TRABALLO FIN DE GRAO

GRAO EN QUÍMICA

UNIVERSIDADE
DE SANTIAGO DE COMPOSTELA

Facultade de Química

Unravelling the formation of guanine in the ISM

Curso 2023/24

Alumna: Daría Pérez López

FACULTADE DE QUÍMICA

Antonio Fernández Ramos, titor e docente do Departamento de Química Física, e Marta Castiñeira Reis, cotitor e docente do Departamento de Química Física, autorizan a presentación do Traballo de Fin de Grao do alumno/a Daría Pérez López na convocatoria de Xullo do curso 2023/2024, o cal foi realizado baixo a súa dirección no CIQUS (Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares).

E para que así conste asinamos o presente informe en Santiago de Compostela o 26 de Xuño de 2024.

Sinatura de Antonio Fernández Ramos

FERNANDEZ RAMOS ANTONIO - 34996049P

Firmado digitalmente por FERNANDEZ RAMOS ANTONIO - 34996049P
Fecha: 2024.06.26 16:08:33 +02'00'

Sinatura de Marta Castiñeira Reis

Marta Castiñeira Reis

Digitally signed by Marta Castiñeira Reis
Date: 2024.06.26 10:28:38 +02'00'

Contents

Abstract

The chemical processes within the Interstellar Medium (ISM) are intricate and involve a diverse array of organic molecules, composed of carbon, nitrogen, and oxygen among other elements. These molecules are believed to play a crucial role in the formation of bio-molecules, such as guanine. Understanding the genesis of such prebiotic molecules in the ISM has been a subject of significant interest for over half a century.

Here, we present an extensive computational investigation to elucidate the possible chemical pathways leading to the formation of guanine in the cold regions of the ISM. With this purpose, the AutoMeKin software was used in combination with MOPAC and Gaussian16, along with the semi-empirical PM7 method and Density Functional Theory (DFT) [MO8HX/6-31+G(d,p)], respectively.

Our findings reveal that guanine can be synthesized via different barrierless free radical pathways. Moreover, we established a correlation between guanine and structural isomers of cytosine, indicating a common potential pathway for the formation of other substantial nucleobases.

This study also highlights the effectiveness of AutoMeKin software in generating tautomers of molecules, showing a strong alignment with both experimental and theoretical data obtained in recent years.

Resumen

Los procesos químicos dentro del Medio Interestelar (ISM, por sus siglas en inglés) son complejos e implican una diversa gama de moléculas orgánicas compuestas por carbono, nitrógeno y oxígeno, entre otros elementos. Se cree que estas moléculas juegan un papel crucial en la formación de biomoléculas, como, por ejemplo, la guanina. Entender el origen de estas moléculas prebióticas en el ISM ha sido un tema de interés significativo durante más de medio siglo.

En este trabajo se presenta una amplia investigación computacional para elucidar los posibles caminos de reacción que conducen a la formación de guanina en las regiones frías del ISM. Con este propósito, se ha utilizado el software AutoMeKin en combinación con MOPAC y Gaussian16, junto con el método semi-empírico PM7 y la Teoría del Funcional de la Densidad (DFT) [MO8HX/6-31+G(d,p)], respectivamente.

Los resultados obtenidos revelan que guanina puede ser obtenida a través de diferentes reacciones químicas que implican radicales libres sin barreras energéticas. Además, se ha establecido una correlación entre la guanina y los isómeros estructurales de la citosina, indicando un posible camino común para la formación de otras bases nitrogenadas importantes.

Este estudio también destaca la efectividad del software AutoMeKin en la generación de tautómeros de moléculas, mostrando una estrecha alineación de nuestros datos con aquellos publicados en la literatura.

Resumo

Os procesos químicos dentro do Medio Interestelar (ISM) son complexos e implican unha variada gama de moléculas orgánicas compostas por carbono, nitróxeno e oxíxeno, entre outros elementos. Crese que estas moléculas xogan un papel crucial na formación de biomoléculas, como a guanina, por exemplo. Comprender a orixe destas moléculas prebióticas no ISM leva sendo un tema de interese significativo durante máis de medio século.

Neste traballo preséntase unha ampla investigación computacional para esclarecer os posibles camiños de reacción que conducen á formación de guanina nas rexións frías do ISM. Con este propósito, utilizouse o software AutoMeKin en combinación con MOPAC e Gaussian16, xunto co método semiempírico PM7 e a Teoría do Funcional da Densidade (DFT) [MO8HX/6-31+G(d,p)], respectivamente.

Os nosos achados revelan que guanina pode ser obtida a través de diferentes reaccións químicas que implican radicais libres sen barreiras enerxéticas. Ademais, estableceuse unha correlación entre guanina e os isómeros estruturais de citosina, indicando un posible camiño común para a formación doutras bases nitroxenadas importantes.

Este estudo tamén destaca a efectividade do software AutoMeKin na xeración de tautómeros de moléculas, mostrando unha forte alineación entre os datos aquí presentados e os precedentes bibliográficos.

1. Introduction

One of the fundamental questions that has captivated the interest of the scientific community for centuries is the origin of life. This inquiry has been approached in two distinct phases mainly, even since then studies on this arena remained active. The first phase is the period from 1859 to 1951, which focused predominantly on theoretical hypotheses, and the second phase is the period from the 1940s to 1965, which concentrated on empirical experiments simulating primitive atmospheric conditions. A timeline is shown in Figure 1.

In 1859, Charles Darwin published "On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life", commonly referred to as "On the Origin of Species". Although Darwin only briefly addressed the origin of life, his thoughts on the evolution of species profoundly influenced the scientific community's perspective on it, in the sense that paralelisms of the evolution of species and life were stablished. Prior to 1860, the theory of spontaneous generation (or abiogenesis) was widely accepted. This theory is understood as the birth of an unusual form of matter from inorganic substances in early geological periods.[1] V. Vernadski hypothesized that the diverse variety of life's forms came from a result of geological flow of substance, leading to the appearance of several species belonging to different categories of organisms and biocenoses.[2]

In 1871, Joseph Dalton Hooker expanded on this concept by proposing an evolutionary abiogenesis theory. Hooker's theory suggested that life originated spontaneously, in line with the evolutionary principles outlined by Darwin.[1]

In 1924, a change of paradigm took place when A. I. Oparin proposed that the primitive atmosphere was reducing and lacked carbon dioxide. Under these specific conditions, microscopic globular entities known as coacervates, which resemble cells, could form. In 1929, John Burdon Sanderson Haldane developed similar ideas, culminating in the Oparin-Haldane hypothesis. This hypothesis established key points, including the concept of a primordial soup or early Earth's oceans as a mixture of organic molecules, the subsequent formation of complex molecules, the eventual development of cell membranes, and the absence of oxygen in the early atmosphere.[1]

During the 1930s and 1940s, Alexandre Dauvillier proposed a photochemical origin of life, while in 1947, John Desmond Bernal suggested that clays could catalyze the genesis of life due to their catalytic properties and crystalline structure, which could support the chemical reactions necessary for life. This period was marked by numerous hypotheses, though experimental evidence remained scarce.[1]

The 1950s marked the beginning of prebiotic chemistry as a scientific field. The Melvin Calvin exper-

iment demonstrated that bombarding carbon dioxide with gamma rays under primitive atmospheric conditions produced formaldehyde, representing the first prebiotic chemistry experiment. In 1952, Harold Urey proposed that carbon dioxide was absent from the primitive atmosphere, advocating for experiments with methane. In 1953, Stanley Miller, under Urey's supervision, conducted experiments that produced amino acids from a mixture of hydrogen, ammonia, methane, and water exposed to electrical discharges over a week. Between the 1950s and 1960s, further experiments were conducted, focusing on the synthesis of amino acids, sugars, and nucleic acid bases, particularly uracil, which are fundamental to DNA and RNA that attempt to cast light on the emergence of prebiotic molecules and the origin of life.[1]

In 1958, Sidney Fox and Kaoru Harada achieved the thermopolymerization of amino acids under prebiotic conditions, synthesizing what they termed "proteinoids". These proteinoids demonstrated the potential to form complex macromolecules under conditions thought to resemble those of early Earth. By 1965, Fox had further advanced this work by creating microspheres based on these proteinoids, which provided a model for the protocell structures that may have been precursors to living cells. [1]

These pioneering efforts by Fox and others delineated the three main stages of the model of the origin of life: first, the synthesis of small organic molecules such as sugars, amino acids, and nucleic acid bases takes place; second, the polymerization of these small molecules into macromolecules should occur; and finally, the assembly of these macromolecules into cell-like structures sets the path to the origin of life. These stages collectively illustrate the possible pathways from simple organic compounds to the complex organization of living cells, providing a framework for understanding the chemical evolution that led to the emergence of life on Earth.[1]

A question that remains unsolved is: *how could the transition from abiotic chemistry to processes such as metabolism, replication, and transfer of genetic information have occurred?* The theory of panspermia is the answer. This theory is based on that the delivery of "seeds of life" takes place from the ISM to Earth. The idea was proposed by S. Arrhenius in the late 19^{th} century. He suggested that compounds roamed for millions of years through space in meteorites, comets or any other celestial bodies and set on Earth, where they found the propitious conditions to develop life, or stimulate its formation through abiogenesis. In the initial times, primitive organisms could have used monomers of basic organic compounds of non-biological source, in a very similar way such as chemosynthesis takes place in Earth.[3] This theory is the basis of the work developed here, that pivots on the premise that it is possible that the formation of the basic molecules of life came to Earth from the Interstellar Medium (ISM).

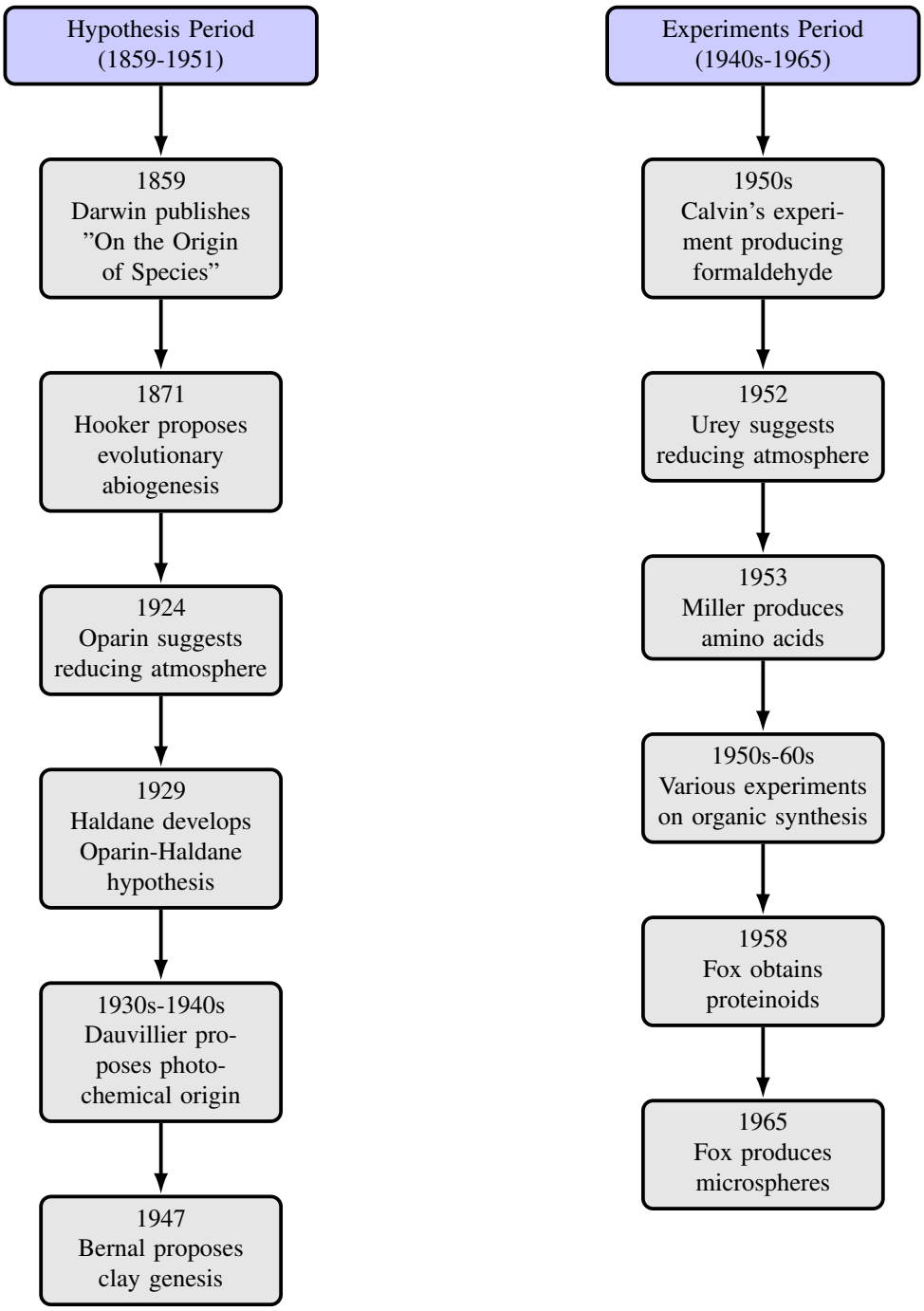


Figure 1: Timeline of key developments in theories and experiments on the origin of life.

The ISM in our galaxy consists of four readily detectable components, three of which are described in terms of clouds of atoms and molecules. The first component is the most luminous one: the HII regions. These regions are areas of ionized hydrogen that emit radiation when protons (H^+) and electrons recombine. These regions are continuously reionized by radiation from nearby stars and are prominently visible near stars with temperatures exceeding 30000 K, such as O-type stars.[1]

The second component comprises dark clouds, which are extremely cold and do not emit visible radiation. However, they do emit far-infrared (at 100 μm) and millimeter molecular line radiation. The absorption of background starlight by dust within these clouds creates the dark bands observed along the Milky Way in regions such as Scorpio and the Cygnus Rift, as well as the famous dark spot in the southern sky known as the Coal Sack. Lastly, the third component, detectable only with sophisticated instrumentation, consists of diffuse clouds. These clouds are lower in mass compared to HII regions or dark clouds and are warmer than dark clouds, yet not warm enough to emit significant radiation. They contain insufficient dust to cause noticeable extinction to the human eye.[4]

Molecules in space are identified using specialized telescopes. The Atacama Large Millimeter Array (ALMA) [5] is a network of highly precise and sensitive ground-based radio telescopes that perform space measurements and obtain high-resolution spectra. ALMA has discovered numerous complex molecules that are significant in prebiotic chemistry, including formaldehyde,[6] glycolaldehyde,[7] ethylene glycol,[8] methyl formate (related to formic acid),[9] and dimethyl ether.[10] Additionally, ALMA has identified molecules containing deuterium, peptide bonds (such as formamide),[11] and methylamine,[12] the precursor of glycine, the simplest amino acid. These emerging molecules are cataloged in a comprehensive database,[13] which currently lists over three hundred identified molecules.

In this context, a pertinent question arises: could pyrimidine and purine nucleobases (guanine, cytosine, uracil, thymine, and adenine) have been formed in the ISM, particularly in its cold regions? This inquiry is driven by the discovery of several nucleobase precursors in various cosmic environments, including circumstellar envelopes, interstellar molecular clouds, interstellar ice, comets, and meteorites. Examples of such precursors include acetonitrile (NH_2CH_2CN) [14] and the proto-sugar glycolaldehyde ($OHCH_2CHO$),[15, 16] both of which are precursors to glycine, as well as nucleobase precursors like HCN,[17] HNCO,[18] NCCCH,[19] $HCONH_2$,[20] and NH_2CONH_2.[21]

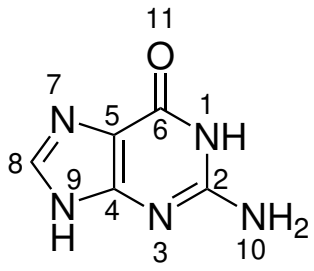
In 2013, Saladino *et al.* demonstrated that a significant variety of molecules, including amino acids and purine and pyrimidine nucleobases, can be synthesized from formamide in meteorite-catalyzed chemical

[1] The Great Nebula in Orion is the most well-known HII region.

reactions in the condensed phase.[22] Furthermore, research by Orlando's group established that purine nucleobases can form through the irradiation of formamide solutions with ultraviolet light at 254 nm.[23] Over the past decade, numerous theoretical studies have been published on the formation of cytosine, thymine, uracil, and adenine under space conditions. In contrast, guanine has received relatively less attention.

Guanine or 2-amino-6-oxypurine (see Figure 2) is a purine molecule characterized by a planar two-ring system with a pentacyclic structure resembling an imidazole ring and a hexacyclic structure resembling pyrimidine. This molecule is found in both DNA and RNA. Jeilani *et al.* proposed that guanine - along with purine, hypoxanthine, and adenine - forms through reactions involving formamide and CN radicals, with guanine following a 21-step mechanism that includes a 30 kcal/mol energy barrier.[24] Da Silva and Araujo explored the theoretical mechanism for guanine formation,[25] determining that the six-membered ring forms initially and only later the formation of the five membered ring takes place, from precursors such as HCN, CCO, HNCNH, and H_2NCN.

Additionally, it is believed that guanine can be formed through the oligomerization of hydrogen cyanide in the ISM. This purine has been synthesized, in the laboratory, from 4-aminoimidazole-5-carboxamide, which is produced by the hydrolysis of 4-aminoimidazole-5-carbonitrile through the addition of cyanogen (NCCN) or potassium cyanate to an aqueous solution.[26] Additionally, guanine has been formed from iced ammonium cyanide solutions maintained at -20 ºC [27] or -78 ºC [28] for 25 and 27 years, respectively, prepared by mixing HCN and ammonia (NH_3). Further synthesis methods include using a plasma at high temperatures with a mixture of CO, N_2, and H_2O in the gas phase,[29] and employing catalysts such as Al_2O_3, SiO_2, or Ni-Fe alloy in a mixture of CO, H_2, and NH_3 at high temperatures and in the gas phase.[30, 31] Overall, these synthetic methods point to a variety of potential reactant sets for guanine.



$C_5H_5N_5O$

Figure 2: Cartoon of the guanine molecule.

2. Objective

The primary goal of this work is to elucidate the potential formation mechanisms of guanine in the ISM, with a particular focus on its cold regions (cold molecular clouds). With this goal, automated mechanistic search methods, that allow us to explore the potential energy surface (PES) for the decomposition of guanine, are employed. Note that the decomposition PES of guanine is the same as its formation PES, the main difference is the reading direction or perspective (see Figure 3).

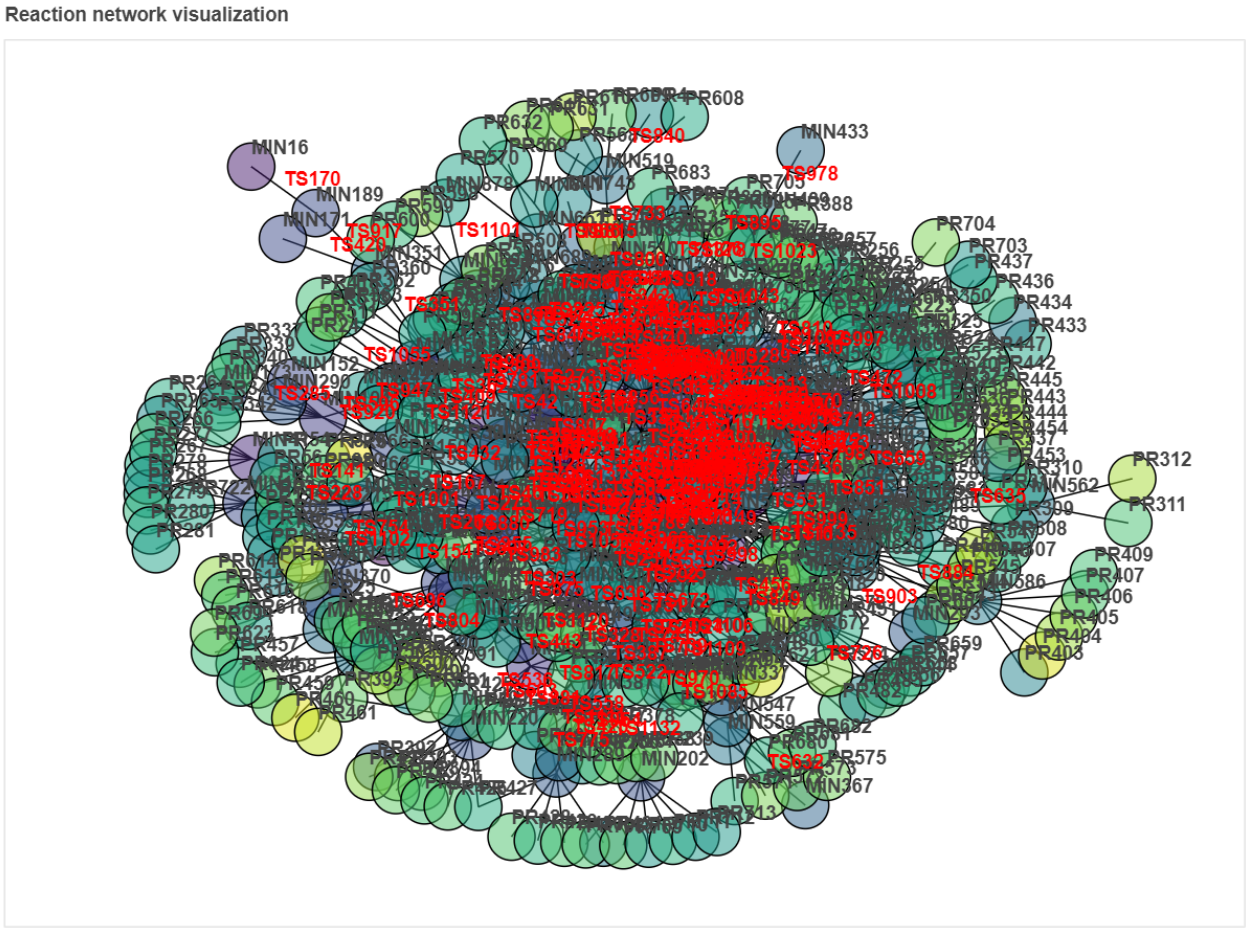


Figure 3: Picture of the reaction network obtained for the decomposition of guanine. Circles represent products (**PR**) or minima (**MIN**) and the connections between the minima are the transition states (**TS**).

3. Computational methods, mechanistic search protocol and other conventions to consider

To search for the stationary points (minima and transition states) that knit the PES of guanine decomposition, we used AutoMeKin. [32] This program runs high-energy trajectories and, from them, snapshots are taken in order to obtain guess transition state structures that are further optimized at a low electronic structure level. For the low-level calculations, we have employed the PM7 semiempirical method [33] implemented in MOPAC. [34] For each transition state successfully obtained, we follow the minimum energy path (MEP), so the minima (reactants and products) associated with that particular transition state are unequivocally identified.

Once the low-level network is created, the low-level transition state structures are re-optimized at a high level, and new MEP calculations are performed to generate a more accurate reaction network. In this case, all the stationary points were re-optimized at the M08HX functional [35] with the 6-31+G(d,p) basis set. All the low-level and high-level optimization calculations were performed with MOPAC and $Gaussian16$ [36] software, respectively. High-level Hessian calculations were performed to check the nature of the stationary points.

For the location of barrierless channels, AutoMeKin employs the Nudged Elastic Band algorithm (NEB). [37] This algorithm is used at the low level PES. The processes found barrierless at this level are presumed to remain barrierless independently of the computational method explored.

The location of reaction paths and sub-networks inside the reaction network generated by AutoMeKin was performed with `amk_tools`. [38, 39] The energy profiles presented in this Bachelor Thesis are adapted from these plots.

The reaction networks for the different potential energy surfaces were built incorporating some restrictions. During the barrierless fragmentation of the molecule, AutoMeKin only examines the cleavage of a single bond with a bond order smaller than 1.5 (MOPAC calculates the bond order). Barrierless fragmentations that involve the cleavage of multiple bonds, are not considered.

Importantly, we have included a spin restriction during the dissociation into fragments. In particular, fragments with an even number of electrons have singlet spin states, so intersystem crossing to different electronic states (for instance, triplet states) is not explored.

Moreover, we only describe and include in our analysis, with detailed pathways, the energy profiles

compatible with the cold regions of the ISM,[13] while the incompatible ones are excluded (see Figure 4). Additionally, we classify the reaction pathways based on the nature of the reactants: pathways involving reactants that have been detected in the ISM are considered more likely and are, therefore, prioritized in our study.

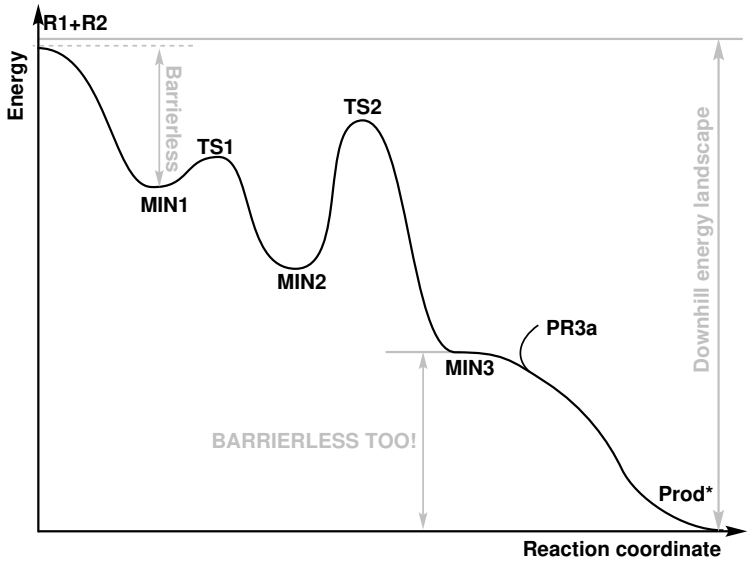


Figure 4: Simplified scheme of a reaction profile compatible with the conditions present in the cold regions of the ISM.

Once the products are classified as detected or not in the ISM and as compatible or not with the conditions present in the cold regions of the ISM, the same procedure is applied iteratively with the products found, until molecules with less than four atoms are reached.[2] A flowchart is shown in Figure 5.

[2]The criterion of the dimension of the molecules is based on the number of atoms present in molecules in ISM because they are usually small.

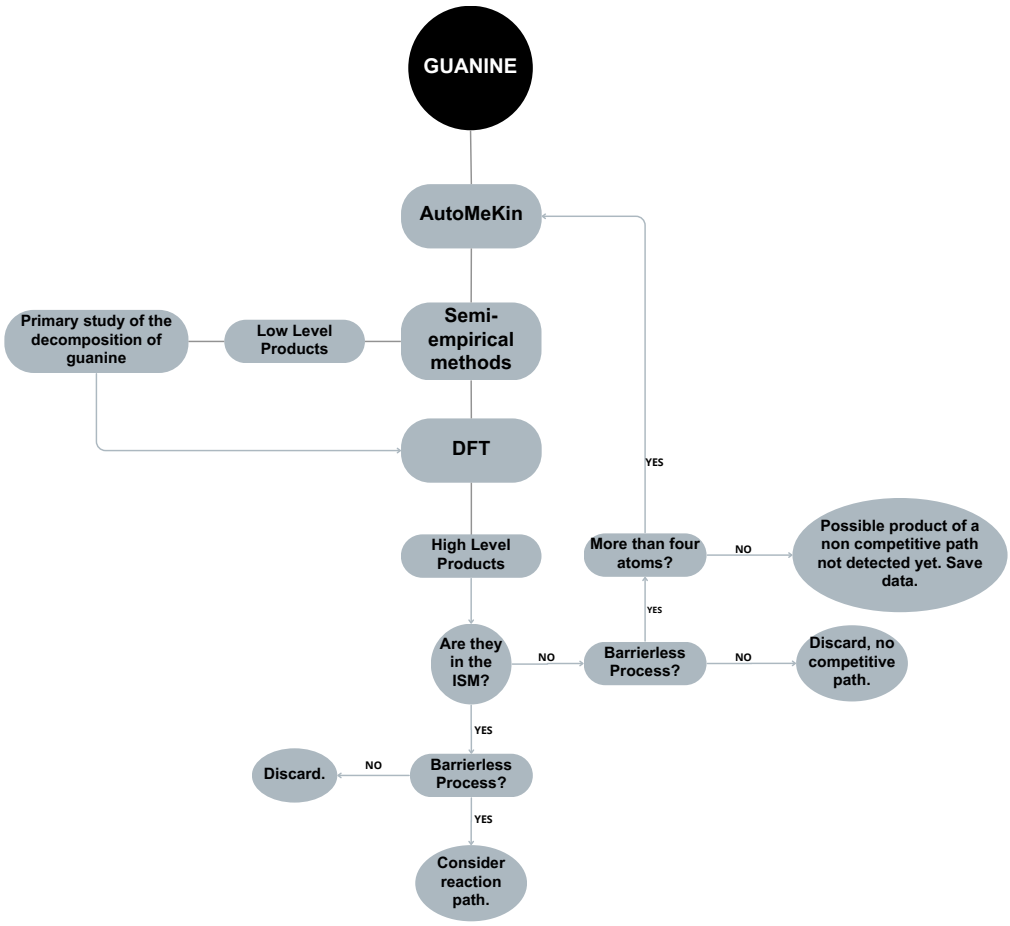


Figure 5: Working protocol flowchart.

To refer to guanine tautomers we have used the nomenclature of Jones *et al.*, [40] it mainly consist on refering to guanine isomers using as marker the position where a hydrogen atom (H) is located. Hence, G-7H referes to an isomer that has a hydrogen at position 7 while G-9H represents guanine with a hydrogen atom bonded to nitrogen at position 9. Refer to Figure 2 for illustration on atom labeling. This nomenclature is clearly not unequivocal nor it identifies completely the isomers. That is why this 3-letter-code is always accompanied by detailed specifications of the molecules, such as IUPAC names or numbering of the molecule.

In this manuscript, the stationary points are labelled as **MIN**, **TS** and **PROD** followed by a number. This is the nomenclature that the AutoMeKin software uses by default. Minima and transition states are numbered by energetic order while products are numbered according to the moment they were found.

4. Results and discussion

The decomposition of guanine rendered a total of 903 minima, 1171 transition states and 731 reaction products. This number of stationary points result in an intractable number of reaction paths describing the decomposition of guanine, as one would naturally expect. Initially, we focused on the analysis of the minima found, since they constitute structural isomers of guanine. Then, we explored the products formed as well as the paths that connect this products with guanine. In the following lines we describe our findings on these two matters.

4.1. Analysis of the most stable minima obtained

Guanine (see Figure 2) has many possible isomers, since it can experience keto-enol tautomerism (marked in blue) and other hydrogen transfers that inter-convert the imino (-NH) and amino (-NH_2) groups (red, see Figure 6). Some of the guanine tautomers (keto pair guanine-7H, guanine-9H and enol guanine-7H, guanine-9H) have been identified experimentally with IR-UV spectroscopy,[41] UV spectroscopy,[42] and in gas phase by photoelectron spectroscopy.[43, 44] Previous computational studies show that there are up to 15 isomers of guanine with close relative energies.[45, 46] These findings account for the difficulties assigning the spectral lines of these species and the existence of different theoretical studies that attempt to localize more guanine tautomers.[40, 47, 48]

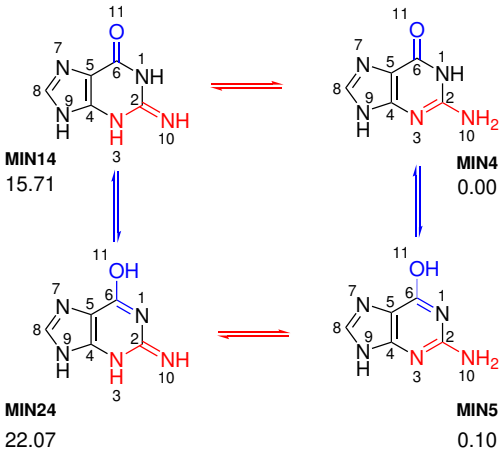


Figure 6: Imino-amino (red) and keto-enol (blue) tautomerism for guanine minima.

Previous computational studies show that there are four low-energy tautomers (amino-oxo and amino-hydroxy). From these tautomers, the amino-oxo G-7H (**MIN3**, see Figure 7) was assigned to the most stable

gas-phase one.[48] We have found that **MIN3** has a relative energy[3] of -0.09 kcal/mol and that there are two other isomers of guanine that are more stable than **MIN3**, namely **MIN1** (-0.74 kcal/mol) and **MIN2** (-0.37 kcal/mol). We assign this difference in energy to the fact that **MIN1** features a fully-conjugated-6-memebered ring, and a cyano moiety. On the other hand, **MIN2** has two fused cycles (one of five and the other of six members) and lacks the cyano motive. Alternatively, **MIN3** has a scaffold with the five and six-membered rings but the oxo moiety is now on its keto form. Additionally, the five-membered ring shows a 1,3-hydrogen shift w.r.t. **MIN2**.

D.B. Jones *et al.*[40] observed, using DFT-B3LYP/TZVP, that the two most stable guanine tautomers are G-7H and G-9H (**MIN3** and **MIN4**, respectively) but they found that their relative stabilities were opposed to the ones we obtain. They also found that the non-planar form of both keto isomers have lower energy than the planar ones and that G-7H has a slightly lower total energy than G-9H. Referring to Figure 7, it can be observed that **MIN4** has a energy of 0.00 kcal/mol, this correlates with Jones *et al.*[40] being the energy for G-7H of -0.09, a bit lower than G-9H.

Alternatively, literature precedents suggest that imino tautomers are less stable than the amino counterparts.[47] Our results are in-line with these suggestions (see Figure 6 and Figure 7). We also found that the imino tautomers exhibit lower stability w.r.t. the corresponding amino tautomers. Specifically, the imino tautomer **MIN14**, derived from **MIN4**, possesses an energy of 15.71 kcal/mol, whereas its amino counterpart **MIN4** exhibits a stability of 0.00 kcal/mol. Similarly, **MIN24**, the imino tautomer of **MIN5**, has an energy of 22.07 kcal/mol, in contrast with the stability of 0.1 kcal/mol observed for **MIN5**. This observation aligns with previous findings by Sabio *et al.*, indicating that imino tautomers tend to be less stable than their amino counterparts.

On the other hand, we found that keto forms exhibit greater stability than their enol counterparts. For instance, **MIN3**, in its keto form, has an energy of -0.09 kcal/mol, while its enol analogue **MIN10** possesses an energy of 10.8 kcal/mol. In contrast to this trend, **MIN4**, in its keto form exhibits an energy of 0.0 kcal/mol, which differs slightly from the energy of **MIN5**, of 0.1 kcal/mol.

M. Sabio *et al.*,[47] by using semiempirical AM1 and Hartree-Fock (STO-3G and 3-21G basis sets), proposed that guanine moiety exists mainly when it is substituted at position 9, although sometimes it exists as enol tautomer. We have found that the G-9H substituted forms are more stable than the other ones, for example **MIN4** with an energy of 0.00 kcal/mol or **MIN5** with an energy of 0.1 kcal/mol. Nonetheless, species that do not follow G-9H pattern have been also found. Its energy exceeds 10 kcal/mol (e.g. **MIN18**,

[3] The energy reference is guanine.

MIN19, **MIN25**, **MIN28**, **MIN30**) (See Figure 7).

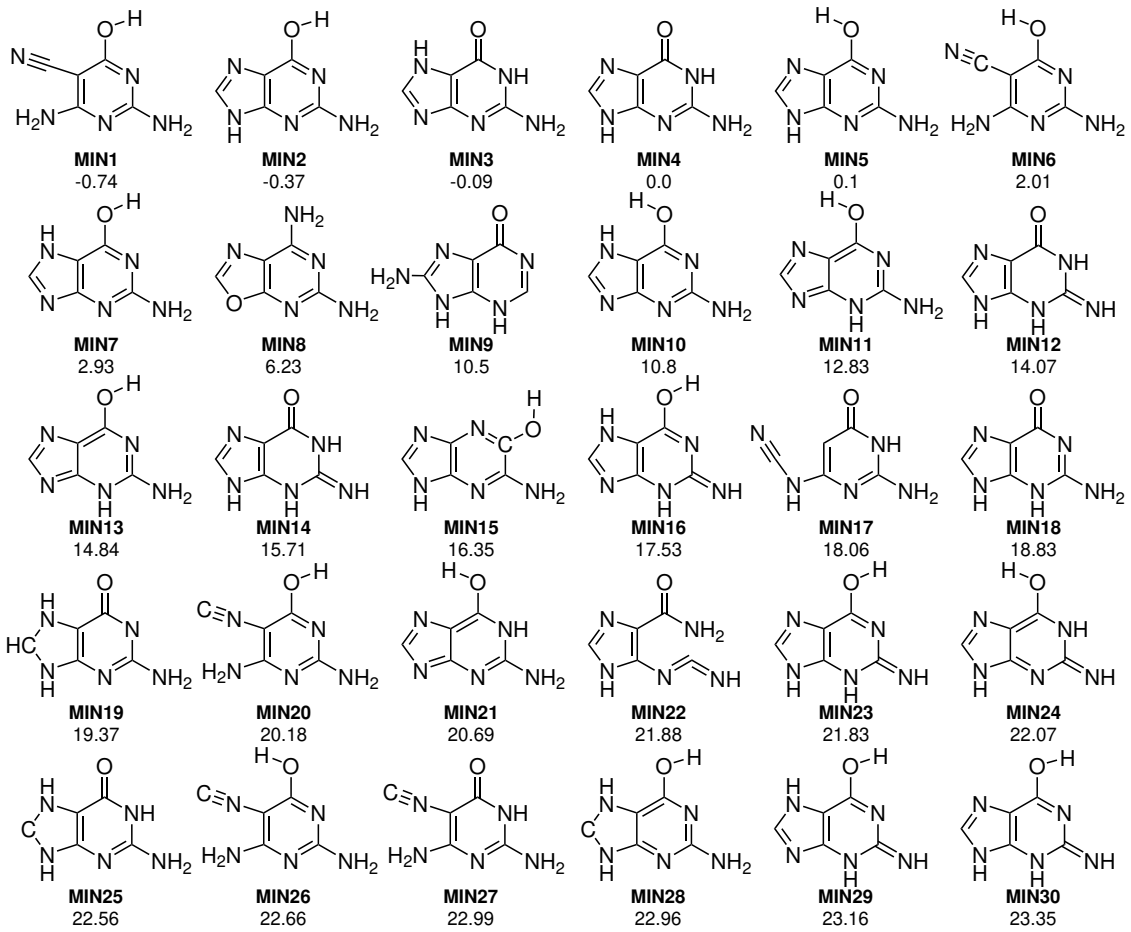


Figure 7: The thirty most stable isomers found for guanine.

4.2. Analysis of the products and reaction paths obtained in the decomposition of guanine

As previously indicated, 731 products were obtained in the decomposition of guanine. Of these 731 products, 352 can be accessed via barrierless mechanisms, compatible with the reaction criteria described in the ***Computational methods and mechanistic search protocol*** section. A selection of these products is presented in Figure 9.

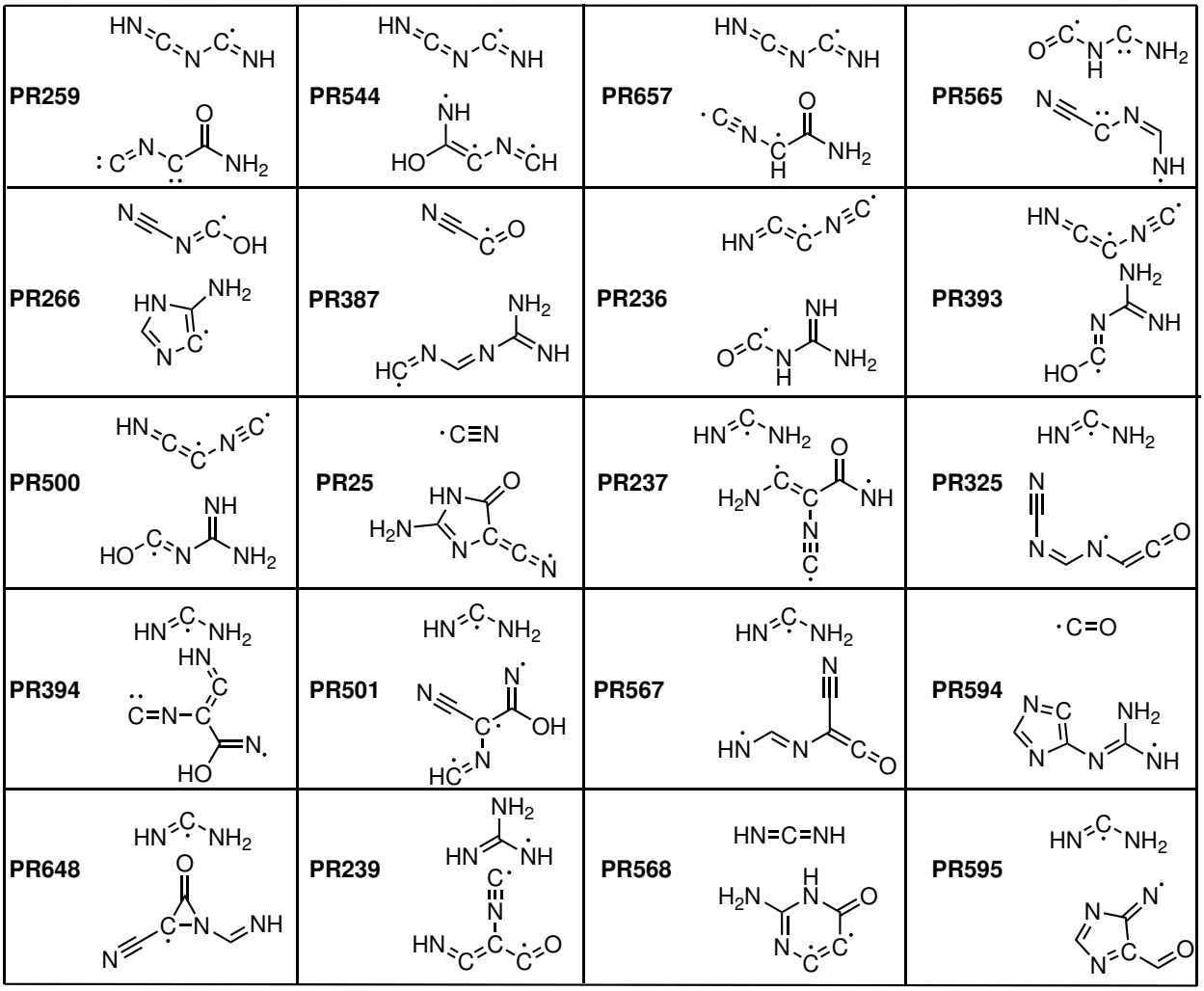


Figure 8: Display of 20 distinct products obtained from the decomposition of guanine.

We were pleased to find that all products obtained in the decomposition of guanine were radicals. Notice that radical species have been suggested to participate in prebiotic chemistry, although they are not the only chemical reactions possible (see Table 4) . For example, Hugen *et al.*[49] studied several radical reactions for the formation of pyrimidine bases, cytosine, thymine and uracyl under cold conditions by using DFT. All this reactions have some common points that are the presence of H-migrations and intramolecular radical cyclizations.

In the upcoming lines, we will discuss the analysis of the paths that can account for the reaction of these radicals towards the formation of guanine.

4.2.1. Radical reorganizations towards guanine

The number of radical-radical reactions found were intractable (over thousands). Hence, we decided to classify the reaction paths according to the number of elementary-steps involved, *i.e.* whether the reaction involved one, two or more elementary steps in the evolution of reactants to products. We will limit our discussion to those transformations with up to four elementary steps (see Figure 9), for the sake of space.

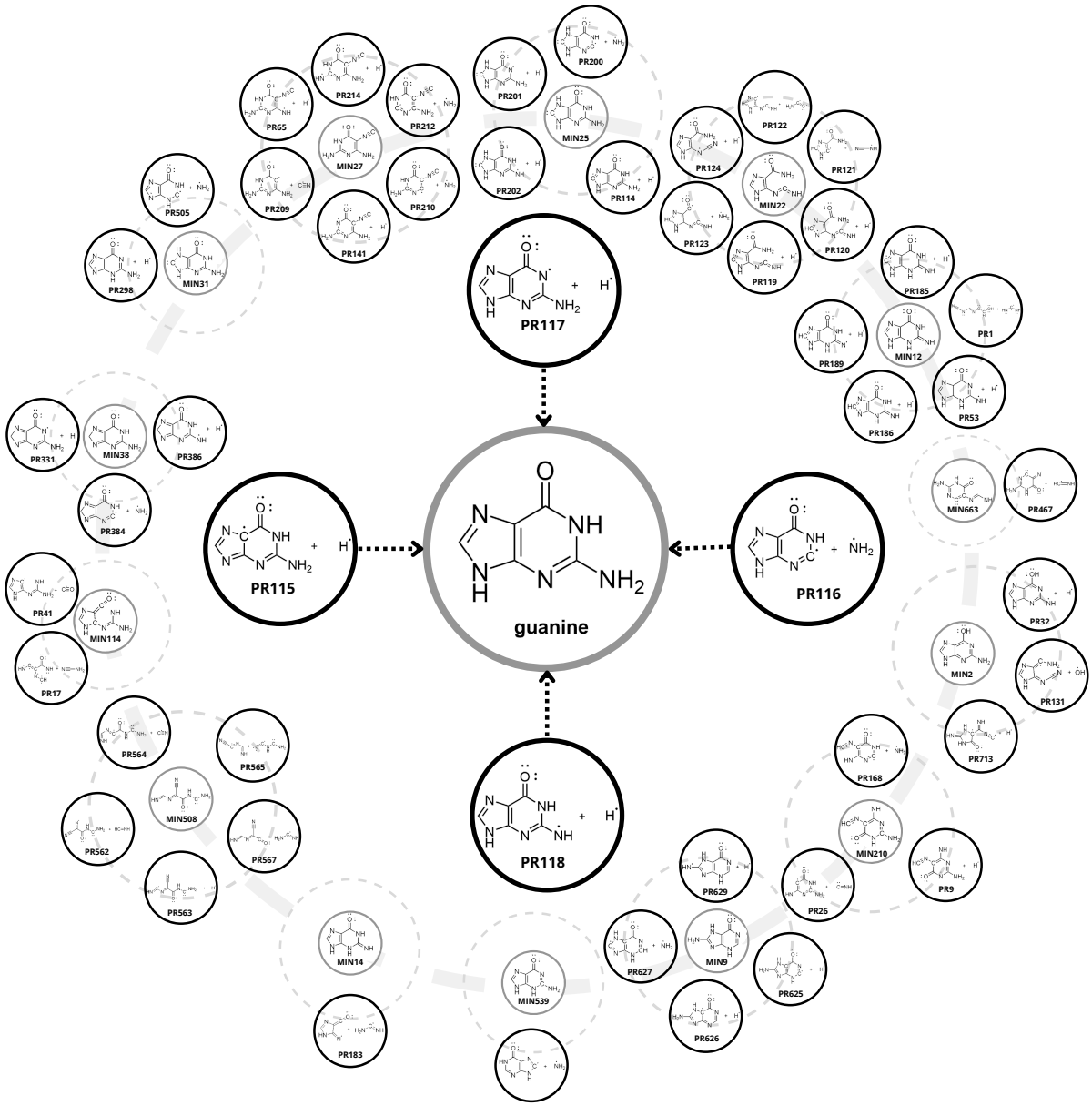


Figure 9: Scheme with 1-step and 2-step reactions for the formation of guanine.

4.2.1.1 Discussion of some selected reaction paths: Reaction of carbon monoxide and (E)-N'-(1H-imidazol-5-yl)-λ^2-azanecarboximidamide radical (A)

Carbon monoxide, a molecule detected in the ISM,[50] can react with **A** through a barrierless radical-radical collapse of the carbon monoxide with C5 at **A** (see Figure 10). This initial radical collapse results in the formation of **MIN114**. This intermediate can further react through an intramolecular cyclization resulting in the formation of a new C-N bond. This step takes places through **TS126** (52.87 kcal/mol) and results with the formation of guanine (**MIN4**).

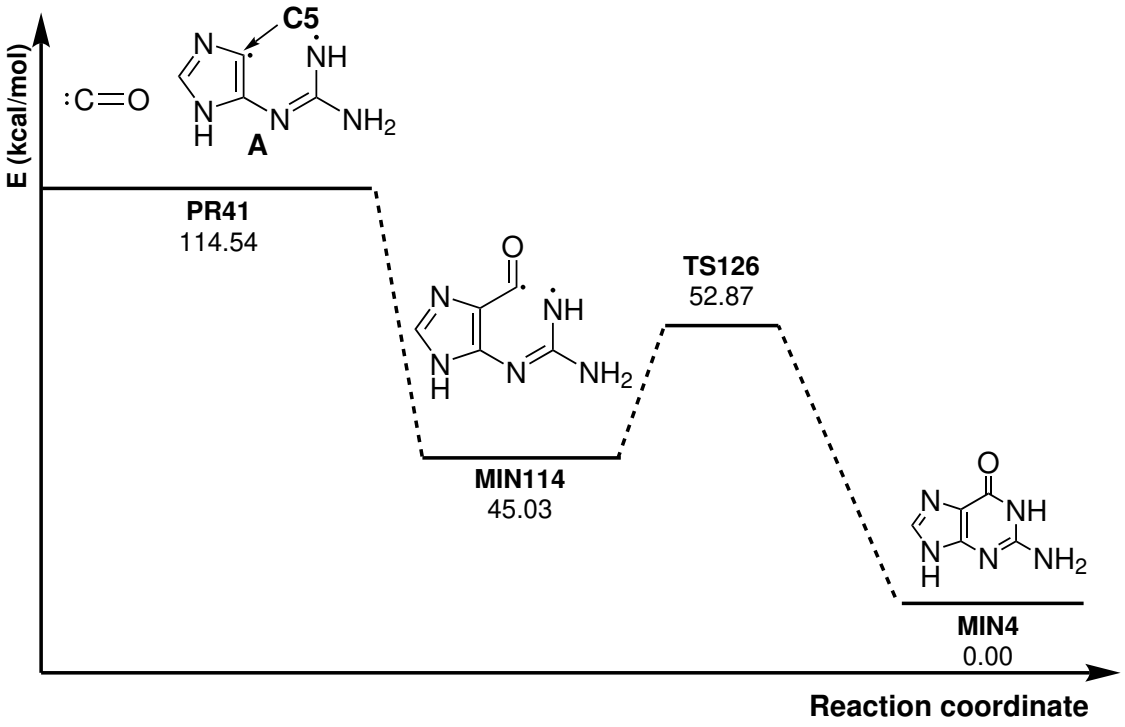


Figure 10: Reaction mechanism and energy profile for the two-step-formation of guanine from(E)-N'-(1H-imidazol-5-yl)-λ^2-azanecarboximidamide and carbon monoxide radicals.

It should be highlighted that **PR41** can further decompose into a guanidine derivative and a dehydrogenated imidazol derivative. Note that guanidine is considered a key precursor of species such as cytosine and uracil. Instances of reports that demonstrate the relevance of guanidine in the formation of guanine are the ones from Miller *et al.*[51] who synthesized pyrimidines by reacting 10^{-3} M cyanoacetaldehyde and 1 M guanidine in aqueous solution at 0°C and pH 8.1. After two months, this reaction yielded cytosine, uracil, isocytosine, and 2,4-diaminopyrimidine (DAP), with the highest yield for uracil. Moreover, computational study of Joong Chul Choe [52] suggested a possible mechanism for the reaction of cyanoacetaldehyde and

guanidine towards guanine, both with and without water-catalysis.

Additionally, we obtained that once guanidine is released from **A**, an imidazole-like molecule is obtained. This result is of great relevance since this core builds up many biological and bioactive heterocyclic compounds. Examples include adenine, another nitrogenous base, histamine, molecule involved in the immune system response, and histidine, an amino acid.

4.2.1.2 Discussion of some selected reaction paths: reaction of 1H-5λ^3-imidazole-4-carboxamide radical (B) and cyanoamidogen

We have found that guanine can be formed alternatively, from λ^2-azanecarbonitrile, also known as cyanoamidogen in a two-step reaction following mechanism showed in Figure 11. The mechanism starts with the reaction of **PR121** (129.00 kcal/mol) to form **MIN22** (21.81 kcal/mol). Once this intermediate is reached, it can further progress to guanine via **TS490** (92.33 kcal/mol) through the formation oc a C-N bond that ensures the formation of the six-membered cycle.

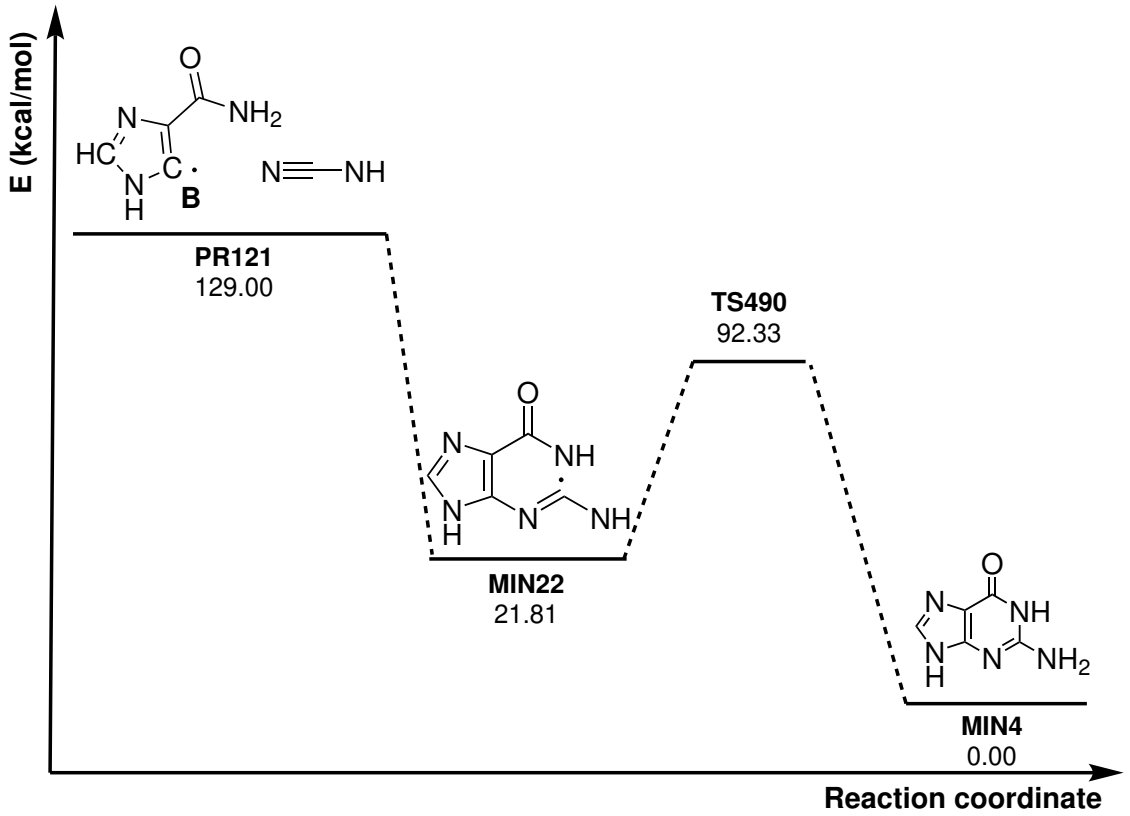


Figure 11: Energy profile of the formation of guanine from 1H-5λ^3-imidazole-4-carboxamide radical and cyanoamidogen.

At this point, it should be highlighted that the cyanoamidogen radical has been detected in the ISM,[53] pointing to its possible role in interstellar chemistry, specifically in the production of prebiotic compounds. HNCN was detected toward the G+0.693-0.027 molecular cloud, situated in the Galactic Center's Sgr B2 complex.[54]

4.2.1.3 Discussion of some selected reaction paths: 2,6-diaminopyrimidin-4(3H)-one (C) and cyanide radicals

The formation of guanine from **PR121** (152.96 kcal/mol) involves the radical collapse of the two components of **PR121**, that yields **MIN27** (22.93 kcal/mol, see Figure 12). The intermediate **MIN27** then transforms into guanine (**MIN4**) via **TS322** (78.24 kcal/mol) through the formation of a new C-N bond. The energy penalization of this step is 53.31 kcal/mol.

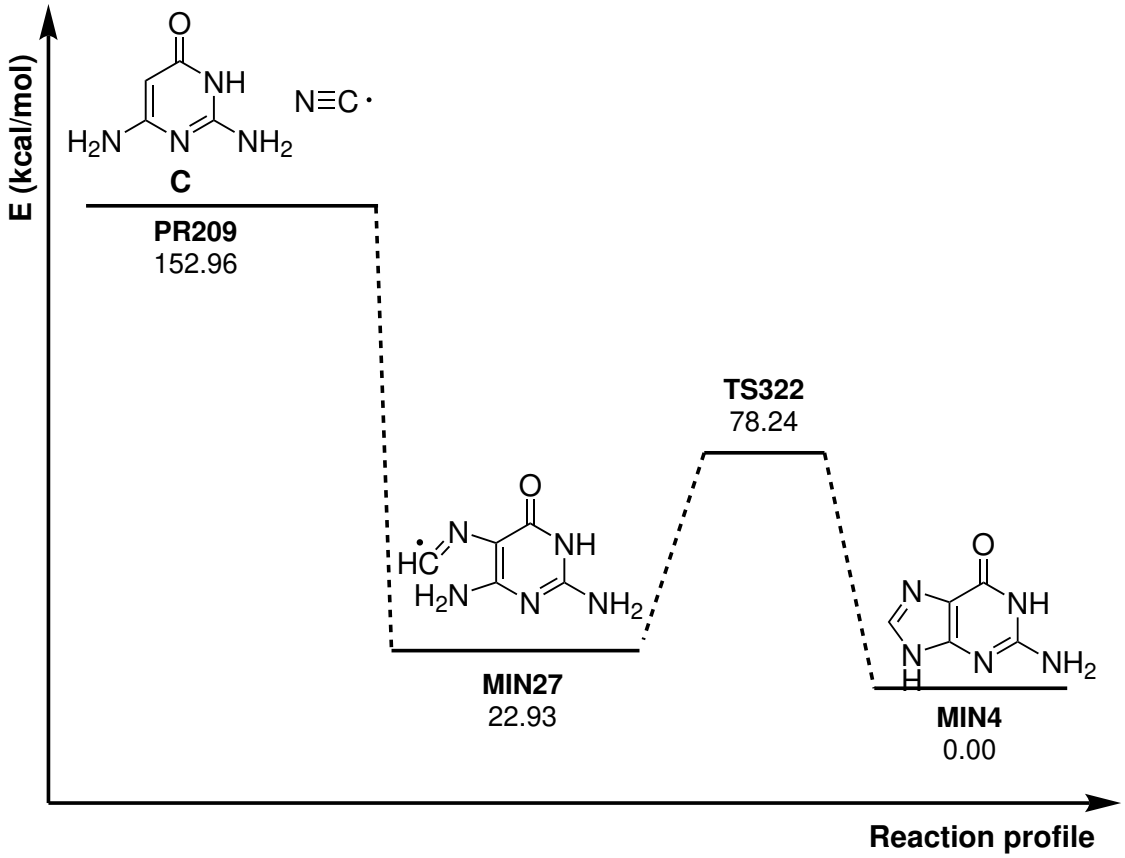


Figure 12: Energy profile of the two-step-chemical-reaction for the formation of guanine from 2,6-diaminopyrimidin-4(3H)-one and cyanide radical.

It is noteworthy that if the 2,6-diaminopyrimidin-4(3H)-one radical (**C**) loses the NH_2 radical, it would be

an isocytosine radical. Isocytosine is an isomer of cytosine, also known as 4-amino-3H-pyrimidin-2-one. Its chemical structure is obtained by interchanging the amino and oxo groups in the pyrimidine backbone.[55] This structural similarities, potentially point toward a reaction path for the formation of both nucleobases in the ISM.

4.2.2. From guanine to cytosine?

We were intrigued by the presence of products matching the molecular formula of cytosine $C_4H_5N_3O$ (see Figure 13). This result suggests that there could be a connection of the paths that lead to guanine with those that explain the formation of cytosine.

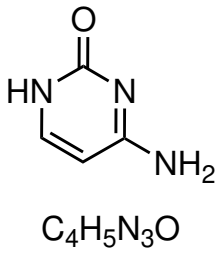


Figure 13: Molecule of cytosine.

We have found a set of 3 products with molecular formula $C_4H_5N_3O$, namely **PR128**, **PR517** and **PR556**.

4.2.2.1 Discussion of some selected reaction paths: from PR128 to guanine

This process starts with a radical-radical collapse of the two components of **PR128** (168.49 kcal/mol). This reaction results in the formation of **MIN34** (25.56 kcal/mol) barrierlessly. Subsequently, a hydrogen transfer occurs, yielding **MIN22** (21.81 kcal/mol) via transition state **TS46** (34.26 kcal/mol). Then, guanine is formed via a cyclization step that involves **TS490** (92.33 kcal/mol).

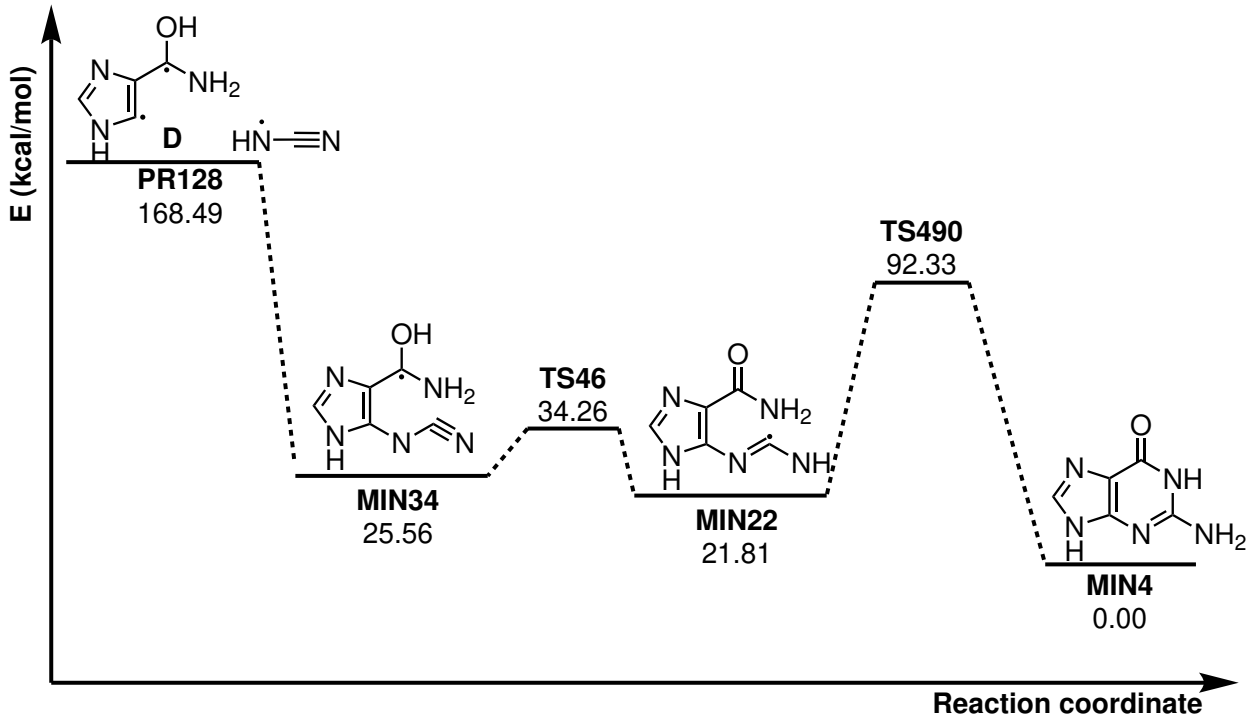


Figure 14: Description of Reaction Pathways Mechanisms and Energy Profile for Structural Isomer **PR128** of Cytosine.

Extended decomposition analysis of PR128-D

We envisioned that **PR128-D** would further decompose into a formamide derivative and an indole-like-core. Formamide is the simplest molecule containing a peptide bond, making it a possible precursor of complex molecules in prebiotic conditions. Saladino *et al.*[56] demonstrated that it could provide all the compounds needed for the formation, under prebiotic conditions, of nucleic polymers. Formamide has been detected in ISM towards the hot core of Sgr B2,[20] Sgr B2(N),[57] Hale-Bopp comet and within the solid phase of the ice grains in protostellar objects NGC 7538 IRS9 and W33A,[58] based on the qualitative comparison of the spectra of isocyanic acid at 10K subjected to ultraviolet photolysis under vacuum conditions with the experimental spectra.[59, 60]

AutoMeKin, proved our hypothesis correct, not only **D** can fragment into a formamide isomer, but also into an indole derivative (see Figure 15).

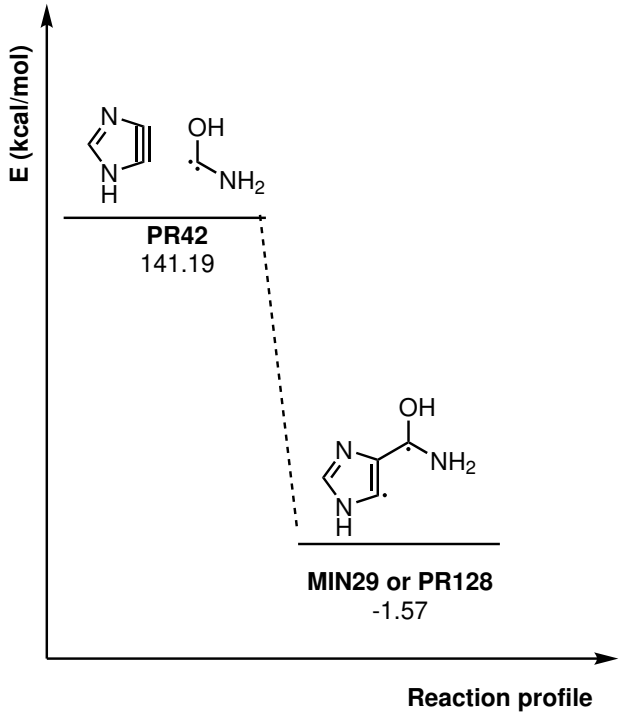


Figure 15: Depiction of a reaction path for the decomposition of **PR42**.

4.2.2.2 Discussion of some selected reaction paths: from PR517 to guanine

The reaction mechanism initiates with a radical-radical reaction, resulting in the formation of **MIN329** (72.08 kcal/mol). Subsequently, a hydrogen transfer between N9 and C8 (refer to Figure 2) leads to the formation of **MIN210** (59.13 kcal/mol), through transition state **TS1076** (137.93 kcal/mol). Finally, a cyclization step takes place yielding guanine (**MIN4**). The energy penalization of this last step is 3.78 kcal/mol.

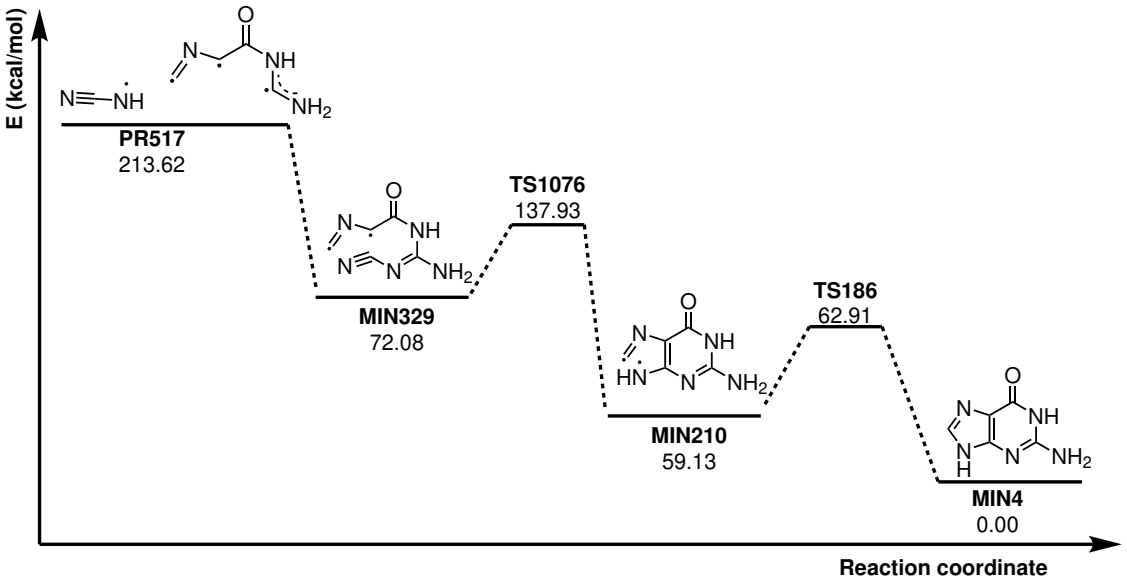


Figure 16: Reaction path for the formation of guanine from **PR517**.

4.2.2.3 Discussion of some selected reaction paths: from PR556 to guanine

The reaction mechanism starts, once again, with a radical-radical reaction resulting in the formation of **MIN554** (93.45 kcal/mol). Subsequently, a hydrogen transfer occurs via transition state **TS880** (123.49 kcal/mol), leading to the generation of **MIN264** (64.91 kcal/mol). Following this, cyclization takes place, yielding **MIN2** (-0.38 kcal/mol), facilitated by transition state **TS210** (67.12 kcal/mol). Finally, isomerization from enol to keto form takes place. As discussed in section **Analysis of the most stable minima obtained**, it is a favoured process due to the higher stability of the keto form, resulting in the formation of guanine (**MIN4**) through transition state **TS54** (35.63 kcal/mol).

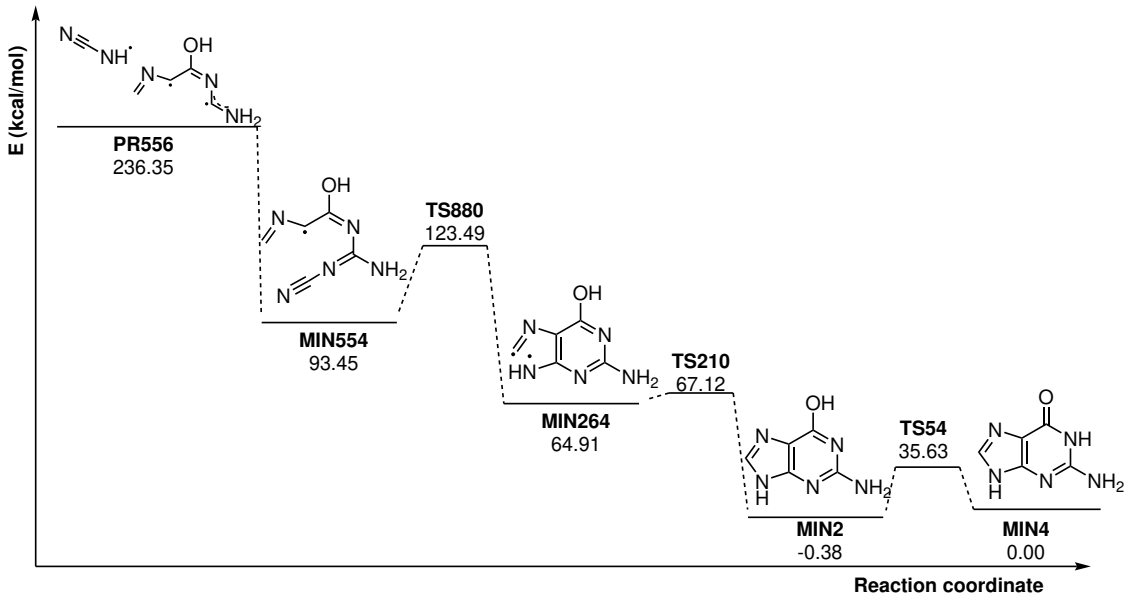


Figure 17: Energy profile for the reaction of **PR556** towards guanine.

4.2.3. Exploring radical-radical multi-step decomposition-derived products

4.2.3.1 Discussion of some selected reaction paths: reaction of (1H-imidazol-5-yl)-λ^2-azanecarbonitrile radical and formamide isomer to form guanine

We have found a four step mechanism (see Figure 18) that describes the formation of guanine from **PR129**. This mechanism starts with a barrierless radical-radical collapse (**PR129**, 164.56 kcal/mol). This step is followed by an H transfer between the amino and imino froups of **MIN34**. The penalty in energy of this hydrogen transfer is unusually low (8.7 kca/mol).

Once **MIN34** (25.56 kcal/mol) has evolved to **MIN22** (21.81 kcal/mol) through transition state **TS46** (34.26 kcal/mol), this intermediate can progress to guanine via a C-N bond formation with the concomitant closure of the ring. This cyclization step takes place via **TS490** (92.33 kcal/mol).

This product (see Figure 18) is fragmented into an isomer of formamide,[11] H_2NCOH which is not detected yet in the ISM.

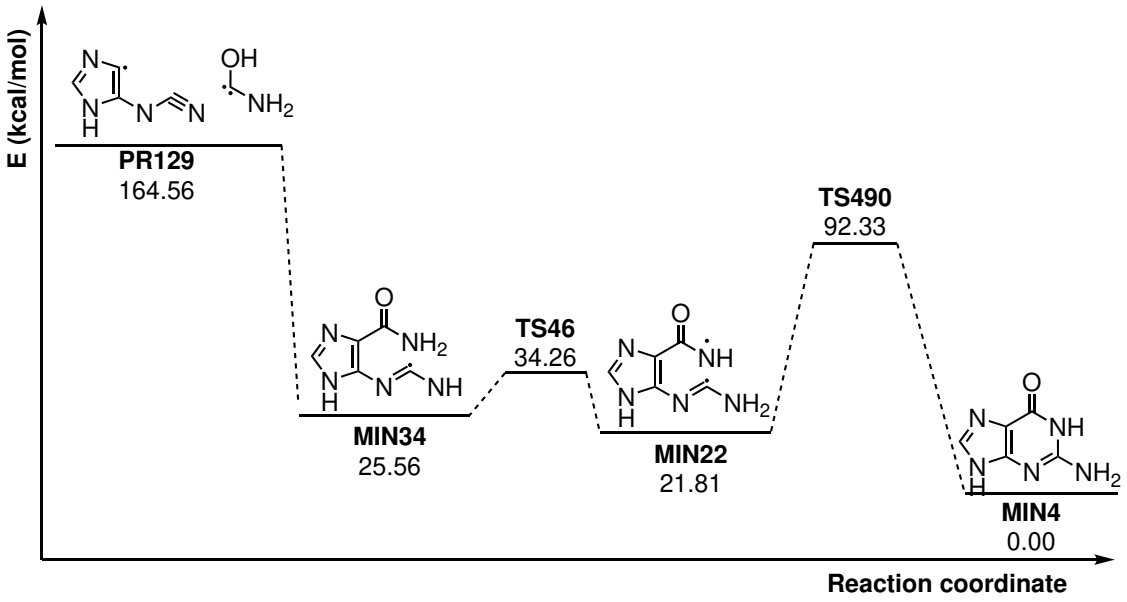


Figure 18: Energy profile of the evolution of **PR129** to guanine.

4.2.3.2 Discussion of some selected reaction paths: isocytosine and carbodiimide radicals

This product (see Figure 19) consist in carbodiimide (HNCNH) and aminopyrimidin-4(3H)-one radicals.

The reaction proceeds via a four-step mechanism (see Figure 19). Initially, a barrierless radical-radical reaction takes place between the two components of **PR568** (115.23 kcal/mol), leading to **MIN661** (104.79 kcal/mol). This barrierless step is followed by a cyclization step (**TS718**, 110.09 kcal/mol) that involves the formation of a C-C bond between the hexacycle and the carbon center of the dangling substituent and forming **MIN175** (54.48 kcal/mol).

Subsequently, the cyclized structure undergoes an intramolecular rearrangement that involves a ring opening from 4 centers to 5 centers (see Figure 2), through transition state **TS714** (109.95 kcal/mol), resulting in **MIN25** (22.57 kcal/mol). Finally, a hydrogen transfer between N7 and C8 occurs through transition state **TS213** (67.61 kcal/mol), yielding guanine, **MIN4**.

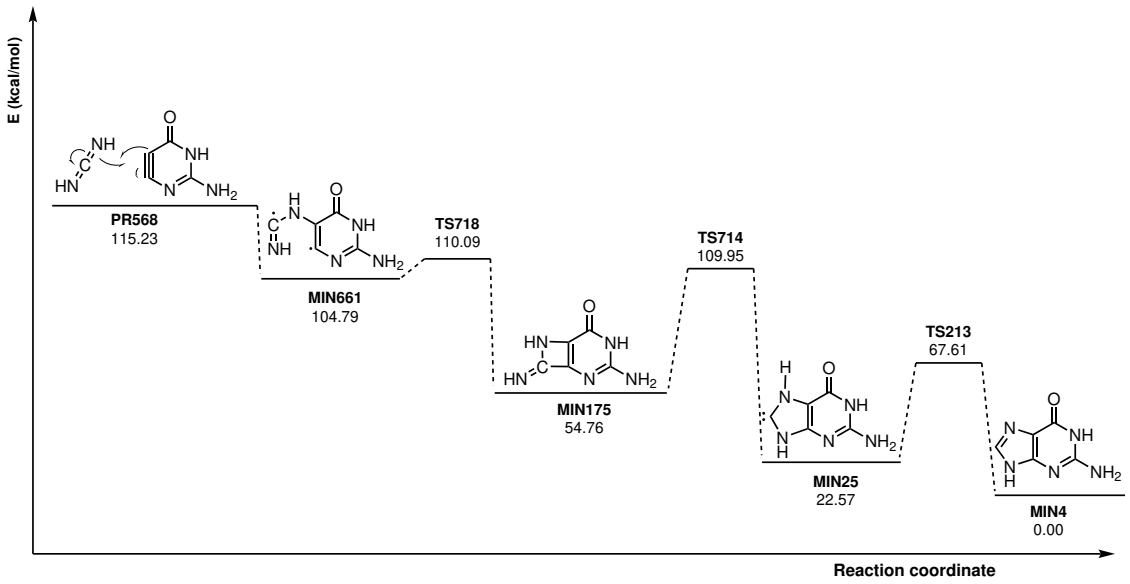


Figure 19: Reaction mechanism for the synthesis of guanine corresponding to the reaction mechanism involving isocytosine and carbodiimide radicals.

Carbodiimide is the second most stable isomer of empirical formula CH_2N_2, being the first cyanamide, for about 4 kcal/mol. Therefore, cabodiimide is expected to exist in the ISM in less concentration than cyanamide.[61] However, both are considered as key molecules in prebiotic chemistry.

5. Conclusions and future work

Overall, we have described here several different paths that can account for the formation of guanine in the ISM, and more specifically in its cold regions. The paths discussed range from 1-step to 4-steps but this is only a selection of the complete decomposition PES of guanine.

While we have not found a complete set of precursors (detected in the ISM) due to the time constrains of this Bachelor Thesis, we do have found that guanidine, indole and formamide derivatives might play a key role in the formation of guanine. These findings, allow us to hypothesize that these molecules will eventually be found in the ISM.

Further studies will be carried out to get a complete understanding on the formation of this molecule. Specifically:

- Decomposition of the found products, to further understand how these species are generated in the ISM.
- Analysis of the stability of the wave-function of all stationary points here described. Note that we have described not only radicals, but also carbenes which might show multireference character.
- Energy refinement, via recomputation of the PES at CASPT2 level.

Conclusiones y trabajo futuro

En general, se han descrito diferentes caminos que pueden explicar la formación de guanina en el ISM, y más específicamente en sus regiones frías. Los caminos aquí discutidos se ciñen a aquellos que tienen hasta 4 etapas elementales, pero esta es solo una selección de la PES completa de la descomposición de guanina.

Aunque no se han encontrado un conjunto completo de precursores (detectados en el ISM), debido a las limitaciones temporales de este Trabajo Fin de Grado, hemos encontrado que los derivados de la guanidina, indol y formamida podrían desempeñar un papel clave en la formación de la guanina. Estos hallazgos nos permiten hipotetizar que, con el tiempo, estas moléculas serán encontradas en el ISM.

Se llevarán a cabo estudios adicionales para obtener una comprensión completa sobre la formación de esta molécula. Específicamente:

- Descomposición de los productos encontrados, para entender mejor cómo se generan estas especies en el ISM.
- Análisis de la estabilidad de la función de onda de todos los puntos estacionarios aquí descritos. Cabe destacar que hemos descrito no solo radicales, sino también carbenos que podrían mostrar carácter multirreferencia.
- Optimización de la energía, mediante la recomputación de la PES con un nivel CASPT2.

Conclusións e traballo futuro

Globalmente, neste traballo explicamos distintos camiños que poden explicar a formación de guanina no ISM, e máis concretamente nas súas rexións frías. Os camiños discutidos inclúen só camiños de 1 a 4 etapas elementais, pero esta é só unha selección da PES completa da descomposición de guanina.

Aínda que non encontramos un conxunto completo de precursores (detectados no ISM) debido ás limitacións temporais desta Traballo Fin de Grado, atopamos que os derivados da guanidina, indol e formamida poderían xogar un papel clave na formación de guanina. Estes achados permítennos hipotetizar que, co tempo, estas moléculas serán atopadas no ISM.

Realizaranse estudos adicionais para obter unha comprensión completa sobre a formación desta molécula. Concretamente:

- Descomposición dos produtos atopados, para entender mellor como se xeran estas especies no ISM.
- Análise da estabilidade da función de onda de todos os puntos estacionarios aquí descritos. Sinalamos que describimos non só radicais, senón tamén carbenos que poderían mostrar carácter multirreferencia.
- Refinamento da enerxía, mediante a recomputación do PES cun nivel CASPT2.

A. Appendix

A.1. Exploring the Cosmos: From the Bing Bang to Life's Origin Theories

This section of this work presents a brief summary of the fundamental building blocks of the universe, based on the book "Astrochemistry: The Physical Chemistry of the Universe" of Andrew M. Shaw.[62] Chapter I, V and VIII, titled "The Molecular Universe", "The Interestellar Medium" and "Prebiotic Chemistry" have contributed to the development of the theory presented here.

A.1.1. The Standard Model: Big Bang Theory

Approximately 13.722 ± 0.059 billion years ago,[63, 64] the Universe and time itself began. It is well established that stars and galaxies are expanding, implying that in the distant past, a common point or singularity in space-time, known as the Big Bang, likely existed, marking the origin of everything.

The Standard Model of Cosmic Evolution is a straightforward theory that incorporates six parameters to fit all cosmological data: the age of the universe, the density of atoms, the scale dependence of amplitude fluctuations, and the epoch of the first star formation.[65] To understand the universe's evolution, the Wilkinson Microwave Anisotropy Probe (WMAP) measured temperature differences across the sky in the cosmic microwave background (CMB), constructing a cosmic timeline. Each observation by the WMAP satellite captures the universe's history, beginning from the Planck epoch, characterized by uncertainty, through a massive inflation by a factor of 180e (where e is the base of natural logarithms).

Prior to this inflation, the Universe was a hot, dense mixture of charged particles, primarily electrons and protons, which constantly interacted with photons. These interactions prevented light from traveling long distances without scattering. However, as the Universe cooled, protons and electrons combined to form neutral hydrogen atoms, significantly reducing the number of charged particles available to interact with photons. This transition allowed photons to travel freely, rendering the universe transparent. The radiation formed by high-energy photon collisions during this period is known as the cosmic microwave background (CMB), which we observe today as key evidence of the Big Bang and the early structure of the universe. This afterglow provides a snapshot of the universe 375,000 years after the Big Bang (see Figure 20).

Temperature plays a crucial role in the universe's phases of evolution and its subsequent cooling, leading to several critical periods, as detailed in Table 1. These intervals are predictions of the Big Bang Theory or the Standard Model of Cosmic Evolution.[66] Additionally, a temperature map of the early universe can be

seen in Figure 21.[67]

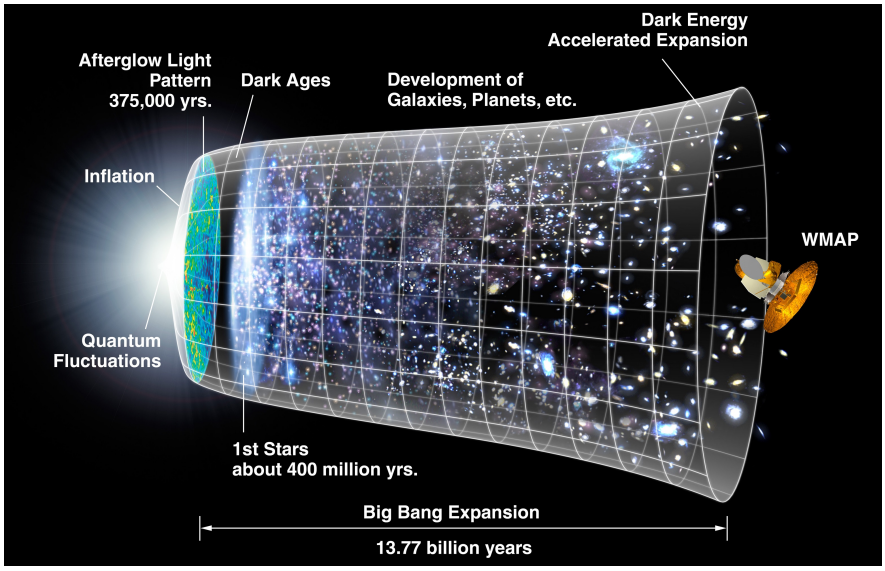


Figure 20: Cosmic timeline observed today by studying the microwave background.[68]

Table 1: History of the Universe, based on the work developed by Ratra and Vogeley.[66]

Time (s)	Observation	Temperature (K)
10^{-43}	Gravity splits from strong, weak, and electromagnetic forces.	10^{32}
10^{-35}	Strong forces separate. Expansion of the Universe.	10^{27}
10^{-12}	Weak and electromagnetic forces split. Neutrons and protons are formed by the collision of photons.	10^{15}
10^{-2}	Electrons and positrons are formed by the collision of photons.	10^{11}
1	Universe becomes transparent to neutrinos.	10^{10}
180	Nucleosynthesis of hydrogen, helium, deuterium, and some lithium begins.	10^{9}
$3 - 7 \times 10^{5}$	Universe is now transparent to radiation and cosmic background is emitted. Light atoms formation.	3000
10^{9}	Galaxies are formed.	20
Present	Stars and galaxies are created.	2.276

In the early Universe, where temperature was high, collisions between high-energy photons produced particle-antiparticle pairs, such as protons and antiprotons. After about 180 seconds, when the temperature was 10^9 K, atomic nuclei like hydrogen, deuterium, helium, and some lithium began to form. The first three minutes of the universe were chemically non interesting, with no atoms or molecules present in it. Over the next 10^6 seconds, light atoms continued to form, marking the period nucleosynthesis, during which matter was created.

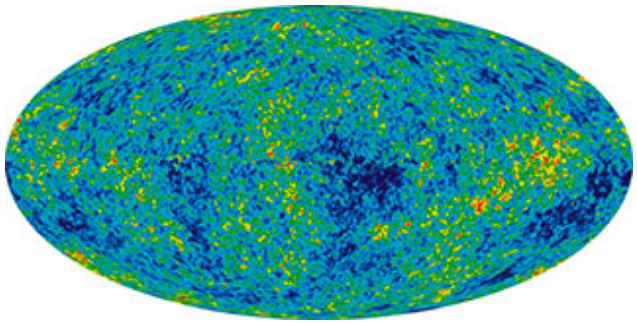


Figure 21: Picture of the infant universe, based on the data collected by WMAP for nine years. Credit: NASA / WMAP Science Team.[67]

Most of the atomic matter in the universe consists of hydrogen and helium produced during the Big Bang, with some helium formed later. The relative cosmic abundances of elements important for life are shown in Table 2.[65] All elements heavier than hydrogen, helium, and lithium were created through fusion processes within stars. The cosmic abundances are assumed to be like the composition of the Sun.

Table 2: Abundance of Chemical Elements Taking Hydrogen as the Unity.

Element	**Relative Abundance**	**Element**	**Relative Abundance**
H	1	S	1.6×10^{-5}
He	0.085	P	3.2×10^{-7}
Li	1.5×10^{-9}	Mg	3.5×10^{-5}
C	3.7×10^{-3}	Na	1.7×10^{-6}
N	1.2×10^{-3}	K	1.1×10^{-7}
O	6.7×10^{-3}	Si	3.6×10^{-6}

A.2. The Genesis of Galaxies, Stars, Planets, and Life

After the initial photon collisions and matter formation, gravity pulled it together, leading to the creation of the first proto-stars before any starlight existed. This period is known as the epoch of first star formation or dark ages. Within 1 billion years, the first massive proto-galaxies formed. Continued gravitational contraction in confine regions led to the galaxies that can be seen nowadays, including the Milky Way (see Figure 22).[69] It was formed within 1 billion years after the Big Bang and has a mass of about 10^9 solar masses. It originated from a large, slowly rotating cloud of hydrogen and helium. As this cloud collapsed, the conservation of angular momentum caused matter near the axis to rotate rapidly, spreading out to form a flat, spiral disc approximately 120000 light-years in diameter and about 3300 light-years thick. The Sun is located about 30000 light-years from the center of it. The core of the galaxy is known as the nuclear bulge

and contains old stars. It is thought to be extremely massive.

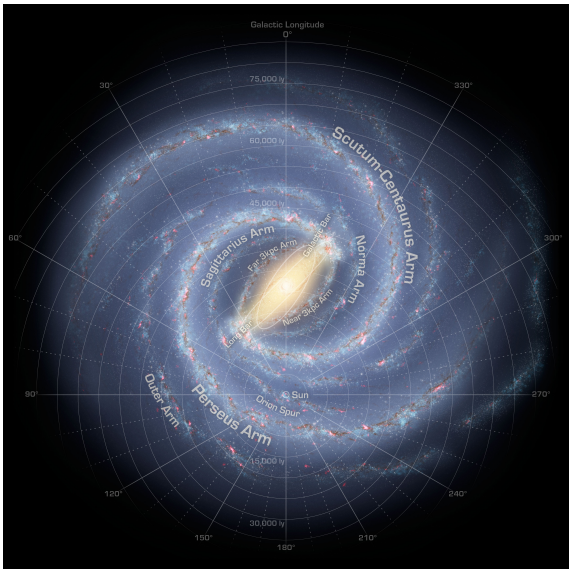


Figure 22: Milky Way map developed using infrared images of NASA's Spitzer Space Telescope. Credits: NASA/JPL-Caltech/R. Hurt (SSC/Caltech).[69]

The Sun (see Figure 23) formed about 4.5 Gyr ago (Gyr is gigayear, or 10^9 years) from its own gas cloud called the solar nebula. This was primarily composed of hydrogen but also contained all the heavier elements observed in the Sun's spectrum. The elemental abundance on Earth and all the planets was determined by the composition of the solar nebula, as already said before and summarized in the Table 2 which ultimately provided the molecular list of chemical elements necessary for life.

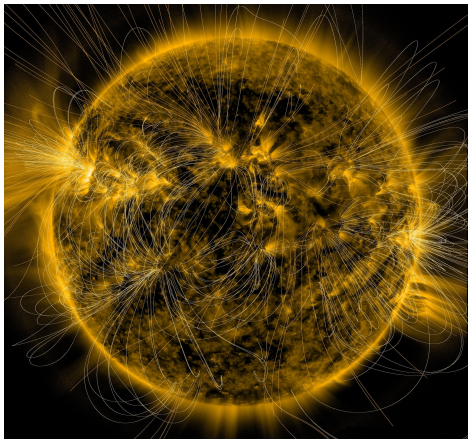


Figure 23: Suns magnetic field captured by NASA's Solar Dynamics Observatory on March 12, 2016. Credits: NASA/SDO/AIA/LMSAL.[70]

The solar system started from a slowly rotating nebula that contracted around the proto-sun, creating the

actual system of planets. New astronomical discoveries have found other stars with comparable planetary systems, indicating that planet formation around stars is a common process. In our solar system, most mass is in the Sun, with rocky inner planets (Mercury, Venus, Earth, Mars) and gas giants (Jupiter, Saturn, Uranus, Neptune).

The Earth (see Figure 24) is approximately 4.55 Gyr, determined through radioisotope dating. In the first billion years, it experienced significant impact events that could sterilize it and caused mass extinctions. Regardless of this, geological fossil records indicate that life existed around 3.5 billion years ago.

Figure 24: Earth. Picture took by Apollo Mission. Credits: NASA.[71]

The earliest life forms were simple although they already possessed complex structures that involved membranes and genetic information. It is startling that life may have developed in as little as 100 million years and at most 0.5 billion years, evolving from a primordial soup to organisms capable of adapting to their surroundings. But what it is life?. NASA defined it as:

"Life is a self-sustained chemical system capable of undergoing Darwinian evolution."

Andrew M. Shaw defines it as:

"A system that is capable of metabolism and propagation of information."

Now, cellular life could have spontaneously originated from the primordial soup. Encapsulation process led to the fist specialization within the surroundings, leading to, firstly, external compartmentalisation and then, internal. Genetic information is stored in DNA or RNA. They are huge polymeric molecules that contain all the information for the replication of the blocks that form all organisms: proteins. The four bases,

G, A, T, and C, pair together as A–T and G–C, known as Watson–Crick base pairs. They form the double helix structure of the DNA molecule together with the desoxyribose sugar and phosphate backbone. This information is propagated from generation to generation.

Life is very varied and this reflects the capacity of adaptation and survival ability in extreme and complex environments, as in the case of bacteria. Organisms can be classified based on the length of its genome. This suggests a common ancestor: hyperthermophile bacterium. Contrary to popular belief, rather than being descended from apes, there is evidence that we may be descending from bacteria. Experiments measuring the survival of bacteria in space indicate that bacterial spores or dried cells could survive in space for at least eighteen years. This allows the possibility of life being transferred from one planet to another. Recent analysis of the meteorite ALH84001 suggests the presence of structures resembling fossilized organisms within it.[72] This meteorite likely originated from Mars, ejected by a collision and reaching Earth in a rapid transit, possibly within 60,000 years. This scenario implies the intriguing possibility that our ancestors could have been Martian bacteria.

A.2.1. Exploring the Origins of Life: Theoretical Perspectives

There is not one definitive theory explaining how life began on Earth, although numerous theories have been proposed. Current ideas have a series of paradoxes that includes the organic inventory of molecules, self-replicating molecules, RNA preceding proteins, and the evolution of species. Theories (see Figure 25) that can be tested 'experimentally' fall into two categories: panspermia (or extraterrestrial), suggesting life was seeded randomly throughout the universe, and abiogenesis (or terrestrial), proposing life originated independently on Earth or other habitable planets. **While theories of terrestrial origin are more commonly accepted, the recent discovery of possible habitable planets and the potential for life within any solar system increases the probability of panspermia.**

Terrestrial theories split into the organic or inorganic origin of life. Starting with organic molecules, how they could originate in prebiotic conditions? The continuous production of organic material within a planet offers a direct connection from prebiotic conditions for the development of full organisms. Anyways, it is probable that habitable conditions let organic material from external sources to reach Earth (delivered by meteorites or comets). Energy for life could have been obtained from organic sources, such as photosynthesis or inorganic processes such as the ones occurring in hydrothermal vents. It is even conceivable that inorganic surfaces catalysed the formation of the initial self-replicating or primitive molecules.

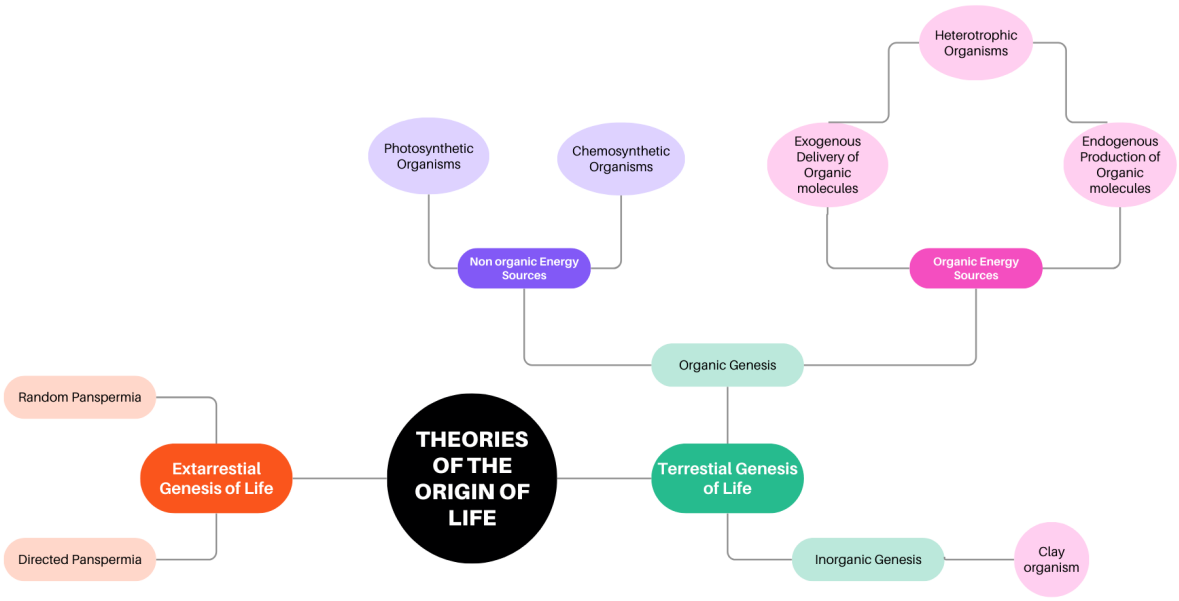


Figure 25: Theories of the Origin of Life. Adapted from Andrew M. Shaw.[62]

Extraterrestrial theories propose that life could emerge whenever the chemistry of it could develop, either randomly or guided by an unknown force. The discovery of extrasolar planets within habitable zones, where liquid water can exist, suggests that conditions are fulfilled for the spontaneous beginning of life. However, it's crucial not to assume that what we observe nowadays has a direct correlation with the prebiotic Earth or any other planet. Mass extinction events are common in Earth's fossil record. There were five major events. Among these, the most well-known is the meteorite impact that led to the extinction of the dinosaurs.

In the early times of Earth, impacts were frequent, this is evidenced by the craters on the Moon's surface. Also, fossil records indicate early collisions that significantly reduced the diversity of species, potentially eliminating up to 90 percent of them. Some organisms may have survived by obtaining biologically active molecules from the debris of other species.

The evolution of modern life, as it is known now, may differ substantially from the spontaneous emergence of life through prebiotic chemistry, leading to diverse challenges for astrobiologists.

A.3. Delving into Interstellar Medium Chemistry

In this section, the information was gathered from the encyclopedia of Astrobiology,[1] specially from the Section - Astrochemistry: Models and Observations: V. Wakelam and Section - Prebiotic Chemistry: R. Saladino.

According to it, the definition of Interstellar Chemical Processes is:

"Interstellar chemical processes manipulate matter by either creating or destroying molecular material. These processes, acting on gas and dust, include heating, irradiation, shocks, and cosmic ray bombardment, as well as chemical reactions such as ion-neutral and neutral-neutral reac- tions and surface chemistry."

All material present in the ISM is cycled during the different phases of stellar life, from the diffuse medium to the formation of a new star or planetary system, until the death of them. Physical and chemical processes that take place conform the complexity of each phase. Generally, low densities and moderate temperatures characterize the ISM and its characteristics differ with respect to Earth and the local environments. As discussed earlier, these environments can be divided into different regions based of its density and its temperature, they are summarized in Table 3:

Table 3: Regions of the ISM characterized by their atomic density and temperature. Adapted from Wooden *et al.*[73] and Encyclopedia of Astrobiology.[1]

Region	**Ordinary name**	**Temperature (K)**	**Density of atoms (cm^{-3})**
Hot ionized medium	Coronal gas	10^6	0.003
Warm ionized medium	Diffuse ionized gas	10^4	> 0.1
Warm neutral medium	Intercloud H I	10^3-10^4	0.1
Atomic cold neutral medium	Diffuse clouds	100	10-100
Molecular cold neutral medium	Molecular clouds, dense clouds, dark clouds	0-50	10^3-10^5
Hot molecular cores	Protostellar cores	100-300	10^3-10^7

Now a general outlook for the several interstellar processes taking place in the ISM is going to be examined.

A.3.1. Gas-phase Chemistry

Stellar atmospheres (especially old stars) expel matter in the form of matter and dust. Most of the molecules formed do not survive the extreme conditions of the ISM and convert into smaller molecules by photo-dissociation or fragmentation. Other molecules will remain and become part of the material of the

surroundings. In molecular clouds, whereas density is high and there is protection against UV radiation, chemistry will take place significantly. Anyways, in other parts, where low density predominates, two-body collisions are the main chemical processes. Radiation, either UV and EUV (extreme ultraviolet radiation) play a crucial role in the ISM chemistry. This reactions can be classified in:

A.3.1.1 Photochemistry

The ISM regions with a high UV activity, such as hot massive starts, are the so called photodissociation regions. They are dominated by the formation and destruction of hydrogen.[74] The recombination of electrons and charge transfer mechanism keep the ionization balance. In this terms, gas density and pressure, intensity of the radiation present, the scattering due to dust and the abundance of the different elements will be the variables affecting this reacitons. Inside this regions, subregions can be discerned depending of the optical depth, this is a dimensionless quantity that measures how absorbing or scattering a part of medium is for an electromagnetic wave that propagates on it. The most exposed region to radiation has atomic hydrogen, helium, oxygen and ionized carbon.[75] High temperatures are achieved (100-1000K). The next region is characterized by the formation of molecular hydrogen although carbon remains in its ionized form. In the more inner or shielded regions, carbon is converted to CO and temperatures are lower to 10-100K due to the own rotational emission of this molecule.

A.3.1.2 Ion-Neutral Chemistry

Low temperatures influence the ISM chemistry, it requires exothermic reactions and low densities restrict reactions to mainly two bodies collision. This is fullfilled by reactions involving ions (either cations or anions) and tend to dominate gas-phase chemistry. Cosmic ray bombardment, this is, high energy electrons and nuclei that are accelerated by supernova explosions and massive stellar winds that travel across the Galaxy,[76] influences in the formation of positive ions. Ionization, dissociative recombinations , radiative association and isotope exchange reactions are significant ion-neutral chemical reactions. Anions are formed by the combination of an electron with a neutral specie.

Neutral-Neutral Chemistry It involves interaction between not charges species, they are typical endothermic with a large activation energy due to the fact that it is the interaction of two species with no unpaired electrons. Nevertheless, the barrier can be lowered by the presence of a radical or an ion and then this type of reactions become interesting in the processes of formation and destruction of molecules.

A.3.1.3 Surface Chemistry

Stars in evolution form dust and grains that are dispersed in the ISM. At low temperatures, molecules can collide or add to its surface and form an ice layer, leading to the possibility of the formation of different molecules, specially hydrogenated species than, afterwards are emitted in gas phase. This processes dominate the synthesis of molecular hydrogen, althought the limited knowledge of the composition and structure of this grains and dust do not let to find exact mechanisms for the chemical reactions such as the formation of H_2. They are considered as the main source of hydrogenation of C, O, N and S to form CH_4, NH_3, H_2O and H_2S. Methanol can also be produced by CO surface reactions.[77]

Finally a summary table is given:

Table 4: Examples of different types of reactions. Adapted from the Encyclopedia of Astrobiology. [1]

Reaction	**Example**
Ion-neutral	$H_3^+ + O \longrightarrow H_2O^+ + H$
Neutral-neutral	$OH + H_2 \longrightarrow H_2O + H$
Photoionization	$H_2O + h\nu \longrightarrow H_2O^+ + e^-$
Photodissociation	$H_2O + h\nu \longrightarrow OH + H$
Photodetachment	$C_6H + h\nu \longrightarrow C_6H^+ + e^-$
Radiative association	$H_2 + S^+ \longrightarrow H_2S^+ + h\nu$
Dissociative recombination	$H_3O^+ + e^- \longrightarrow H_2O + H$
Associative detachment	$C_6H + H \longrightarrow C_6H_2 + e^-$

References

[1] Gargaud, M.; Amils, R. *Encyclopedia of Astrobiology*; Encyclopedia of Astrobiology; Springer, 2011; Vol. 1.

[2] McMenamin, M.; Langmuir, D.; Margulis, L.; Vernadsky, V.; Ceruti, M.; Golubic, S.; Guerrero, R.; Ikeda, N.; Ikezawa, N.; Krumbein, W.; others *The Biosphere*; A Peter N. Nevraumont book; Springer New York, 1998.

[3] Marov, M. Y. Astronomical and Cosmochemical Aspects of the Life Origin Problem. *Astron. Rep.* **2023**, *67*, 764–789.

[4] Puzzarini, C. Gas-phase Chemistry in the Interstellar Medium: The Role of Laboratory Astrochemistry. *Front. Astron. Space Sci.* **2022**, *8*.

[5] Bae, J. *et al.* Molecules with ALMA at Planet-forming Scales (MAPS): A Circumplanetary Disk Candidate in Molecular-line Emission in the AS 209 Disk. *ApJL* **2022**, *934*, L20.

[6] Müller, H. S. P. Formaldehyde in Space. https://cdms.astro.uni-koeln.de/classic/molecules:ism:h2co, 2019; Accessed: 2024-06-11.

[7] Müller, H. S. P. Glycolaldehyde in the ISM. https://cdms.astro.uni-koeln.de/classic/molecules:ism:glycolaldehyd?s[]=glycolaldehyde, 2021; Accessed: 2024-06-11.

[8] Müller, H. S. P. On Ethylene glycol in the ISM. https://cdms.astro.uni-koeln.de/classic/molecules:ism:glycol?s[]=ethylene&s[]=glycol, 2019; Accessed: 2024-06-11.

[9] Müller, H. S. P. On the Detection of Methyl Formate, HC(O)OCH3, in the ISM. https://cdms.astro.uni-koeln.de/classic/molecules:ism:methylformiat?s[]=methyl&s[]=formate, 2019; Accessed: 2024-06-11.

[10] Müller, H. S. P. On the Detection of Dimethyl Ether, $(CH_3)_2O$, in the ISM. https://cdms.astro.uni-koeln.de/classic/molecules:ism:dme?s[]=dimethyl&s[]=ether, 2021; Accessed: 2024-06-11.

[11] Müller, H. S. P. Reports on Formamide, HC(O)NH, in the ISM. https://cdms.astro.uni-koeln.de/classic/molecules:ism:formamid?s[]=formamide, 2019; Accessed: 2024-06-11.

[12] Müller, H. S. P. On Methylamine, CH_3NH_2, in the ISM. https://cdms.astro.uni-koeln.de/classic/molecules:ism:menh2?s[]=methylamine, 2023; Accessed: 2024-06-11.

[13] CDMS (Cologne Database for Molecular Spectroscopy) CDMS Molecules Database. https://cdms.astro.uni-koeln.de/classic/molecules, Accessed: 2024-06-09.

[14] Belloche, A.; Menten, K. M.; Comito, C.; Müller, H. S. P.; Schilke, P.; Ott, J.; Thorwirth, S.; Hieret, C. Detection of amino acetonitrile in Sgr B2(N). *A&A* **2008**, *482*, 179–196.

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

[16] Halfen, D. T.; Apponi, A. J.; Woolf, N.; Polt, R.; Ziurys, L. M. A Systematic Study of Glycolaldehyde in Sagittarius B2(N) at 2 and 3 mm: Criteria for Detecting Large Interstellar Molecules. *ApJ* **2006**, *639*, 237–245.

[17] Snyder, L. E.; Buhl, D. Observations of Radio Emission from Interstellar Hydrogen Cyanide. *ApJL* **1971**, *163*, L47.

[18] Johansson, L. E. B.; Andersson, C.; Ellder, J.; Friberg, P.; Hjalmarson, A.; Hoglund, B.; Irvine, W. M.; Olofsson, H.; Rydbeck, G. The spectra of Orion A and IRC + 10216 between 72.2 and 91.1 GHz. *A&AS* **1985**, *130*, 227.

[19] Ziurys, L. M. The chemistry in circumstellar envelopes of evolved stars: following the origin of the elements to the origin of life. *PNAS* **2006**, *103*, 12274–12279.

[20] Rubin, R. H.; G. W. Swenson, J.; Benson, R. C.; Tigelaar, H. L.; Flygare, W. H. Microwave Detection of Interstellar Formamide. *ApJ* **1971**, *169*, L39–L44.

[21] Belloche, A.; Garrod, R. T.; Müller, H. S. P.; Menten, K. M.; Medvedev, I.; Thomas, J.; Kisiel, Z. Re-exploring Molecular Complexity with ALMA (ReMoCA): interstellar detection of urea. *A&A* **2019**, *628*, A10.

[22] Saladino, R.; Botta, G.; Delfino, M.; Mauro, E. D. Meteorites as catalysts for prebiotic chemistry. *J. Chem.* **2013**, *19*, 16916–16922.

[23] Barks, H.; Buckley, R.; Grieves, G.; Di Mauro, E.; Hud, N.; Orlando, T. Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation. *ChemBioChem* **2010**, *11*, 1240–3.

[24] Jeilani, Y. A.; Nguyen, H. T.; Newallo, D.; Dimandja, J.-M. D.; Nguyen, M. T. Free radical routes for prebiotic formation of DNA nucleobases from formamide. *Phys. Chem. Chem. Phys.* **2013**, *15*, 21084–21093.

[25] da Silva, J. B. P.; de Araujo, A. P. M. A New Mechanism of Guanine-Isomer Formation from Species Previously Observed in the Interstellar Medium. *ACS Earth and Space Chem.* **2017**, *1*, 376–383.

[26] Sanchez, R. A.; Orgel, L. E. Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Bio.* **1970**, *47*, 531–543.

[27] Levy, M.; Miller, S. L.; Oró, J. Production of guanine from NH_4CN polymerizations. *J. Mol. Evol.* **1999**, *49*, 165–168.

[28] Miyakawa, S.; Cleaves, H. J.; Miller, S. L. The Cold Origin of Life: Implications Based On The Hydrolytic Stabilities Of Hydrogen Cyanide And Formamide. *Orig. Life Evol. Biosph.* **2002**, *32*, 195–208.

[29] Miyakawa, S.; Yamanashi, H.; Kobayashi, K.; Cleaves, H. J.; Miller, S. L. Prebiotic synthesis from CO atmospheres: implications for the origins of life. *Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 14628–14631.

[30] Hayatsu, R.; Studier, M. H.; Matsuoka, S.; Anders, E. Origin of organic matter in early solar system—VI. Catalytic synthesis of nitriles, nitrogen bases and porphyrin-like pigments. *GCA* **1972**, *36*, 555–571.

[31] Yang, C. C.; Oró, J. Synthesis of adenine, guanine, cytosine, and other nitrogen organic compounds by a Fischer-Tropsch-like process. 1971.

[32] Martínez-Núñez, E.; Barnes, G. L.; Glowacki, D. R.; Kopec, S.; Peláez, D.; Rodríguez, A.; Rodríguez-Fernández, R.; Shannon, R. J.; Stewart, J. J. P.; Tahoces, P. G.; Vazquez, S. A. AutoMeKin2021: An open-source program for automated reaction discovery. *J. Comput. Chem.* **2021**, *42*, 2036–2048.

[33] Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.

[34] Stewart, J. J. P. MOPAC2016. web-site: HTTP://OpenMOPAC.net, 2016.

[35] Zhao, Y.; Truhlar, D. G. Exploring the limit of accuracy of the global hybrid meta density functional for main-group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

[36] Frisch, M. J. *et al.* Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

[37] Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.

[38] Garay-Ruiz, D. Visualization tools for AutoMeKin. 2021; https://github.com/dgarayr/amk_tools, Institute of Chemical Research of Catalonia.

[39] Garay-Ruiz, D.; Álvarez Moreno, M.; Bo, C.; Martínez-Núñez, E. New Tools for Taming Complex Reaction Networks: The Unimolecular Decomposition of Indole Revisited. *ACS Phys. Chem. Au* **2022**, *2*, 225–236.

[40] Jones, D. B.; Wang, F.; Winkler, D. A.; Brunger, M. J. Orbital based electronic structural signatures of the guanine keto G-7H/G-9H tautomer pair as studied using dual space analysis. *Biophys. Chem.* **2006**, *121*, 105–120.

[41] Chin, W.; Mons, M.; Dimicoli, I.; Piuzzi, F.; Tardivel, B.; Elhanine, M. Tautomer contributions to the near UV spectrum of guanine: towards a refined picture for the spectroscopy of purine molecules. *Eur. Phys. J. D At. Mol. Opt. Phys.* **2002**, *20*, 347.

[42] Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. Guanine tautomerism revealed by UV–UV and IR–UV hole burning spectroscopy. *J. Chem. Phys.* **2001**, *115*, 4604.

[43] Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L.; LeBreton, P. Ultraviolet photoelectron studies of the ground-state electronic structure and gas-phase tautomerism of hypoxanthine and guanine. *J. Phys. Chem.* **1980**, *84*, 1006.

[44] LeBreton, P.; Yang, X.; Urano, S.; Fetzer, S.; Yu, M.; Leonard, N.; Kumar, S. Photoemission properties of methyl-substituted guanines: photoelectron and fluorescence investigations of 1,9-dimethylguanine, 06,9-dimethylguanine, and 9-methylguanine. *J. Am. Chem. Soc.* **1990**, *112*, 2138.

[45] Giese, B.; McNaughton, D. Density functional theoretical (DFT) and surface-enhanced Raman spectroscopic study of guanine and its alkylated derivatives Part 1. DFT calculations on neutral, protonated and deprotonated guanine. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5161–5170.

[46] Piacenza, M.; Grimme, S. Systematic quantum chemical study of DNA-base tautomers. *J. Comput. Chem.* **2004**, *25*, 83–99.

[47] Sabio, M.; Topiol, S.; Lumma, W. C. An investigation of tautomerism in adenine and guanine. *J. Phys. Chem.* **1990**, *94*, 1366–1372.

[48] Hanus, M.; Ryjáček, F.; Kabeláč, M.; Kubař, T.; Bogdan, T. V.; Trygubenko, S. A.; Hobza, P. Correlated ab initio study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment and in aqueous solution. Guanine: Surprising stabilization of rare tautomers in aqueous solution. *J. Am. Chem. Soc.* **2003**, *125*, 7678–7688.

[49] Nguyen, H. T.; Jeilani, Y. A.; Hung, H. M.; Nguyen, M. T. Radical Pathways for the Prebiotic Formation of Pyrimidine Bases from Formamide. *J. Phys. Chem.* **2015**, *119*, 8871–8883.

[50] Cologne Database for Molecular Spectroscopy CO (Carbon Monoxide) in Extragalactic Sources. https://cdms.astro.uni-koeln.de/classic/molecules:extragalactic:co_extragalactic?s[]=carbon&s[]=monoxide, Accessed: 2024-06-12.

[51] Millerl, S. L.; Jones, G. H. A production of Amino Acids under Possible Primitive Earth Conditions. *Ann. Entomol. Soc. Amer* **1953**, *117*, 528–529.

[52] Choe, J. Formation of Cytosine and Uracil from Cyanoacetylaldehyde and Guanidine: A Computational Study. *BKCS* **2020**, *41*, 382–384.

[53] Cologne Database for Molecular Spectroscopy HNCN (Cyanoamidogen) in Interstellar Medium. https://cdms.astro.uni-koeln.de/classic/molecules:ism:hncn?s[]=cyanoamidogen, Accessed: 2024-06-12.

[54] Rivilla, V. M.; Jiménez-Serra, I.; García de la Concepción, J.; Martín-Pintado, J.; Colzi, L.; Rodríguez-Almeida, L. F.; Tercero, B.; Rico-Villas, F.; Zeng, S.; Martín, S.; Requena-Torres, M. A.; de Vicente, P. Mon. Not. R. Astron. Soc. *RAS* **2021**, *506*, L79–L84.

[55] Podolyan, Y.; Gorb, L.; Blue, A.; Leszczynski, J. A theoretical investigation of tautomeric equilibria and proton transfer in isolated and hydrated thiocytosine. *THEOCHEM* **2001**, *549*, 101–109.

[56] Costanzo, G.; Saladino, R.; Crestini, C.; Ciciriello, F.; di Mauro, E. Formamide as the main building block in the origin of nucleic acids. *BMC Evol Biol.* **2007**, *7*.

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

[58] Bockelée-Morvan, D. *et al.* New molecules found in comet C/1995 O1 (Hale-Bopp). Investigating the link between cometary and interstellar material. *AAP* **2000**, *353*, 1101–1114.

[59] Raunier, S.; Chiavassa, T.; Marinelli, F.; Allouche, A.; Aycard, J. Reactivity of HNCO with NH3 at low temperature monitored by FTIR spectroscopy: formation of $NH_4^+OCN^-$. *Chem. Phys, Lett.* **2003**, *368*, 594–600.

[60] Schutte, W. A.; Boogert, A. C. A.; Tielens, A. G. G. M.; Whittet, D. C. B.; Gerakines, P. A.; Chiar, J. E.; Ehrenfreund, P.; Greenberg, J. M.; van Dishoeck, E. F.; de Graauw, T. Weak ice absorption features at 7.24 and 7.41 MU M in the spectrum of the obscured young stellar object W 33A. *AAP* **1999**, *343*, 966–976.

[61] Jabs, W.; Winnewisser, M.; Belov, S. P.; Lewen, F.; Maiwald, F.; Winnewisser, G. The structure of carbodiimide, HNCNH. *Mol. Phys.* **1999**, *97*, 213–238.

[62] Shaw, A. M. *Astrochemistry: The Physical Chemistry of the Universe*, 2nd ed.; WLY, 2022.

[63] Bennett, C. L.; Larson, D.; Weiland, J. L.; others Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: final maps and results. *ApJ Suppl S.* **2013**, *208*, 20.

[64] Ade, P. A. R. *et al.* Planck2013 results. XVI. Cosmological parameters. *A&A* **2014**, *571*, A16.

[65] Spergel, D. N. The dark side of cosmology: Dark matter and dark energy. *J. Sci.* **2015**, *347*, 1100–1102.

[66] Ratra, B.; Vogeley, M. S. The Beginning and Evolution of the Universe. *PASP* **2008**, *120*, 235–265.

[67] NASA Wilkinson Microwave Anisotropy Probe (WMAP) NASA WMAP: Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmology Results. 2022; https://map.gsfc.nasa.gov/media/121238/index.html.

[68] NASA Wilkinson Microwave Anisotropy Probe (WMAP) NASA WMAP: Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations. 2022; https://map.gsfc.nasa.gov/media/060915/index.html.

[69] NASA/JPL-Caltech/R. Hurt (SSC/Caltech) The Milky Way Galaxy. 2022; https://science.nasa.gov/resource/the-milky-way-galaxy/.

[70] NASA/SDO/AIA/LMSAL Image Detail: Active Region on the Sun, 2017. 2022; https://science.nasa.gov/image-detail/amf-gsfc_20171208_archive_e000393/.

[71] NASA Earth From Space. 2022; https://explorer1.jpl.nasa.gov/galleries/earth-from-space/#gallery-10.

[72] Frankel, R. B.; Buseck, P. R. Magnetite biomineralization and ancient life on Mars. *COCHBI* **2000**, *4*, 171–176.

[73] Wooden, D. H.; Charnley, S. B.; Ehrenfreund, P. In *Comets II*; Festou, M. C., Keller, H. U., Weaver, H. A., Eds.; 2004; p 33.

[74] Hollenbach, D. J.; Tielens, A. G. G. M. Photodissociation regions in the interstellar medium of galaxies. *Rev. Mod. Phys.* **1999**, *71*, 173–230.

[75] Sternberg, A.; Dalgarno, A. Chemistry in Dense Photon-dominated Regions. *APJS* **1995**, *99*, 565.

[76] Bergin, E. A. In *Physical processes in circumstellar disks around young stars*; Garcia, P., Ed.; University of Chicago Press: Chicago, 2009.

[77] Charnley, S. B.; Tielens, A. G. G. M.; Rodgers, S. D. Deuterated Methanol in the Orion Compact Ridge. *ApJ* **1997**, *482*, L203.