions between the molecules and observe what new species are formed as a result. In short, our objective has been to perform the equivalent of an experimental study while satisfying the conditions outlined in (i−vi) above. Remarkably, we have found, as will be shown in the Results and Discussion, that just the interaction of HCN and H_2O was sufficient to eventually lead to the formation of the experimentally reported precursor molecules to RNA and proteins:cyanamide,[27,30−32,54,55] glycolaldehyde,[27,30,31,56−61] an oxazole derivative,[27,62,63] and the amino acid glycine[36,53,64−68] (as shown in Figure 1B). Furthermore, analysis of the data allowed us to determine the mechanistic pathways by which HCN and H_2O reacted together to yield intermediates and, eventually, the RNA and protein precursors. We subsequently subjected these pathways to a full static quantum chemical study with density functional theory (DFT) and thus obtained all the barriers ($\Delta G^{\#}$) for the reactions involved in these processes, as well as the energies (ΔG) of the reactions. As will be discussed in the Results and Discussion, this has led to results that not only reveal interesting pathways for the formation of the precursor molecules beginning from aqueous HCN but also indicate that these mechanistic routes would have been thermodynamically and kinetically feasible.

RESULTS AND DISCUSSION

The AINR approach makes use of collisions between the molecules of HCN and H_2O, and this provides the energy required to cross the activation barriers for each of the elementary steps of the reactions. The simulations have been done on systems having nearly homogeneously mixed HCN and H_2O molecules as the starting reactants. Sixteen H_2O and 15 HCN molecules were taken together, and the system was allowed to evolve for 750 ps. For more details, please see the Computational Methods section. Collisions between the molecules gave rise to new species. It should be noted that homogeneous mixtures of HCN and H_2O do not represent the exact ratios of HCN and H_2O present in a localized region of

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

the early Earth oceans or hydrothermal vents. The reason such mixtures were employed was to maximize the possibility of interactions between HCN and H_2O in the AINR. This would increase the probability of obtaining different products during the simulations. The goal of the AINR simulations was to obtain mechanistic pathways for the formation of different RNA and protein precursors beginning from HCN and H_2O, pathways that could then be studied carefully with a static DFT and QM approach to ascertain their feasibility. Studying homogeneous mixtures of HCN and H_2O afforded the best possibility of realizing this goal.

Figure 2 below illustrates how a system starting with a mixture of HCN and H_2O molecules evolves with time. An

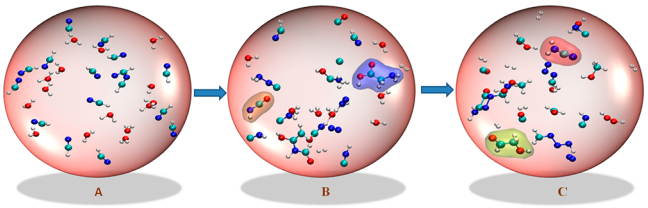


Figure 2. Snapshots of AINR simulations. (A) the beginning, 0.0 ps: only HCN and H_2O present. (B) after 100 ps, glycine (blue surface) has formed, along with molecules such as isocyanic acid (pale yellow surface). (C) after 250 ps: glycoaldehyde (green surface) and cyanamide (orange surface) have formed, along with other oligomeric species. Color scheme: oxygen: red, carbon: teal, hydrogen: gray and nitrogen: blue.

mp4 file (Supporting Information, Movie S1) of a movie made of a part of an AINR simulation is included in the Supporting Information.

The AINR approach thus leads to the discovery of new species from the starting compounds, and the analysis of the data through the connectivity graph (shown in Figure S1, Supporting Information) allows the exploration of new mechanistic pathways. The next section discusses the results that have been obtained by this approach.

Analysis of the Reaction Pathways Leading to the Formation of Specific Compounds. *Formation of the Protein Precursor: Glycine.* The use of the nanoreactor produces as an output many different pathways to new species from the starting reactants. In most of the pathways, formaldehyde **3**, urea **26**, formaldimine **2**, and glycolonitrile **4** were seen to be formed as intermediates. This suggests that these species were the key intermediates en route to the formation of the target molecules, as has also been noted by experimentalists.[29,30,35,60,63,69,70] Apart from this, a lot of diverse organic species were also observed to have formed during the simulations (Figure S2). Moreover, small molecules such as CO **13**, CO_2 **11**, and H_2 were produced (Figure 3), and these were seen to take part in the synthesis of comparatively larger organic molecules. Figure 3 below describes how the relevant intermediates formic acid, CO, and CO_2 are formed from HCN and H_2O, leading from HCN, **1**, through the intermediate species **7**, formamide **8**, formic acid **9**, to carbon monoxide, CO, **13**. The complete free energy profile is shown in Figure S3, Supporting Information.

Figure 3 shows that species such as ammonia and dihydrogen were created in the AINR from the interaction between the HCN and water, and they turned out to be important substrates in subsequent reactions. This is interesting because it suggests that HCN and water would

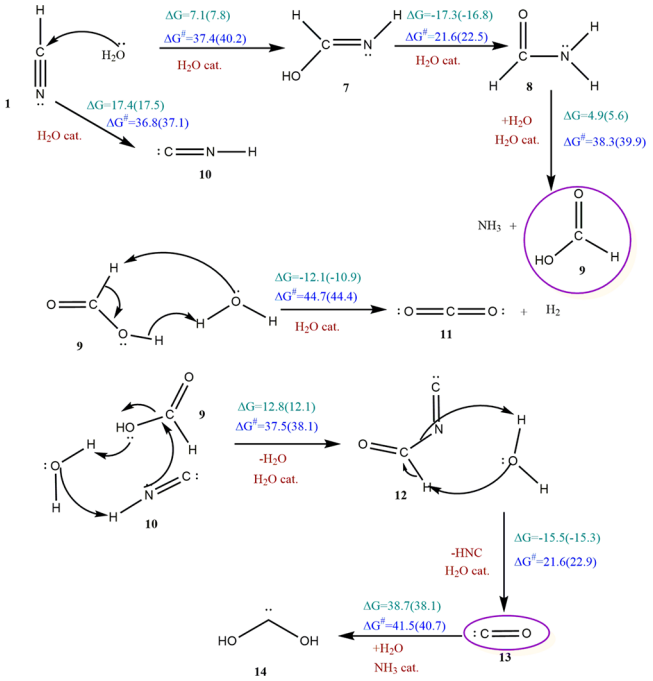


Figure 3. Sequence of elementary reaction steps derived from the AINR: the formation of HCOOH, CO_2, and CO starting from HCN and H_2O. Molecules labeled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO-(ε = 80.0) and the B3LYP-D3/TZVP+ COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses) levels of theory in kcal/mol.

have created all the necessary reactants in subsequent steps. However, it could be argued that since the concentration of HCN would have been low such species would have been formed in low concentrations as well, which would have further reduced the yield of the subsequent products. The counter argument to this is that ammonia and other reactant species were also present separately in the oceans at that time because the protective haze of the Titan-like atmosphere would have prevented the photochemical degradation through UV of molecules such as ammonia in the atmosphere, and such molecules could have dissolved separately in the oceans as well and could be thus available for the reactions shown later in Figure 4.

Now, as we continue along this path, we find that important intermediate species: formaldehyde **3** and formaldimine **2** are formed during glycine **23** synthesis via several elementary steps (see Figure 4A and energy profile in Figure S5). These intermediates lead to the formation of glycine **23**, the precursor to proteins. A perusal of the three most feasible pathways found for glycine formation (two are shown in Figure 4B and one in Figure S7, Supporting Information) shows that **2** is present as an intermediate in all the cases. All the pathways are seen to proceed via stepwise elementary reaction steps. In one of the pathways, **2** reacts with hydrogen cyanide **1** to produce **20**, which, through further stepwise hydrolysis, leads to **23**. This is the well-known Strecker synthesis pathway.[64] That the AINR finds the same is gratifying and can serve as a validation for the computational approach. What is also satisfying is that the pathway does not involve protonation, which therefore agrees with condition (vi), mentioned earlier. We do note, though, that the AINR has also found another pathway,

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

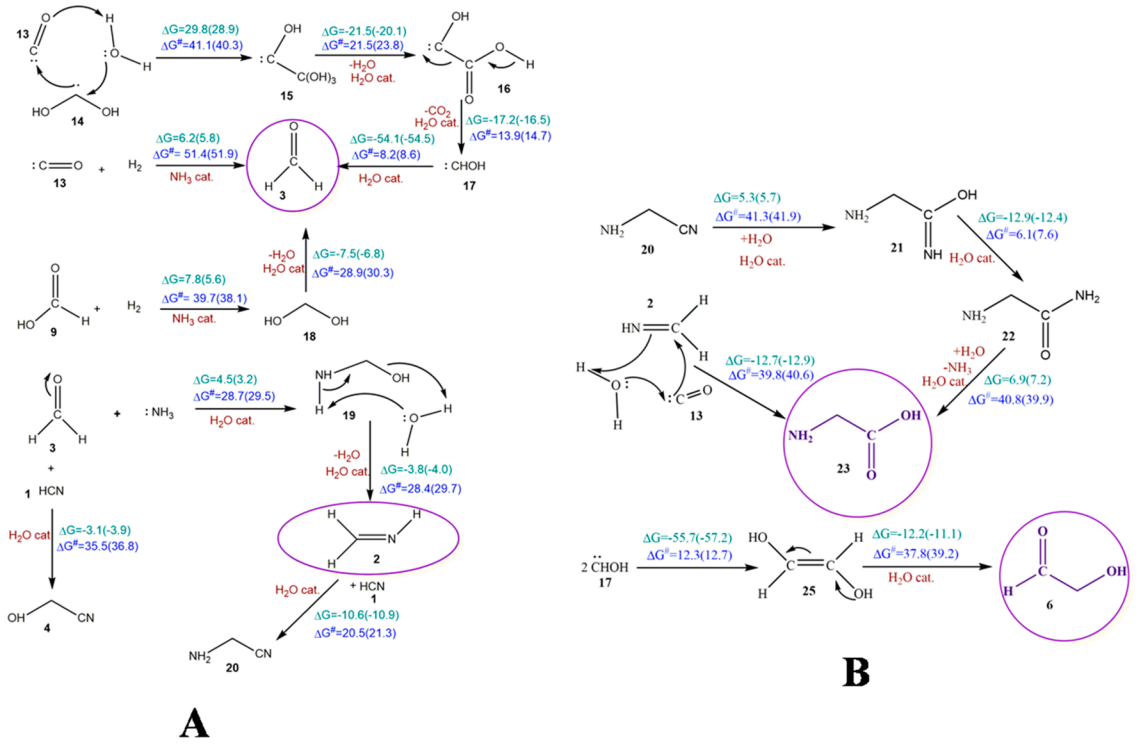


Figure 4. (A) The sequence of elementary reaction steps derived from the AINR: the formation of formaldehyde, formaldimine, glycolonitrile, and aminoacetonitrile. (B) The formation of the target species: glycine and sugar. The values are in kcal/mol.

involving a trimolecular reaction between **13**, **2**, and water, with a five-membered transition state, which also leads to **23** (encircled in violet in Figure 4B).

Pathways for RNA Building Units. The previous section discussed feasible pathways for the formation of the important protein precursor: glycine **23**. In this section, we discuss how the AINR also yields feasible pathways for the formation of important RNA precursors such as cyanamide **29**, glycoaldehyde (sugar) **6**, and the oxazole derivative **33**. The experimental synthesis of the sugar **6**, one of the most important RNA precursors, is challenging, but several reports have emerged recently where this has been achieved. Recently, Sutherland and co-workers[60] synthesized sugar from HCN and H_2O in the presence of a copper cyanide catalyst, through a photoredox cycle. They proposed a mechanistic pathway for the formation of the sugar **6** in the absence of the copper cyanide catalyst (Figure S8A, Supporting Information), involving two reduction steps, through which HCN **1** would be reduced to **2** and **4** to imine **5** in the presence of an H_2O catalyst molecule. We have calculated the barriers for these two processes (at the same level of theory as the mechanistic pathways investigated in the current work) and found them to be 83.2 and 85.8 kcal/mol respectively (see Figure S8B, Supporting Information). Interestingly, the results obtained from the current computational studies reveal a completely different pathway for sugar formation, avoiding the reduction steps. The feasible pathway derived from the nanoreactor for the formation of **6**, where two molecules of hydroxymethylene carbene **17**, the formation of which was discussed in the previous section (see Figure 4A), dimerize with a low barrier to form **24**, which further tautomerizes and leads to the formation of **6** (encircled in violet in Figure 4B). This pathway is found to be facile, with reduction not involved in any of the steps, with the slowest step seen to be ∼37.0 kcal/mol, which is considerably lower (by more than 40.0 kcal/mol) than the barriers for the pathways (83.2 and 85.8 kcal/mol) that have been proposed in the literature.[60] The implication of this is that the process of forming sugar **6** would not have needed the presence and intervention of metal catalysts but would have been possible under thermal conditions, at temperatures of 80.0−100.0 °C. For other pathways, please see Figure S9, Supporting Information.

It is also interesting to note that Schreiner's group has recently reported[70] that carbene **17** is the intermediate en route to the formation of **6**. The fact that the AINR discovers the same intermediate, carbene **17**, to be important for the formation of **6** as this recent, independent, experimental study by Schreiner and co-workers is quite remarkable. We note, however, that the pathway that has been proposed for the transformation of **17** to **6** in the absence of solvent and base by Schreiner and co-workers (shown in Figure S10, Supporting Information) is different from the one that we have discussed here. Since both the solvent and the base are relevant in our simulations (and in early Earth), it is quite likely that the facile pathways reported here transforming **17** to **6** represent avenues by which this important RNA precursor was formed in prebiotic Earth.

Cyanamide **29**, another precursor of RNA, is formed via two important intermediate species isocyanic acid **28**, or urea **26**. **28** reacts with ammonia in the presence of H_2O as a proton shuttling catalyst to produce **27**, which is further dehydrated in the presence of an NH_3 molecule acting as a catalyst, leading to the formation of **29** (shown in Figure 5 and the energy profile in Figure S12). There are several other possible pathways for the formation of **28**, which have been discussed in Figure S13, Supporting Information.

During the nanoreactor simulations, apart from simple and complex acyclic organic compounds, numerous cyclic com-

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

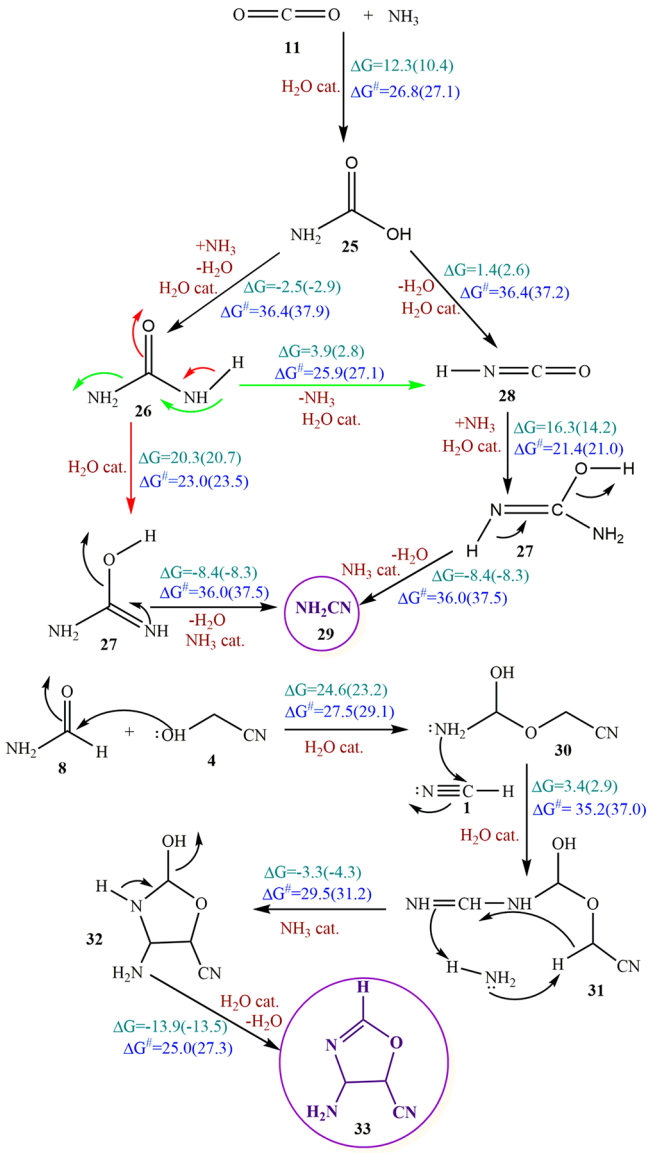


Figure 5. Formation of the target species: cyanamide and the oxazole derivative. The values are in kcal/mol.

pounds such as oxazole, imidazole, as well as isoxazole derivatives were also seen to have formed (see Figure S14, Supporting Information). Among these heterocyclic compounds, oxazole derivatives are among the more important, because 2-amino-oxazole is known to be an important precursor for RNA synthesis.[27,62,63] In our current work, we have shown a very feasible pathway for the formation of one of the oxazole derivatives **33** (encircled in violet in Figure 5) which is formed during the reaction between **8** with **4** (shown in Figure 5, with the energy profile in Figure S5, Supporting Information). For more information on the formation of cyclic products, please see Figures S7 and S15, Supporting Information.

Implications of the Current Work. HCN, as has been noted in the literature, may have occupied a "unique position in terrestrial pre-biological chemistry".[40] The current work shows that just the interaction between HCN and water as the starting reactants would have been sufficient to eventually lead to the precursors of RNA and proteins. This is significant because it shows that the reactions could have happened ubiquitously in the water bodies all over the Earth. What is also important is that all the conditions specified as (i–vi) earlier were adhered to during the simulations.

There are, however, important questions that need to be addressed. First, there is the issue of low concentrations of HCN in water in early Earth conditions, which would have reduced the concentration of the subsequently formed RNA and protein precursors. A resolution to this problem is suggested by a recent molecular dynamics report which indicates that in dilute systems the HCN concentration is an order of magnitude larger in the surface layer than in the bulk liquid phase.[71] Such HCN concentration effects at the surface of water bodies would have facilitated the chemistry described here. Furthermore, certain hydrothermal vents at the bottom of the ocean are in the vicinity of cold seawater, as well as ice. Recent reports suggest that HCN could be stabilized and concentrated at water–ice interfaces as well.[72,73] Water containing this more concentrated HCN could then have seeped into the hydrothermal vents in the vicinity of the cold water–ice and undergone hydrolysis at higher temperatures inside the vents.

The other issue is with regard to the eventual products of the HCN hydrolysis. Our results show that the formation of precursor molecules of RNA and protein would have been feasible, but the question then is would these precursors have been formed in sufficient concentrations to then react with each other, in order to lead to greater complexity? One solution that can be provided to this problem is to invoke the idea of a "warm little pond" that had been suggested by Charles Darwin in 1871,[74] i.e., to consider shallow ponds, lakes, estuaries, or tidal lagoons in prebiotic times that would have had temperatures of about 100.0 °C. The reactions discussed here could have happened in such water bodies, and then evaporation of the water would have led to increased concentration of the products formed. Such a scenario would have led to greater interactions between the molecules formed, and thus, to more complex molecules. It is also possible that such precursors could have seeped out with water from hydrothermal vents and been concentrated at water–ice interfaces in the vicinity of the vents, which would then have allowed subsequent, more complex molecules to have emerged.

Another clarification that should be made is with regard to the specified condition (iii) in the Introduction, regarding the need to have chemical reactions occurring without the need for metal catalysts. This condition increases the probability of the chemical reactions taking place all over the Earth's oceans, and not just in the few, select regions of the Earth where metal based catalysts were available. However, in the regions where metal based catalysts were available in early Earth, their presence would have been beneficial and accelerated the formation of the precursors for RNA and proteins.

The other salient points gleaned from the AINR studies are as follows:

(i) As the pathways found for sugar formation (see Figure 4B) indicate, lower barriers have been obtained for chemical conversions when an pathway alternative to reduction was found (see the respective steps in the reduction pathways, shown in Figure S9, Supporting Information). This corroborates experimental observations that indicate that reduction was generally avoided in prebiotic chemistry.[75]

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532–1540

(ii) The formation of low valent species such as carbenes (**17**, **12**, **14**, **15**, **16**, NH_2COH) is an important reason why most of the barriers for the mechanistic pathways discovered through the AINR approach were seen to be reasonable to low. This interesting fact echoes previous hypotheses that low valent main group compounds are important intermediates in mechanistic cycles.[70,76,77]

(iii) The AINR was seen to exploit H_2O or NH_3 molecules as proton shuttling catalysts in most of the elementary reaction steps. This, again, has relevance in the context of recent reports,[50,78,79] suggesting that a lot of biology occurs with the mediation of H_2O as a proton shuttling catalyst. Moreover, the role of NH_3 as a proton shuttling agent has also been explored in the literature.[80−83]

CONCLUSION

The current work shows that interaction between only two different molecules—HCN and H_2O—would have been sufficient to give rise to most of the important precursors to RNA and proteins in prebiotic times. Taking advantage of the recently developed AINR method,[50] which has allowed us to discover new reaction pathways, we have shown that cyanamide, glycoladehyde, oxazole derivative, and glycine all could have been formed from only a single carbon and nitrogen source molecule: HCN and a single oxygen source molecule: H_2O, at temperatures of about 80.0−100.0 °C. Pathways that were found to be feasible were seen to avoid the reduction step, corroborating previous experimental reports.[75] Most of the steps of the discovered mechanistic routes have barriers that are low to moderate, with only a few higher barriers of ∼40.0 kcal/mol, which suggests that the reactions could have occurred without the mediation of metal catalysts and through the aid of thermochemistry alone. This insight is valuable because it helps to explain how the reactions could have taken place in the absence of photochemical activity on the surface of Earth's oceans. Furthermore, the RNA and protein precursor molecules were obtained during the simulations in "one-pot", i.e., during a single simulation in the AINR.

These findings make it possible to imagine that the molecules necessary for building larger, more complex entities such as RNA and proteins could have existed and interacted together in at least some of the water bodies present in early Earth. The current work thus indicates that HCN and H_2O could have been the Adam and Eve of chemical evolution—the source of the precursor molecules that formed the basis of life on Earth.

COMPUTATIONAL METHODS

Ab Initio Molecular Dynamics (AIMD) Simulations. The nanoreactor AIMD simulations were performed with the TeraChem 1.9 software package[84−90] using the Hartree−Fock (HF)[91] electronic wave function and the 3-21g(d) Gaussian basis set,[92] to calculate the Born−Oppenheimer potential energy surface. This method has been implemented in TeraChem by Martinez and co-workers.[50] This approach was deemed acceptable because the HF method is well-known for predicting chemically reasonable structures.[93] Also, it should be noted HF was not employed to determine barrier heights and reaction rates: its only role was in the discovery process. This was also the approach employed by Martinez and co-workers in their original AINR paper (employing HF/3-21g), where they replicated the results obtained from the Urey−Miller experiment, as well as from the interaction of acetylene molecules.[50] We note here that we did also attempt discovery in the AINR simulations with DFT using the B3LYP density functional and the 3-21g(d) basis set and did find the preliminary intermediates (formamide, formic acid, formaldehyde and others) through this approach as with HF/3-21g(d). However, this was at greater computational expense and did not appear to give different results from the HF/3-21g(d) approach. Hence, we have limited the discovery process to HF/3-21g(d) in the AINR simulations.

The AINR simulation results that have been discussed here pertain to the case in which 16 H_2O and 15 HCN molecules were taken together in a spherical box of radii 10.0 and 3.5 Å (the system alternated between the two radii, in order for the collisions to take place—see original paper by Martinez and co-workers[50]). This system was allowed to evolve for 750 ps and generated the intermediates and reaction pathways that have been discussed. Additionally, we have also performed several AINR simulations where we varied different parameters, such as (i) the ratio of the reactant species, (ii) the total number of molecules taken in the simulation box, (iii) the spherical boundary conditions, (iv) the temperature, and (v) the total time of the AIMD simulations. The results obtained by changing the parameters (i−v) have been discussed in the Computational Details section in the Supporting Information. In general, they indicate that while most of the intermediates were discovered by varying (i) to (v), the most comprehensive results were obtained from the simulation case described in the manuscript: taking 16 H_2O and 15 HCN molecules in a spherical box of radii 10.0 and 3.5 Å. Moreover, for this case, multiple simulations were also performed from the same initial configuration and were seen to give rise to all the desired intermediates and products (following the same mechanistic pathways), although the time of formation of these species during the simulations was seen to change from simulation to simulation.

Newton's equations of motion were calculated using Langevin dynamics with an equilibrium temperature of 2000.0 K (also the starting temperature of the dynamics). We have used this high temperature in order to increase the average kinetic energy of the reactant molecules and for faster dynamics, allowing the overcoming of noncovalent interactions without the breaking of covalent bonds. This, too, follows the example of the work with the AINR done by Martinez and co-workers.[50] The nanoreactor simulations employ a piston to accelerate the reaction rate. We have employed the augmented direct inversion in the iterative subspace (ADIIS) algorithm[94] available in TeraChem as an alternative tool for self-consistent field calculations at each AIMD step in which the default DIIS algorithm[95] failed to converge. Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event. For further information on the spherical boundary conditions, analysis of simulation trajectories by NetworkX, Numpy, and Graphviz Python libraries, please see the Supporting Information.

The mechanistic pathways obtained from the AINR simulations were then analyzed as follows: (i) all the reactant, intermediate, and transition state structures were optimized with high level density functional theory (DFT) calculations, using the Turbomole 7.0 software package at the B3LYP-D3/TZVP+COSMO(ε = 80.0) level of theory, (ii) single point calculations were then done at both the coupled cluster

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

(RICC2) as well as the Møller−Plesset second order perturbation (RIMP2) levels of theory in order to obtain the electronic energies. Hence, the calculations have been done at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP +COSMO(ε = 80.0) and the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses in the free energy profiles) levels of theory. Further, entropic and internal energy contributions were determined by frequency calculations at the B3LYP-D3/TZVP+COSMO(ε = 80.0) level of theory. Volume corrections were also included for the translational entropy term. Moreover, in addition to these calculations done with Turbomole 7.0, Gaussian 09 was also employed to obtain the free energy profiles, at the B3LYP-D2/6-311++g(d,p)+PCM(ε = 80.0) and the M06-2X/6-311++g(d,p)+PCM(ε = 80.0) levels of theory. The free energy profiles obtained from the Gaussian 09 calculations are provided in the Supporting Information and were seen to be similar to the profiles obtained by employing Turbomole 7.0. More details of the static DFT and QM calculations, along with all the relevant references, are provided in the Supporting Information.

Safety Statement. No unexpected or unusually high safety hazards were encountered.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.9b00520.

Computational details, alternative mechanistic pathways, Cartesian coordinates of all the transition states (PDF)
Movie (MP4)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: k.vanka@ncl.res.in.

ORCID

Kumar Vanka: 0000-0001-7301-7573

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the Centre of Excellence in Scientific Computing (COESC), NCL, Pune, for providing computational facilities. K.V. is grateful to the Department of Science and Technology (DST) (EMR/2014/000013) for providing financial assistance. T.D. and S.G. thank Council of Scientific and Industrial Research (CSIR) for providing Research Fellowship.

■ REFERENCES

(1) Gánti, T. *The Principles of Life*; Oxford Univsersity Press: Oxford UK, 2003.
(2) Dyson, F. *Origins of Lfe*; Cambridge University Press: Cambridge UK, 1999.
(3) Sutherland, D. Studies on the origin of life — the end of the beginning. *Nat. Rev.* **2017**, *1*, 1−7.
(4) Schopf, J. W. *Life's Origin: The Beginnings of Biological Evolution*; University of California Press: Oakland, CA, 2002.
(5) Steel, M.; Penny, D. Origins of life: common ancestry put to the test. *Nature* **2010**, *465*, 168−169.
(6) Oparin, A. I. *The Origin of Life*; World Publishing: Cleveland, 2003.
(7) Courier, D.; Oro, J. *Historical Understanding of Life's Beginnings*; World Publishing: Cleveland, 2002.
(8) Bernal, J. D. *The Origin of Life*; World Publishing: Cleveland, 1967.
(9) Sutherland, J. D. The origin of life—out of the blue. *Angew. Chem., Int. Ed.* **2016**, *55*, 104−121.
(10) Bracher, P. J. Origin of life: primordial soup that cooks itself. *Nat. Chem.* **2015**, *7*, 273−274.
(11) Pascal, R.; Pross, A.; Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **2013**, *3*, 130156−130164.
(12) Bada, J. L. How life began on earth: a status report. *Earth Planet. Sci. Lett.* **2004**, *226*, 1−15.
(13) Bada, J. L. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiment. *Chem. Soc. Rev.* **2013**, *42*, 2186−2196.
(14) Woese, C. *The Genetic Code*; Harper & Row, 1967; pp 179−195.
(15) Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367−379.
(16) Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **1968**, *38*, 381−393.
(17) Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **2002**, *418*, 214−221.
(18) Joyce, G. F.; Orgel, L. E. In *The RNA World*; Gesteland, R. F.; Cech, T. R.; Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: New York US, 2006; pp 23−56.
(19) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99−123.
(20) Schmidt, S.; Sunyaev, S.; Bork, P.; Dandekar, T. Metabolites: a helping hand for pathway evolution. *Trends Biochem. Sci.* **2003**, *28*, 336−341.
(21) Caetano-Anolles, G.; Kim, H. S.; Mittenthal, J. A. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9358−9363.
(22) Copley, R. R.; Bork, P. Homology among $(\beta\alpha)_8$ barrels: implications for the evolution of metabolic pathways. *J. Mol. Biol.* **2000**, *303*, 627−640.
(23) Danchin, A. From chemical metabolism to life: the origin of the genetic coding process. *Beilstein J. Org. Chem.* **2017**, *13*, 1119−1135.
(24) Ferris, J. P.; Hill, A. R.; Liu, R., Jr; Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381*, 59−61.
(25) Verlander, M. S.; Lohrmann, R.; Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J. Mol. Evol.* **1973**, *2*, 303−316.
(26) Szostak, J. W. System chemistry on early earth. *Nature* **2009**, *459*, 171−172.
(27) Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, *459*, 239−242.
(28) Ritson, D. J.; Sutherland, J. D. Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew. Chem., Int. Ed.* **2013**, *52*, 5845−5847.
(29) Bowler, F. R.; Chan, C. K.; Duffy, C. D.; Gerland, B.; Islam, S.; Powner, M. W.; Sutherland, J. D.; Xu, J. Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* **2013**, *5*, 383−389.
(30) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **2015**, *7*, 301−307.
(31) Ritson, D. J.; Battilocchio, C.; Ley, S. V.; Sutherland, J. D. Mimicking the surface and prebiotic chemistry of early Earth using flow chemistry. *Nat. Commun.* **2018**, *9*, 1821−1830.
(32) Anastasi, C.; Buchet, F. F.; Crowe, M. A.; Helliwell, M.; Raftery, J.; Sutherland, J. D. The search for a potentially prebiotic synthesis of

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

nucleotides via arabinose-3-phosphate and its cyanamide derivative. *Chem. - Eur. J.* **2008**, *14*, 2375−2388.
(33) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. Cyanoacetylene in prebiotic synthesis. *Science* **1966**, *154*, 784−785.
(34) Ferris, J. P.; Hagan, W. J., Jr. HCN and chemical evolution: the possible role of cyano compounds in prebiotic synthesis. *Tetrahedron* **1984**, *40*, 1093−1120.
(35) Schlesinger, G.; Miller, S. L. Prebiotic synthesis in atmospheres containing CH_4, CO and CO2. II Hydrogen cyanide, formaldehyde and ammonia. *J. Mol. Evol.* **1983**, *19*, 383−390.
(36) Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528−529.
(37) Bada, J. L.; Cleaves, H. J. *Ab initio* simulations and the Miller prebiotic synthesis experiment. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E342.
(38) Arney, N. G; Domagal-Goldman, D. S; Meadows, S. V. Organic haze as a biosignature in anoxic earth-like atmospheres. *Earth and Planetary Astrophysics* **2017**, arXiv:1711.01675.
(39) Kasting, J. F.; Ackerman, T. P. Climatic consequences of very high carbon dioxide levels in the earth's early atmosphere. *Science* **1986**, *234*, 1383−1385.
(40) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **1967**, *30*, 223−253.
(41) Aubrey, A. D.; Cleaves, H. J.; Bada, J. L. The role of submarine hydrothermal systems in the synthesis of amino acids. *Origins Life Evol. Biospheres* **2009**, *39*, 91−108.
(42) White, R. H. Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250°C. *Nature* **1984**, *310*, 430−432.
(43) Miller, S. L.; Bada, J. L. Submarine hot springs and the origin of life. *Nature* **1988**, *334*, 609−611.
(44) Szori, M.; Jojart, B.; Izsak, R.; Szori, K.; Csizmadia, I. G.; Viskolcz, B. Chemical evolution of biochemical building blocks. Can thermodynamics explain the accumulation of glycine in the prebiotic ocean? *Phys. Chem. Chem. Phys.* **2011**, *13*, 7449−7458.
(45) Florian, J.; Warshel, A. Phosphate ester hydrolysis associative in aqueous solution: Associative versus dissociative mechanisms. *J. Phys. Chem. B* **1998**, *102*, 719−734.
(46) Song, J. L.; Wang, T.; Zhang, X.; Chung, W. L.; Wu, D. Y. A Combined DFT/IM-MS Study on the reaction mechanism of cationic Ru(II)-catalyzed hydroboration of alkynes. *ACS Catal.* **2017**, *7*, 1361−1368.
(47) Lee, B. T.; Mckee, L. M. Mechanistic study of $LiNH_2BH_3$ formation from $(LiH)_4$ + $NH_3BH_3$and subsequent dehydrogenation. *Inorg. Chem.* **2009**, *48*, 7564−7575.
(48) Nguyen, T. M.; Nguyen, S. V.; Matus, H. M.; Gopakumar, G.; Dixon, A. D. Molecular mechanism for H_2 release from BH_3NH_3, including the catalytic role of the Lewis acid BH_3. *J. Phys. Chem. A* **2007**, *111*, 679−690.
(49) Yamakawa, M.; Ito, H.; Noyori, R. The metal−ligand bifunctional catalysis: A theoretical study on the ruthenium(II)-catalyzed hydrogen transfer between alcohols and carbonyl compounds. *J. Am. Chem. Soc.* **2000**, *122*, 1466−1478.
(50) Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044−1048.
(51) Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **2013**, *34*, 1385−1392.
(52) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex chemical reaction networks from heuristics-aided quantum chemistry. *J. Chem. Theory Comput.* **2014**, *10*, 897−907.
(53) Miller, S. L.; Urey, H. C. Organic compound synthesis on the primitive earth. *Science* **1959**, *130*, 245−251.
(54) Turner, B. E.; Liszt, H. S.; Kaifu, N.; Kisliakov, A. G. Microwave detection of interstellar cyanamide. *Astrophys. J.* **1975**, *201*, L149−L152.
(55) Lohrmann, R. Formation of urea and guanidine by irradiation of ammonium cyanide. *J. Mol. Evol.* **1972**, *1*, 263−269.
(56) Decker, P.; Schweer, H.; Pohlamnn, R. Bioids: X. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography−mass spectrometry of n-butoxime trifluoroacetates on OV-225. *J. Chromatogr. A* **1982**, *244*, 281−291.
(57) Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. Borate minerals stabilize ribose. *Science* **2004**, *303*, 196.
(58) Serianni, A. S.; Clark, E. L.; Barker, R. Carbon-13-enriched carbohydrates. Preparation of erythrose, threose, glyceraldehyde, and glycolaldehyde with ^{13}C-enrichment in various carbon atoms. *Carbohydr. Res.* **1979**, *72*, 79−91.
(59) Fischer, E. Reduction von säuren der Zuckergruppe. *Ber. Dtsch. Chem. Ges.* **1889**, *22*, 2204−2205.
(60) Ritson, D.; Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **2012**, *4*, 895−899.
(61) Benner, S. A.; Kim, H. J.; Carrigan, M. A. Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc. Chem. Res.* **2012**, *45*, 2025−2034.
(62) Cockerill, A. F.; Deacon, A.; Harrison, R. G.; Osborne, D. J.; Prime, D. M.; Ross, W. J.; Todd, A.; Verge, J. P. An improved synthesis of 2-amino-1,3-oxazoles under basic catalysis. *Synthesis* **1976**, *1976*, 591−593.
(63) Eschenmoser, A.; Loewenthal, E. Chemistry of potentially prebiological natural products. *Chem. Soc. Rev.* **1992**, *21*, 1−16.
(64) Strecker, A. Ueberein enneuenaus aldehyd-ammoniak und blausäureents tehenden Körpe. *Liebigs Ann. Chem.* **1854**, *91*, 349−351.
(65) Bar-Nun, A.; Bar-Nun, N.; Bauer, S. H.; Sagan, C. Shock synthesis of amino acids in simulated primitive environments. *Science* **1970**, *168*, 470−473.
(66) Matthews, C. N.; Moser, R. E. Peptide synthesis from hydrogen cyanide and water. *Nature* **1967**, *215*, 1230−1234.
(67) Goldman, N.; Reed, E. J.; Fried, L. E.; William Kuo, I.-F.; Maiti, A. Synthesis of glycine containing complexes in impact of comets on early earth. *Nat. Chem.* **2010**, *2* (11), 949−954.
(68) Saitta, A. M.; Saija, F. Miller experiments in atomistic computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13768−13773.
(69) Thaddeus, P. The prebiotic molecules observed in the interstellar gas. *Philos. Trans. R. Soc., B* **2006**, *361*, 1681−1687.
(70) Eckhardt, A. K.; Linden, M. M.; Wende, R. C.; Bernhardt, B.; Schreiner, P. R. Gas-phase sugar formation using hydroxymethylene as the reactive formaldehyde isomer. *Nat. Chem.* **2018**, *10*, 1141−1147.
(71) Fabian, B.; Szori, M.; Jedlovszky, P. Floating patches of HCN at the surface of their aqueous solutions − can they make "HCN World" plausible? *J. Phys. Chem. C* **2014**, *118*, 21469−21482.
(72) Szori, M.; Jedlovszky, P. Adsorption of HCN at the Surface of Ice: A grand canonical Monte Carlo simulation study. *J. Phys. Chem. C* **2014**, *118*, 3599−3609.
(73) Menor-Salván, C.; Marín-Yaseli, R. M. Prebiotic chemistry in eutectic solutions at the water−ice matrix. *Chem. Soc. Rev.* **2012**, *41*, 5404−5415.
(74) Ball, P. *Elegant Solutions: Ten Beautiful Experiments in Chemistry*; The Royal Society of Chemistry: Cambridge UK, 2005; pp , 1−205.
(75) Danger, G.; Plasson, R.; Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **2012**, *41*, 5416−5429.
(76) Mandal, S. K.; Roesky, H. W. Group 14 hydrides with low valent elements for activation of small molecules. *Acc. Chem. Res.* **2012**, *45*, 298−307.
(77) Nesterov, V.; Reiter, D.; Bag, P.; Frisch, P.; Holzner, R.; Porzelt, A.; Inoue, S. NHCs in main group chemistry. *Chem. Rev.* **2018**, *118*, 9678−9842.
(78) Mikulski, R.; West, D.; Sippel, K. H.; Avvaru, B. S.; Aggarwal, M.; Tu, C.; McKenna, R.; Silverman, D. N. Water network in fast

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

proton transfer during catalysis by human carbonic anhydrase II. *Biochemistry* **2013**, *52*, 125−131.
(79) De Vivo, M.; Ensing, B.; Klein, M. L. Computational study of phosphatase activity in soluble epoxide hydrolase: High efficiency through a water bridge mediated proton shuttle. *J. Am. Chem. Soc.* **2005**, *127*, 11226−11227.
(80) Pal, A.; Vanka, K. Small molecule activation by constrained phosphorous compounds: Insight from theory. *Inorg. Chem.* **2016**, *55*, 558−565.
(81) Cord-Ruwisch, R.; Law, Y.; Cheng, K. Y. Ammonium as a sustainable proton shuttle in bio electrochemical systems. *Bioresour. Technol.* **2011**, *102*, 9691−9696.
(82) Meuwly, M.; Karplus, M. Simulation of proton transfer along ammonia wires: An "*ab initio*" and semiempirical density functional comparison of potentials and classical molecular dynamics. *J. Chem. Phys.* **2002**, *116*, 2572−2585.
(83) Jaroszewski, L.; Lesyng, B.; Tanner, J. J.; McCammon, J.A. *Ab initio* study of proton transfer in $[H_3N\text{-}H\text{-}NH_3]^+$ and $[H_3N\text{-}H\text{-}H_2O]^+$. *Chem. Phys. Lett.* **1990**, *175*, 282−288.
(84) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.
(85) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge transfer and polarization in solvated proteins from *ab initio* molecular dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.
(86) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Excited-state electronic structure with configuration interaction singles and Tamm−Dancoff time dependent density functional theory on graphical processing units. *J. Chem. Theory Comput.* **2011**, *7*, 1814−1823.
(87) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating efficient quantum chemistry codes for novel architectures. *J. Chem. Theory Comput.* **2013**, *9*, 213−221.
(88) Ufimtsev, I. S.; Martinez, T. J. Graphical processing units for quantum chemistry. *Comput. Sci. Eng.* **2008**, *10*, 26−34.
(89) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.
(90) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.
(91) Froese, F. C. General Hartree-Fock program. *Comput. Phys. Commun.* **1987**, *43*, 355−365.
(92) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-Consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939−947.
(93) Feller, D.; Peterson, K. A. An examination of intrinsic errors in electronic structure methods using the Environmental Molecular Sciences Laboratory computational results database and the Gaussian-2 set. *J. Chem. Phys.* **1998**, *108*, 154−176.
(94) Hu, X.; Yang, W. Accelerating self-consistent field convergence with the augmented Roothaan−Hall energy function. *J. Chem. Phys.* **2010**, *132*, 054109.
(95) Pulay, P. Convergence acceleration of iterative sequences—the case of SCF iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540