Prebiotic chemistry and origins of life research with atomistic computer simulations

Andrea Pérez-Villa, Fabio Pietrucci, A. Marco Saitta

To cite this version:

Andrea Pérez-Villa, Fabio Pietrucci, A. Marco Saitta. Prebiotic chemistry and origins of life research with atomistic computer simulations. Physics of Life Reviews, 2018, 10.1016/j.plrev.2018.09.004 . hal-01922088

HAL Id: hal-01922088
https://hal.sorbonne-universite.fr/hal-01922088v1

Submitted on 15 Dec 2022

Prebiotic chemistry and origins of life research with atomistic computer simulations

Andrea Pérez-Villa[a], Fabio Pietrucci[a], A. Marco Saitta[a]

[a] *Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, F-75005 Paris, France*

Abstract

Research in origins of life is an intrinsically multi-disciplinary field, aimed at finding answers to the formidably complex problem of understanding the emergence of life from the modern versions of Charles Darwin's celebrated "primordial soup". In the last few years, thanks to the increasing computational power and the development of sophisticated theoretical and numerical methods, several computational chemistry and physics groups have invested this field, providing new microscopic insights on fundamental prebiotic chemistry phenomena possibly occurring in the early Earth and outer space. This review presents the most successful and powerful approaches in computational chemistry, and the main results thus obtained in prebiotic chemistry and origins of life. The aim of this work is both to describe the state-of-the-art in computational prebiotic chemistry, possibly useful both to theorists and experimentalists in origins of life research, and to suggest future directions and new perspectives offered by modern simulation tools.

Keywords: Prebiotic Chemistry, Origins of Life, Atomistic Computer Simulations

---

1. Introduction

The study of the origins of life is a challenging and fascinating subject, that involves the interplay of different disciplines such as biology, chemistry, physics, astronomy, geology, philosophy, and so forth. The emergence of life on Earth, and its possible existence on other planets, has up until recently been an exclusively religious and philosophical question, rather than scientific. It was only in 1871 that Charles Darwin [1], in a famous letter to J. D. Hooker, expressed for the first time the hypothesis that life could have emerged "*in a warm little pond, with all sorts of ammonia and phosphoric salts, lights, heat, electricity*", thus formulating the now well-known idea of a "*primordial soup*". In the 1920s, A. I.

Preprint submitted to Physics of Life Reviews *September 12, 2018*

Oparin and J. B. S. Haldane, working from this idea, independently developed a complete geochemical theory of the abiotic origin of life from the aggregation of simple molecules such as water, ammonia, and methane. Experimental support of this scenario came later. In 1953, Stanley Miller and Harold Urey [2] carried out their famous experiment, showing that an electrical discharge applied to a mixture of simple gases, recreating the assumed primordial atmosphere, was capable of triggering chemical reactions leading to the formation of many organic compounds, and notably of amino acids. Similarly in 1960, Juan Orò managed to synthesize adenine from hydrogen cyanide at a high temperature [3]. With these founding experiments, the modern era of scientific study of the origins of life began.

In the bottom-up molecular perspective indicated by those early results, research in the field of prebiotic chemistry and origins of life (PCOL) has focused on the main molecular entities universally found in current living forms, and on the understanding of how and why those specific entities have been selected by chemical and biological evolution over the so-called prebiotic clutter: *(i)* proteins, *(ii)* nucleic acids, *(iii)* carbohydrates and *(iv)* lipids. In fact, given the overwhelming variety of chemically conceivable organic molecules, the fact of observing only a relative-small subset of them, often displaying a very specific chirality in biological systems, raised several questions.

One of the main challenges of PCOL research is the study of different reactive pathways that might have led to the synthesis of biological molecules in the primordial Earth. Dramatically different prebiotic scenarios have been explored to unveil their formation and their chemical evolution. Among them, interstellar medium, comets [4, 5] and interstellar ice (UV radiated) [6, 7] have shown the presence of organic prebiotic molecules, suggesting an exogenous source of life building blocks, linked with further meteoritic impacts on the Earth surface and atmospheric entry of cometary meteors providing sufficient energy to trigger different synthetic routes of prebiotic molecules on Earth [8, 9].

In addition, several endogenous sources such as the primordial soup in a reducing atmosphere mentioned above [10, 11], hydrothermal submarine vents or surface fields [12], and mineral surfaces [13, 14, 15], have been demonstrated to allow the synthesis of prebiotic molecules that progressively combined to form more complex chemical products [16, 17, 18]. The many plausible prebiotic scenarios mentioned above imply very different conditions of temperature, pressure, pH, solvent, ionic force, etc., that, in turn, profoundly affect the physico-chemical behavior of the systems. This is a crucial point: to make relevant advances in the origin of life field, the reductionist microscope should eventually be combined with a broad vision of the puzzle.

In this paper, we focus on modern atomistic computational approaches, *i.e.* explicitly treating each atom of a given chemical system of interest, ranging from gas-phase quantum chemistry to condensed-phase molecular dynamics. These methods have long demonstrated their capability to explain and predict the chemical behavior of a variety of systems at different scales (of length and time) and different conditions, but only recently begin to be adopted within the PCOL research field [19]. A crucial requirement for those methods to be useful in the context of prebiotic chemical reactions, is to accurately model the electronic structure of the systems to describe the breaking and formation of chemical bonds. To this aim, different available computational chemistry methods such as those based on the molecular orbital theory [20] (Hartree–Fock, Coupled Clusters, Møller–Plesset perturbation theory, and so forth), density functional theory [21], and empirical valence bond approaches [22], are able to accurately describe and predict the reorganization of electronic structure connected to changes in chemical bonding. Thanks to the continuous increase of available computer power, such accurate methods can nowadays be applied to challenging prebiotic problems, bringing valuable insights that are complementary to experiments.

Besides an accurate description of the electronic quantum degrees of freedom, the simulation of prebiotic reactions at realistic experimental conditions often requires, on one side, approaches like molecular dynamics that directly address temperature and, e.g., solvent effects, and on the other side, enhanced sampling techniques (such as umbrella sampling [23], metadynamics [24, 25], transition path sampling [26], etc.), able to overcome the fundamental time scale limitations of computer simulations compared to experiments. These latter approaches allow to discover non-trivial reaction mechanisms, to accelerate the crossing of free energy barriers associated to rare events, and to reconstruct the thermodynamics and kinetics of the process. In particular, aqueous solutions constitute a fundamental prebiotic scenario – that must have hosted several key reactions leading to life – of formidable complexity from the microscopic point of view, and it only recently became accessible to this kind of computer simulations. Breakthrough recent applications include the modeling of relevant steps in amino acid formation [27, 28, 29] and dimerization [30, 31, 32] and in the synthesis of purine and pyrimidine RNA nucleotides [33].

This review is organized as follows: Chapter 2 is devoted to the description of state-of-the-art approaches to atomistic computational chemistry. We first sketch the most important methods conceived to describe the many-body electron problem, generally grouped in the Quantum Chemistry (QC) and Density-Functional Theory (DFT) families. We will then describe the statistical approaches necessary to describe the dynamical and thermodynamical properties of a physico-chemical system, from molecular dynamics (MD) to

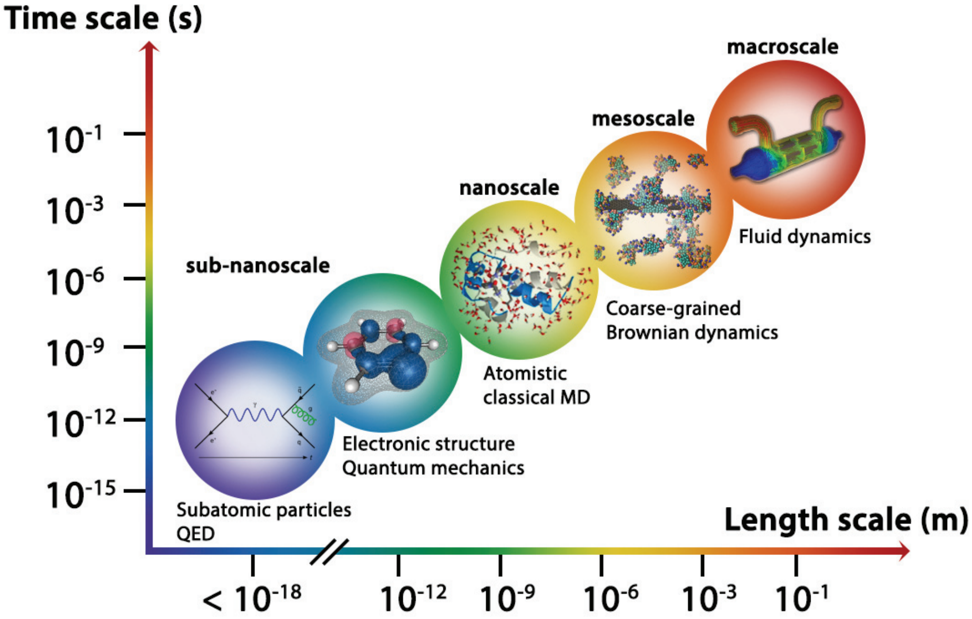

Figure 1: Computer simulations in chemistry. Implementation of theoretical approaches according to the size of the system and the time scale of the chemical processes.

advanced exploration and free-energy methods. Chapter 3 provides a review of the most important PCOL studies based on those computational approaches, underlining how a particular prebiotic scenario (interstellar medium, photochemistry, solution chemistry, surface chemistry, etc...) is typically better described within a particular subset of those numerical methods. Chapter 4 focuses on the emergence of new paradigms in the exploration of chemical reaction spaces, relying more and more on enhanced sampling approaches based on topological reaction coordinates. We present a few selected studies in PCOL, and stress how these novel theoretical tools might open new perspectives in the way we approach the origins of life research field.

2. Atomistic computer simulations in chemistry: standard and advanced approaches

Computational chemistry is the study of chemical systems by means of computer simulations. It comprises different types of approaches such as coarse-grained models, classical molecular dynamics, semi-empirical simulations and *ab initio* quantum chemistry methods (see Figure 1). Every method can assess a specific system size and time scale of chemical processes, displaying different levels of accuracy in terms of molecular geometries and energetics. The evolution of computational power and the growing spectrum of algorithms have progressively allowed the assessment of more complex systems of relevance in different fields, from organic chemistry to materials science to biochemistry. In the following paragraphs, we are going to focus on three main problems: *(i)* accurate quantum description of the system, *(ii)* inclusion of appropriate environment conditions, like temperature, solvent, pH, minerals, etc.; and *(iii)* statistically converged sampling of free energy landscapes and kinetic rates.

2.1. Quantum-chemistry techniques for gas-phase systems at zero temperature

Ab initio quantum chemistry methods are very accurate in computing minima (corresponding to reactants, intermediates, and products) and barriers in the energy landscape of a quantum system containing many atoms and electrons, without explicit temperature effects. *Ab initio* is a latin term that means “from the beginning” or “from first principles”. *Ab initio* quantum chemistry is indeed an approach based on quantum mechanics, not relying on experimental information. This type of methods has in general a elevated computational cost, and typically one is able to model relatively small systems (up to few tens of atoms); however, this is the most established tool to study at the atomic scale the properties and reactivity of molecular systems.

QC calculations aim to the “exact” solution of the Schrödinger equation corresponding to the molecular Hamiltonian, or in the case of the quantum relativistic formulation, of the four-components Dirac equation [34], normally applied to highly massive nuclei such as the lanthanides and/or actinides. Remaining in the non-relativistic limit, solving the Schrödinger equation provides the electronic structure of the system together with the corresponding energetic information. Therefore, it allows to model the rupture and formation of chemical bonds, *i.e.*, it provides insight about chemical reactions. In general, an analytic solution to the Schrödinger equation is not possible for a system bigger than the hydrogen atom. Nevertheless, in the last decades many methods have been developed with the aim of obtaining an accurate, albeit approximate, numerical solution of the Schrödinger equation for systems including tens to hundreds of atoms. Among them, the Hartree–Fock (HF) method was one of the first quantum chemistry approaches to provide useful insight. In the HF equations, the electronic wave-function is approximated as a Slater determinant made up of one spin orbital per electron, where the orbitals are optimized by a numerical iterative procedure (the so-called self-consistent-field, SCF). However, its formulation inaccurately describes states in which electron correlation plays an important role. In this regard, one of the main challenges has been to develop new approaches that include the electron correlation at a reasonable computational cost, typically known as post-Hartree–Fock methods.

In this class of approaches, there are two main categories which are based on *(i)* the variational principle or on *(ii)* perturbation theory. Among the post-HF variational methods, the so-called configuration interaction (CI) method [35] is one of the most general approaches that includes the electron correlation. CI is based on a linear combination of Slater determinants, where the coefficients are estimated through the diagonalization of the Hamiltonian matrix in the subspace of the N-electron determinants. These determinants are formed by "exciting" none, one, two or more electrons from occupied to virtual orbitals from the reference determinant $|\phi_0\rangle$.

$$|\Psi\rangle = c_0|\phi_0\rangle + \sum_{i}^{occ}\sum_{a}^{vir} c_i^a|\phi_i^a\rangle + \sum_{i<j}^{occ}\sum_{a<b}^{vir} c_{ij}^{ab}|\phi_{ij}^{ab}\rangle + \sum_{i<j<k}^{occ}\sum_{a<b<c}^{vir} c_{ijk}^{abc}|\phi_{ijk}^{abc}\rangle + ... \qquad (1)$$

The equation above represents the full CI wave function, that comprises all the possible excitations of a N-electron system. However, the implementation of full CI is only feasible for very small systems, since its computational demand is very elevated and unpractical. The most popular CI approximation is the one including all single and double substituted determinants (known as CISD) to the reference one $|\phi_0\rangle$. This truncated CI approximation is more feasible and is implemented in several codes. However, CISD is not size-extensive which makes it less popular since the accuracy of the calculation decays with the size of the system.

As an alternative to CI (and/or CISD), another class of methods at lower computational cost is available: the Møller–Plesset (MP) approach [20]. The MP method is based on the Rayleigh–Schrödinger perturbation theory, where the Hamiltionian of the system is written as the sum of a pertubation $\hat{V}$ and of an unperturbed Hamiltonian operator $\hat{H}_0$ as:

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \qquad (2)$$

where λ corresponds to an arbitrary parameter that controls the size of the perturbation, and in general is assumed to be small. In MP theory, the wave function and energy are then expanded as a power series in the λ parameter as:

$$\psi_\lambda = \psi_0 + \lambda\psi_1 + \lambda^2\psi_2 + ...$$
$$E_\lambda = E_0 + \lambda E_1 + \lambda^2 E_2 + ...$$

being ψ_0 and E_0 the HF wave function and energy. MPn is obtained by truncating the expansion at order λ^n.

Previous works have demonstrated that MP perturbation theory is not necessarily a convergent theory at higher orders. Its convergence can be slow, fast or simply very erratic depending on the chemical system and/or basis set that is used [36]. For instance, in some applications different molecular properties calculated at MP3 and MP4 level are no better than MP2 ones, even for small molecules. Additionally, the cost of calculating the MP2 energies scales as $O(N^5)$, and typically captures around 80-90% of the correlation energy, making this second-order approach the most popular for different applications in the literature.

Another post-HF method implemented in quantum chemistry softwares is the coupled-cluster (CC) theory. CC is based on an exponential ansatz for the wave function $|\psi_{CC}\rangle = e^{\hat{T}}|\phi_0\rangle$, where $|\phi_0\rangle$ is the HF wave function and $\hat{T}$ is the cluster operator:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + ... + \hat{T}_N$$

This exponential term is the sum of cluster operators of different excitation states, where a single excitation is described by the operator $\hat{T}_1$, the double excitations can be described in either two ways, by $\hat{T}_2$ that describes a simultaneous excitation of two electrons, or by $\hat{T}_1^2$ describing two independent single excitations, and so forth for N-order excitations. This makes CC one of the most expensive methods from the computational point of view, but at the same time one of the most accurate. For practical purposes, the most common CC implementation is based on the single, double, and triple excitations (CCSD(T)) [37] where the triple excitations are treated as a perturbative component. In general, CCSD(T) provides reference quantum chemistry results, with an excellent typical accuracy compared to experiments (below 0.5 kcal/mol [38, 39, 40]), but only for small systems, and limited to solving the electronic problem at fixed geometry (optimized using some more approximate and cheaper method). For instance, in a system of N-electrons a CCSD(T) calculation scales as $O(N^7)$, which limits its implementation to small- and medium-sized molecules, and demands the use of parallel supercomputing platforms already for systems of about 10 to 15 heavy atoms [38].

Another important class of approaches to the electronic structure of a given system is provided by stochastic Quantum Monte Carlo (QMC) methods, long considered prohibitively costly from the computational point of view. They have been recently gaining significant grounds with respect to the most advanced quantum chemistry approaches, because of the more favorable scaling with the number of electronic degrees of freedom. Although historically developed for strongly correlated systems in solid state physics, QMC methods are more and more currently used for molecular systems, and combinations of QMC with molecular dynamics are nowadays carried out, at an accuracy comparable to CCSD(T) and with a far more favorable scaling and parallelization [41, 42, 43].

Over the decades, a less accurate then QC but more affordable class of computational methods for physics and chemistry, based on Density Functional Theory (DFT) [44, 45], have emerged and became extremely popular and widespread. DFT-methods are centered on the Hohenberg–Kohn theorem [46], which proves that the ground state energy of a multi-electron system is univocally determined by a functional of the ground state electronic density $E_{GS} = E[n_{GS}]$. The electronic density is a function of three spatial coordinates only, and not of the coordinates of all N electrons like in the case of the wave function, which makes its treatment more efficiently manageable. As a consequence, and thanks to the availability of efficient approximations and algorithms, DFT-techniques provide a powerful and fast method to get information about the structure and energy of a chemical system, at least in the electronic ground state. Within the DFT framework, the many-body problem of interacting electrons in a static external potential is treated as non-interacting electrons moving in an effective potential. This effective potential includes the external potential and the effects, both of classical-electrostatic nature and of quantum one, of the Coulomb interactions between the electrons (Hartree and exchange-correlation). Indeed, in the DFT formulation, the energy of a system of electronic density $n(\mathbf{r})$ can be thus expressed as:

$$E[n] = T_s[n] + \int d^3r \ V_{ext}(\mathbf{r})n(\mathbf{r}) + E_H[n] + E_{xc}[n] \tag{3}$$

where $T_s[n]$ is the kinetic energy of a non-interacting electron gas of density n, V_{ext} is the external potential acting on the electrons (typically the Coulomb interaction with the atomic nuclei), $E_H[n]$ is the classical electrostatic interaction between electron densities, and $E_{XC}[n]$ contains all the (unknown) quantum part of the electron-electron interactions, most notably the exchange (X) and the quantum correlation (C). The latter term is one of the main issues in DFT approaches; for this reason, different approximations have been tested in small/medium and large-scale applications, with the aim of describing the molecular properties with high accuracy. So far, there are three principal classes of functionals that have been developed to tackle this issue: *(i)* local density approximation (LDA) [47, 48, 49], *(ii)* generalized gradient approximation (GGA) [50, 51, 52, 53, 54], and *(iii)* hybrid functionals [55, 56, 57]. LDA treats the exchange–correlation energy functional, $E_{xc}[n]$, as a simple function of the electronic density $E_{xc}(n)$, and the corresponding energy terms are obtained from the exact result of the homogeneous electron gas having the same (local) density of the real system. Despite the crudity of this approximation, it works surprisingly well for a large variety of systems, and its typical drawbacks are well known and under control. GGAs or hybrid functionals are of course more sophisticated, but they are generally developed starting from LDA.

The great power of DFT relies on the effectiveness by which, once those approximations on the XC functional are introduced, the formidable quantum many-body Hamiltonian can be translated into much simpler single-particle eigenvalue equations known as Kohn-Sham equations:

$$\left[-\frac{\hbar^2\nabla^2}{2m} + V_s(\mathbf{r})\right]\psi_j(\mathbf{r}) = \epsilon_j\psi_j(\mathbf{r}) \tag{4}$$

where the Kohn-Sham effective potential is derived from the energy functional:

$$V_s(\mathbf{r}) = V_{ext}(\mathbf{r}) + e^2\int d^3r\frac{n(\mathbf{r'})}{|\mathbf{r}-\mathbf{r'}|} + V_{xc}(\mathbf{r};[n]) \tag{5}$$

and with the constraints that those single-particle wave functions must be normalized to the total electronic density $n(\mathbf{r})$:

$$n(\mathbf{r}) = \sum_j f_j|\psi_j(\mathbf{r})|^2 \tag{6}$$

where f_i are electronic states occupation numbers.

Further developments brought the second type of functionals based on the gradient-approximation. GGA-functionals are based on the LDA but include a gradient term of the electron density, which contributes to account for deviations from uniformity. The third level of complexity in DFT-based approaches includes the hybrid functionals which also incorporate gradient corrections but include a certain amount of exact Hartree–Fock exchange [58]. More recent advances in developing more accurate functionals gave rise to meta-GGA methods [59], which are essentially an extension of the GGA where the non-interacting kinetic energy density is used as input to the functional as well as the electron density and its gradient. Among the most popular meta-GGA approaches, Minnesota density functionals developed by Truhlar and collaborators have been widely implemented and have shown a good suitability to different chemical applications [60]. Other important developments of special relevance in molecular systems, are the corrections of van der Waals forces. Among them, works by Grimme [61, 62] (DFT-D2) and Tkatchenko-Scheffler [63] (DFT-TS) have proposed an additional energy term that accounts for the dispersive interatomic interactions. Several works by Scuseria and co-workers also developed range-separated functionals such as the screened functional (*e.g* HSE) [64, 65] and long-range-corrected functionals (*e.g.* LC-ωPBE) [66] where the exchange component is decomposed into short-range and long-range parts. In particular, screened functionals decompose the Hartree-Fock exchange at small interelectronic separation (r_{ab}) and the local exchange at large one. While long-range-corrected functionals separates Hartree-Fock exchange at large r_{ab} and local exchange at small r_{ab}.

In the last decades, DFT has became more and more popular throughout computational chemistry, as its cost is generally much lower than the quantum chemistry methods mentioned above, and its accuracy for many disparate applications has been quite satisfactory [67]. In particular, DFT still remains the appropriate electronic structure framework capable of treating explicitly not only relatively small molecules (reactants/products), but also their (usually complex) chemical environment, potentially including a solvent, a mineral or catalyst surface, the presence of impurities, and so on. Moreover, a realistic treatment of chemical reactions, particularly in solution, must generally include temperature effects, in order to achieve a proper description of the thermodynamics and kinetics of those reactions. Once again, the computational efficiency of DFT provides an excellent framework for that, as shown in the following section.

2.2. Molecular dynamics: explicitly including the solvent and temperature effects

In many chemical problems, especially in prebiotic chemistry, it is paramount to assess the effects of the environment. This might comprise the presence of solvent molecules, the effect of ions and/or pH, the interaction with mineral surfaces and so forth. In addition, for several applications is essential to take into direct account the effect of temperature, beyond normal-modes extrapolations like the quasi-harmonic approximation, and sometimes pressure. Hence, in several cases performing QC simulations at absolute zero temperature and gas-phase might not be a very suitable choice [68]. In this regard, several strategies have been proposed in order to improve the description of a system's environment. Among them, Tomasi and co-workers [69] about 30 years ago proposed a computational method known as polarizable continuum model (PCM) able to describe electrostatic solvation effects. PCM essentially computes the molecular free energy of solvation (G_{sol}) as the sum of three terms: the electrostatic interaction between solute-solvent (G_{es}), the dispersion and repulsion interactions solute/solvent (G_{dr}), and the cavitation energy (G_{cav}) that describes the work necessary to create a void cavity within the dielectric to host the solute. PCM models the solvent as a polarizable continuum, rather than individual explicit molecules, and it is an approach that can be easily coupled with QC simulations at low computational cost. This method however presents some limitations when non-electrostatic effects dominate the solute-solvent interactions [70]. For this reason, when solvent molecules are involved in a chemical process, their explicit inclusion in the modeling is crucial to provide a more accurate description [71, 72, 73].

Progress in computing power and the simultaneous developments of new theoretical methods as well as their *in silico* implementation open a new window to assess applications of higher complexity which years ago were intractable. Ever more, theory goes hand in hand with experiments providing both the interpretation and the prediction of chemical properties. In this direction, complicated systems like bulk solutions, solid/liquid interfaces, proteins, DNA and RNA polymers are gradually tackled by computational chemistry, for which different alternative methods have been developed. Among them, there are the QM/MM methods treating a relatively small region of the system at *ab initio* QC level (the QM region), while the rest of the system is typically treated by means of classical molecular mechanics (the MM region).

Molecular mechanics (MM) methods are based on classical laws of motion and empirical interatomic potentials, thus bypassing the explicit solution of the electronic problem and assuming valid the Born-Oppenheimer approximation [74] due to the different time scales for nuclear and electronic motions. MM comprises different types of simulation techniques, such as normal mode analysis [75], Monte-Carlo simulations [76], Langevin dynamics [77] and molecular dynamics (MD) [78]. In classical MD simulations electrons are not treated explicitly, and the atoms are modeled as inert point particles. This prevents in most cases the study of reactivity, since the formation and breaking of chemical bonds are not accounted for. Interactions between atoms are described by constructing a force field (FF), in a functional form that comprises parameters fitted from experiments and/or *ab initio* calculations, to account for the interaction between different types of atoms immersed in a specific chemical environment. The positions and momenta of all particles are calculated by solving Newton's equation of motion at discrete time steps (typically of the order of 1 femtosecond). This type of simulations can be applied to get both thermodynamic and kinetic properties, since they reproduce the thermal dynamics of a system according, e.g., to the Boltzmann distribution. Classical MD is computationally cheaper compared with QC methods, so that much larger systems can be assessed for a longer time scale (see also Figure 1). MD simulations are a widespread tool for the study of biological macromolecules, in applications like ligand binding to receptors, protein-protein interaction, and the folding of protein and nucleic acids polymers.

A compromise to simulate chemical reactions at a lower computational cost compared to a fully quantum mechanical descriptions is represented by methods that exploit an approximate form of the electronic Hamiltonian. A first example is the empirical valence bond (EVB) model, that was initially aimed to assess enzymatic chemical reactions in solution [22]. The EVB method is based on the valence bond of ionic-covalent resonance forms, and estimates a Hamiltonian for the isolated solute molecules as well as a Hamiltonian for the reaction in solution by adding the solvation energies to the diagonal matrix elements of the ionic resonance forms. More concisely, EVB calculates the Hamiltonian of reacting systems by resonance states as classical valence bond structures, which describe the reactant, intermediate(s), and product states of the reaction [79, 80, 81]. With this method, the energy of ground state can be obtained by diagonalizing the EVB Hamiltonian, and the associated activation free energies can be evaluated by adiabatically changing the system from one diabatic state to another [81].

A second example is represented by the density functional tight binding (DFTB) approach [82, 83, 84]. The corresponding Hamiltonian is derived within the framework of DFT, combining a second-order expansion of the Kohn-Sham energy about a reference electronic density with electrostatic interactions and an empirical short-range term. This approach is able to account for covalent bond formation/dissociation as well as charge transfer reactions, at a computational cost much smaller than DFT. In prebiotic chemistry, it was recently applied to the reconstruction of precise free-energy landscapes for glycine dimerization in explicit water solution [32], exploiting the cheap computational cost compared to DFT to sample extensively the free energy landscape as a function of general-purpose topological reaction coordinates [28] (discussed in Section 4).

Among the most powerful strategies to simulate complex chemical systems, a central place is held by *ab initio* molecular dynamics (AIMD) [85, 86, 87]. AIMD combines the quantum treatment of the electronic degrees of freedom, with the numerical integration of Newton's equation of motion, where quantum-calculated forces act on the atoms. From the computational point of view, AIMD methods are quite expensive; however, they provide a more accurate description of the electrons compared to the semi-empirical approaches described above, thus AIMD is well-suited to simulate the formation and rupture of chemical bonds in chemical reactions. In general, the typical time scale for individual AIMD trajectories ranges between tens and hundreds of picoseconds, with typical time steps in the range 0.1-0.5 fs, implying that the quantum-derived forces of the system must be calculated up to millions of times. It is important to note, however, that enhanced sampling techniques allow the efficient simulation of chemical reactions despite of the short *in silico* time scale (see Section 2.3). Since the computational cost grows with the number of electrons, AIMD is usually applied to relatively small systems (from tens to about one thousand atoms). The electronic problem is usually solved within the DFT framework, since more accurate QC methods are computationally very expensive for the above system sizes and time scale. Most often, plane-wave basis sets are used to represent the orbitals, and pseudopotentials avoid the explicit treatment of core electrons, chemically inert, thus significantly reducing the computational cost. In this framework, the interaction of valence electrons with core electrons and nuclei is replaced by an effective potential that correctly reproduces valence wave functions outside a small cutoff radius from the nucleus [88, 89]. In this way, no artifact appears in the interatomic interactions compared to all-electron calculations, as long as the cutoff radius is below the region where the actual valence electrons are involved in chemical bonding. Moreover, well-constructed pseudopotentials are transferable, *i.e.*, they can be employed in calculations of different chemical environments resulting in a comparable accuracy.

The straightforward implementation of AIMD in the DFT framework is generally referred to as Born-Oppenheimer molecular dynamics (BOMD) [85, 86], as atoms move on the Born-Oppenheimer potential energy surface under the action of forces that can be easily calculated through the Hellman-Feynman theorem. However, this implementation historically suffered from a serious drawback: MD is quite sensitive to the quality of forces. If the forces, *i.e.*, the derivatives of the energy, are not calculated with high accuracy, the MD simulation will quickly deviate from the "right" trajectory, which will induce a drift of quantities that should be conserved (like *e.g.* the energy) from their values. The error on DFT forces is linear in the self-consistency error of the charge density (while for the DFT energy it is quadratic). As a consequence, a very good and expensive convergence to self-consistency is required at every time step. In 1985 Car and Parrinello (CP) proposed a different approach [90]. They introduced an effective Lagrangian to treat both the electronic and ionic degrees of freedom:

$$L = \frac{\mu}{2}\sum_k \int d\mathbf{r}|\dot{\psi}_k(\mathbf{r})|^2 + \frac{1}{2}\sum_i M_i \dot{\mathbf{R}}_i^2 - E_{tot}(\{\mathbf{R}\},\{\psi\}) + \sum_{k,l}\Lambda_{k,l}(\int \psi_k^*(\mathbf{r})\psi_l(\mathbf{r})d\mathbf{r} - \delta_{kl}) \quad (7)$$

which generates the following, semiclassical, equations of motion for both the electronic and the ionic degrees of freedom:

$$\mu\ddot{\psi}_k = H\psi_k - \sum_l \Lambda_{k,l}\psi_l, \qquad M_i\ddot{\mathbf{R}}_i = -\frac{\partial E_{tot}}{\partial \mathbf{R}_i} \quad (8)$$

where M_i are the atomic masses, $\mathbf{R}_i$ are the atomic positions, ψ_k are the single-particle wave functions, $E_{tot}(\{\mathbf{R}\},\{\psi\})$ is the DFT total energy, μ is a fictitious "classical" electronic

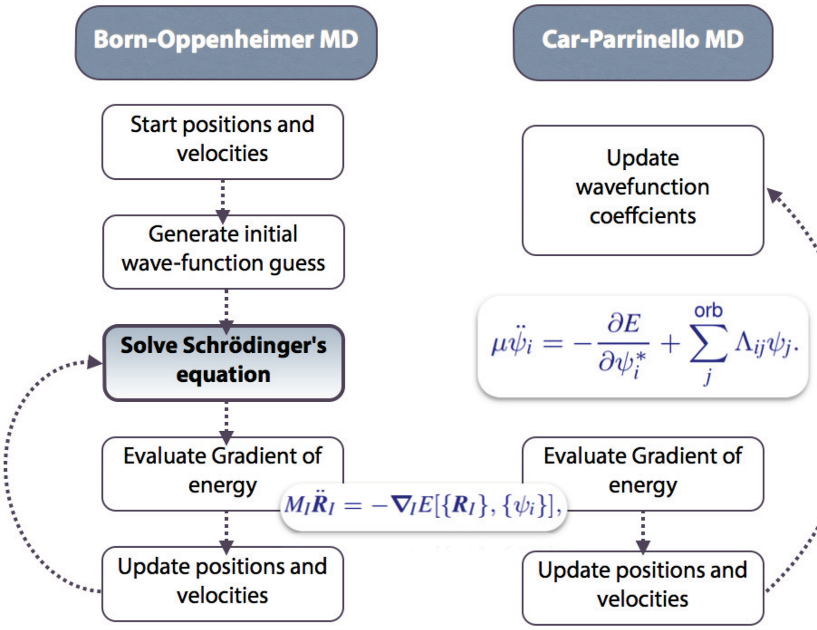

Figure 2: Simulation protocols for Born-Oppenheimer MD (BOMD) and Car-Parrinello MD (CP MD). Typically, BOMD and CP MD are performed within DFT framework. Schemes adapted from Wang's lectures [91]

mass, and $\Lambda_{k,l}$ are Lagrange multipliers needed to enforce the wave function orthonormality along the trajectories. The simulation starts by bringing the electrons on the "correct" BO surface (that is, to their ground state) at fixed ions and proceeds, using standard MD technology, on both electronic and ionic degrees of freedom. With appropriate values of μ and of the timestep δt, the electrons always remain close to the BO surface, while the ions follow a trajectory that is close to the one they would follow in the BOMD.

The important advantage of CP MD with respect to BOMD has historically been its much more favorable scaling. As a consequence, for a couple of decades CP MD has been the only realistic possibility to perform AIMD on interesting systems, particularly chemical reactions (a diagram of both protocols is depicted in Figure 2). Nowadays, the increasing computational power largely compensates the BOMD typical drawbacks. As a consequence, the vast majority of AIMD simulations now adopts the straightforward BOMD approach.

2.3. The time-scale problem: enhanced sampling of rare events

In principle, AIMD simulations are ideally suited to study the mechanisms, thermodynamics, and kinetics of chemical reactions at a given pressure and temperature. An extremely long trajectory of the system, generated from accurate forces, would ergodically sample fluctuations within metastable states as well as all possible transitions among them. However, DFT-based AIMD simulations nowadays are typically limited to the subnanosecond time scale. Of course, because of free energy barriers, the experimental time scale of most chemical reactions is immensely slower, ruling out straightforward computer simulation. To cope with time-scale limitations, but also to get fundamental insight that is not automatically provided by a bare long MD trajectory, a number of different so-called enhanced sampling approaches were introduced in the last few decades.

Two problems need to be addressed: accelerated exploration (i.e., the need to escape rapidly from local minima) and precise sampling (i.e., the extensive accumulation of samples in each relevant region of the configuration space, to estimate equilibrium thermodynamic and kinetic properties). Some enhanced sampling techniques are especially targeted to either of the issues, while others try addressing both at once. Depending on the field of application, three main strategies are employed: exploiting easier barrier-crossing at high temperature, adding suitably designed external biasing forces, or generating a large number of short unbiased reactive trajectories [92].

The techniques most relevant to the chemistry problem addressed in this review exploit the second approach, that can be schematically understood in the following framework. First, a generalized coordinate s tracking the progress of a reaction is selected. In general, s is a function of the coordinates of all atoms of the system (including the solute, solvent, interface minerals, and so on). At given N, V, T conditions in the canonical ensemble, within an irrelevant additive constant the logarithm of the equilibrium probability distribution $P(s)$ of the reaction coordinate corresponds to the (Helmoltz) free energy profile $F(s)$ of the reaction, often assuming the familiar shape of a double-well curve:

$$F(s) = -k_B T \log P(s) = -k_B T \log \frac{\int d\mathbf{R} e^{-U(\mathbf{R})/k_B T} \delta(s - s(\mathbf{R}))}{\int d\mathbf{R} e^{-U(\mathbf{R})/k_B T}} \tag{9}$$

where $\mathbf{R}$ collectively indicates the set of all atomic coordinates, the Hamiltonian is assumed separable with respect to coordinates (potential energy U) and momenta (kinetic energy), and the delta function selects the region of configuration space corresponding to the value s of the reaction coordinate. One can determine the probability ratio between, e.g., reactants and products (equilibrium constant) by integrating in the corresponding regions of configuration space, and the result translates into a free-energy difference by virtue of the previous formula. As usual, free-energy differences can also be decomposed into the sum of an energetic and an entropic component (note that N, P, T conditions can be treated along similar lines, however to date most MD simulations of chemical reactions are performed at fixed volume).

In several enhanced sampling strategies, an external bias potential $V_B(s)$ is added to the Hamiltonian of the system in the course of the simulation, resulting in biasing forces on the atoms in addition to natural forces, so that the total force on atom I can be written as

$$\mathbf{F}_I = -\frac{\partial U}{\partial \mathbf{R}_I} - \frac{\partial V_B}{\partial \mathbf{R}_I} = -\frac{\partial U}{\partial \mathbf{R}_I} - \frac{\partial V_B}{\partial s}\frac{\partial s}{\partial \mathbf{R}_I} \tag{10}$$

By suitably designing the bias potential, it is possible to enhance the sampling of selected regions of the free-energy landscape: for instance, in the limiting case $F(s) + V_B(s) \approx 0$, the landscape would be flattened, barriers hampering reactions would disappear, and all s values would become equally likely.

Of course, in practice $F(s)$ is unknown at the beginning of the simulation, so it cannot be explicitly employed to define $V_B(s)$. Nevertheless, different ingenious recipes have been devised to construct a bias potential: some algorithms exploit a constant bias $V_B(s)$, others a time-dependent $V_B(s,t)$: analysing the results of the latter is typically more cumbersome. The first class includes umbrella sampling [23], hyperdynamics [93], accelerated MD [94], boxed MD [95], and more. The second class includes local elevation [96], conformational flooding [97], metadynamics [24], self-healing umbrella sampling [98], adaptive biasing force [99, 100], targeted MD [101], steered MD [102], adiabatic bias MD [103, 104], and more. In favorable cases, the equilibrium free-energy profile can be recovered by analyzing the simulations, despite the fact that they were performed in presence of an external bias. However, it is important to stress that the efficiency of all these techniques depends critically upon the choice of the collective variable s (see below for a discussion).

Probably the two most widespread enhanced sampling techniques for chemical reactions are umbrella sampling and metadynamics. In umbrella sampling, a well-defined function $V_B(s)$ is chosen before the start, aimed to focus the sampling in a predefined region of configuration space. A common example is a harmonic potential restraining an inter-atomic distance within the neighborhood of a fixed value. Repeating similar simulations for different distances allows, e.g., to sample thoroughly the breaking of a chemical bond. Eventually, the biased probability distributions of all the simulations are corrected mathematically to remove the bias and combined together in a free energy profile using some variant of the weighted histogram analysis method [105, 106, 107, 108, 109, 110]. Umbrella sampling is a rather robust approach, widely used in different fields: it has the advantage that additional simulation can be easily added where needed, systematically improving the statistics, and that windows can be run simultaneously in parallel. On the other hand, applications are mostly limited to 1D or 2D free energy landscapes, the initial $V_B(s)$ often needs optimization to enforce overlap of s distributions between windows, and atomic trajectories must be carefully inspected – a non-trivial task – to assess convergence.

In metadynamics, the form of the bias is not guessed in advance, rather $V_B(s,t)$ is constructed on-the-fly as a sum of Gaussian contributions ("hills") added along the trajectory of the collective variable $s(t)$ at regular intervals.[24, 25, 111] The Gaussians, small both in height and width compared to the local minima, tend to fill the latter bringing the system to overcome large barriers in a short simulation time. The basic rationale is that the system tends to spend time in low free energy regions (since by definition they are more probable), thus the regular addition of Gaussians tends to automatically fill all local minima. Metadynamics enjoyed a broad diffusion thanks to the simplicity of the algorithm and to the considerable amount of work that clarified theoretical foundations and convergence properties [112, 113, 114, 115, 116], with several effective variants introduced. Besides accelerating rare transitions, metadynamics also allows, in favorable cases, to reconstruct the equilibrium free energy landscape from $V_B(s,t)$, like a cast shaped from a mould [112, 113, 114, 115, 116].

As stressed above, a key ingredient in the study of activated processes is the reaction coordinate. In a reaction connecting two metastable states A and B, it can be defined as the committor function $p_B(\mathbf{R})$ that associates each configuration $\mathbf{R}$ of all the atoms with the probability to reach B before A [117, 26, 118, 119, 120, 121]. Such probability can be computed launching from a given configuration a large number of MD trajectories, each with different initial atomic velocities drawn from the equilibrium distribution, and counting how many reach B before reaching A (typically, less than 1 ps is necessary to relax to either state). The committor function varies smoothly between zero and one, assuming a value of 0.5 at the transition state. Therefore it is a suitable progress indicator for the transformation, including kinetic information about the fate of configurations, and is a good definition of reaction coordinate. The applicability is general, going from $T > 0$ down to $T = 0$ (where it reduces to the so-called intrinsic reaction coordinate [122, 123]), encompassing barriers of energetic and/or entropic origin, purely diffusive processes, etc. In practice, the committor can be estimated for a limited number of configurations along a reactive pathway, sampled with the help of metadynamics, umbrella sampling or any other technique. The procedure allows to identify transition state structures, and to obtain unbiased reactive trajectories for further analyses of the reaction mechanism.

Unfortunately, since the committor function $p_B(\mathbf{R})$ cannot be operatively estimated for all possible configurations, the search for approximate definitions of the ideal reaction coordinate is a key challenge of computational chemistry, especially in solutions. The resulting difficulties traditionally limited the widespread diffusion of MD-based studies of reactions. Section 4 describes a recent simple, intuitive, and general formulation of reaction coordinates [28]. This new approach, together with modern computers that allow to simulate hundreds of atoms on the nanosecond time scale, is able to open the way to a systematic reconstruction of chemical reaction mechanisms and free-energy landscapes in solution, as demonstrated for several case studies in prebiotic chemistry.

3. Computer simulations emerging as a tool for prebiotic chemistry

As discussed in the previous chapter, sophisticated computational *ab initio* methods, based on quantum physics, are commonly used for general chemistry problems. However, their application in the field of prebiotic chemistry has not been widespread, likely because the chemical complexity of prebiotic scenarii and the huge uncertainties on the primordial environmental conditions have long been discouraging.

In recent years, the increase in computational power and the development of advanced enhanced sampling methods, capable to efficiently explore prebiotic chemical networks, fostered the application of AIMD simulations to prebiotic systems. In fact, it has become evident that it is feasible to model chemical reactions in many different relevant environments, such as bulk solutions, at the interface with mineral surfaces, at various pressure/temperature hydrothermal conditions, and so forth. This intrinsic versatility have recently made *ab initio* calculations, and particularly AIMD, highly attractive in the PCOL research field.

In the following of this chapter, we will distinguish several different prebiotic contexts (as illustrated in Figure 3), which are typically within the realm of a specific subset of the *ab initio* approaches previously described.

3.1. Interstellar medium

The space between different star-systems, also known as the interstellar medium (ISM), represents a precious source of matter and radiation. In ISM, different temperatures and densities of chemical material are found, and the latter experiences several chemical processes. This is of great interest in prebiotic chemistry, since the chemical reactions occurred in ISM can represent a significant exogenous source of organic material on Earth. Thanks to advances in spectroscopy, many different types of molecules have been detected in the ISM [124]. However, temperatures in the medium are typically below 100 K, which implies a very slow kinetics for the collision and combination between molecules. For this reason,

Highlights

- Life Origin is a pluridisciplinary field where computer simulations are essential.
- Endogenous and exogenous environments played a role in emergence of life on Earth.
- The range of possible conditions requires different theoretical approaches.
- Predictive modeling of prebiotic environments is nowadays possible.
- Progress in methods open new perspectives of simulations in prebiotic chemistry.

Figure 3: Pictorial representation of some exogenous and endogenous prebiotic chemical sources. Exogenous material is typically synthesized in the interstellar medium and carried by different interstellar bodies such as comets and asteroids that subsequently impact the planet crust. Endogenous prebiotic material on the Earth's surface, can be formed under a reductive atmosphere (Miller synthesis [2]), from volcanic activity, different hydrothermal sources, and so forth.

the exploration of plausible chemical pathways in ISM that can support the formation of the observed species has become an issue of increasing interest. In the vast and varied ISM, different conditions have to be considered, since some reactions occur in gas phase, while other are mediated by different types of solid surfaces. A possible source of energy is represented also by proton irradiation events [125, 126, 127]: however, for this kind of phenomena atomistic simulations are still lacking.

High-energy/low-temperature gas-phase reactions represent a first, fundamental class of synthesis of organic molecules of great prebiotic interest. From the computational point of view, they are typically very well described by accurate QC methods [128], capable to provide a rationalization at a precise electronic level.

Electronic structure and kinetics calculations performed by Barone and collaborators addressed formamide synthesis starting from NH_2 + H_2CO at B2PLYPD3 and CBS-QB3 levels [129]. From this study, it was found that this reaction is barrierless and it is very likely to happen under low temperature conditions typical from ISM objects. Moreover, the unveiled reaction ($NH_2 + H_2CO \rightarrow NH_2CHO + H$) does not require grain-surface chemistry, being possible to occur under gas-phase conditions.

Other works by Spezia and collaborators, employing *ab initio* simulations have also addressed the formation of formamide in the ISM exploring different ion–molecule reactions in gas phase [130]. More recent works, following a similar framework addressed the formation of protonated glycine by ion–molecule collisions in gas-phase, through the reaction between acetic acid and NH_3OH^+ or $NH_2OH_2^+$ (as suggested by previous experimental essays [131]), when sufficient translational kinetic energy is present. These AIMD simulations revealed the role of collision energy to yield the products of the reaction and its relevance in the ISM chemistry [132].

A step beyond pure gas-phase chemistry, dust grain surfaces represent a very attractive scenario for the concentration and protection of organic material subsequently distributed on the early Earth. In addition, icy-surfaces (such as the cometary ones) are also of great relevance since they can trigger a specific reactivity due to the presence of water molecules.

In this direction, some works have been focused on the formation of amino acids, saturated hydrocarbons, as well as other organic precursors like formamide. Previous studies have explored the synthesis of glycine on icy particles combining NH_3, formaldehyde and HCN [29] by means of DFT calculations (hybrid B3LYP) following Strecker-type pathways or through radical surfaces of water-ice dust particles [133]. These studies have shown the initial formation of NH_2CH_2CN compound, which is stored in the icy dust particles. Once delivered on the early Earth surface, this product could have led to glycine formation upon contact with the Hadean ocean. Other DFT-based works have studied the role of interstellar silica surfaces to understand the exogenous delivery of prebiotic compounds [134]. In particular, the interaction of glycine molecules on SiO_2 surfaces drives the formation of chemical complexes, with high stability, that could act as a protective agent for the substrates, upon delivery on the planet surface.

Other efforts in ISM chemistry focused on the study of tunneling effects, which are known to happen at very low temperatures. The inclusion of this type of effects proves relevant to explain how certain chemical reactions can take place in such special conditions. QM/MM simulations showed the hydrogenation of isocyanic acid (HNCO) on interstellar icy-surfaces to produce formamide – a crucial prebiotic precursor [135]. This type of calculations implemented tunneling rate calculations based on the instanton theory [136], where reaction rates are determined by the partition function of a quantum mechanical ensemble expressed by a Feynman path integral.

3.2. Photochemical reactions

Photochemistry has a special interest in prebiotic chemistry, since different important reactions could have been driven by the interaction of light (electromagnetic radiation) with the initial prebiotic substrates. An important feature of photochemical processes is their potential to yield products not easily accessible through thermal activation. Electromagnetic radiation have also the potential to induce chemical reactions at very low temperatures, like the ones found in the ISM (including bodies like comets and asteroids), leading to the possible formation of prebiotic exogenous material. Different studies during the last decades addressed all these questions.

When the Sun was a very young star and it was not as bright as it is now, the primitive Earth had a reductive atmosphere (lacking oxygen and ozone) that allowed the impact on the planet's surface of short wavelength light, photochemically very effective [137]. Moreover, it has been recently observed that homochirality in the development of living forms might be linked to some photochemical processes, based on the observation of a direct effect of circularly polarized light on selective enantiomer degradation [138]. Studying the photochemical response of different prebiotic products, as well as their photostability, can bring new explanations of the chemical selectivity of biological building blocks in the development of more complex organisms.

Highly-correlated QC simulations have been performed on photochemical reactions with the aim of providing molecular details from experimental assays. In general, second order perturbation theory correction (*e.g.* CASPT2 [139]) and multireference configuration interaction (MRCI) [140] have satisfactory reproduced excited-state potential energy surfaces, and state crossings. However, the multireference and multiconfigurational approaches mentioned above are highly demanding in computational resources. In alternative, coupled-cluster based approaches like CC2 and ADC(2) [141], usually very robust, have been applied to study excited states, also obtaining very reliable results. Another widespread option is represented by time-dependent DFT (TDDFT) simulations, which are computationally cheaper compared with coupled-cluster methods at the price of a lower accuracy, thus requiring a careful choice of the functional.

In the late 60's, Ferris and Orgel performed the synthesis of purine nucleobases starting from a HCN tetramer known as *cis*-diaminomaleonitrile (*cis*-DAMN). This molecule, after a photoisomerization process to the *trans* form is subsequently converted to 4-aminoimidazole-5-carbonitrile (AICN) [142]. Once AICN is formed, it reacts with HCN, C_2H_2 or NCO^- to produce purine nucleobases (as depicted in Figure 4), such as adenine, isoguanine and 2,6-diaminopurine [143]. Moreover, AICN in presence of water produces AICA, another purine precursor as shown is Figure 4. Previous works by Boulanger et al. [144] applied QC simulations to the photochemical conversion of DAMN to AICN. DFT and time-dependent DFT calculations indicate that the reaction evolves following a series of photoexcitations and internal rearrangements via an excited azetene specie. The most crucial stages of the reaction were optimized with high correlated calculations such as CC2 and CASPT2 second-order perturbation theory. From QM/MM simulations, the time scale associated to photoenergy dissipation was additionally determined. In general, the energy dissipation in water is a very fast phenomena (typically below 2 ps), preventing the possibility of hot-ground-state driven reactions.

The AICN molecule has significant interest in the context of the RNA world hypothesis, being a direct precursor of nucleobases present in RNA nucleosides and nucleotides. Moreover, although produced from a sequence of photoreactions, it is a stable molecule upon UV-light exposure. Previous works have shown solvent effects on the photochemistry of AICN exploiting second-order perturbation theory and multireference CI (MR-CISD) simulations [145]. In particular, the photochemistry of AICN–water clusters, including explicitly six water molecules, have shown that charge transfer to solvent follows the formation of an H_3O^+ cation, occurring in the excited state $\pi\sigma*$ hypersurface. The mechanistic features unveiled by these calculations suggest the possible photostability of analogous organic molecules in bulk water.

Previous works by Sutherland and collaborators [146] have proposed a synthetic pathway toward pyrimidine ribonucleotides, passing through 2-aminooxazole. Due to significant UV irradiation on primitive Earth, the study of the photostablity of this ribonucleotide precursor is essential. QC simulations at second order perturbation theory were performed by Szabla et al. [147], elucidating the possible nonradiative deactivation channels of this compound. Photochemical gas-phase processes of 2-aminooxazole should be dominated by the photo-detachment of the hydrogen atom of the NH_2 group, leading either to ultrafast

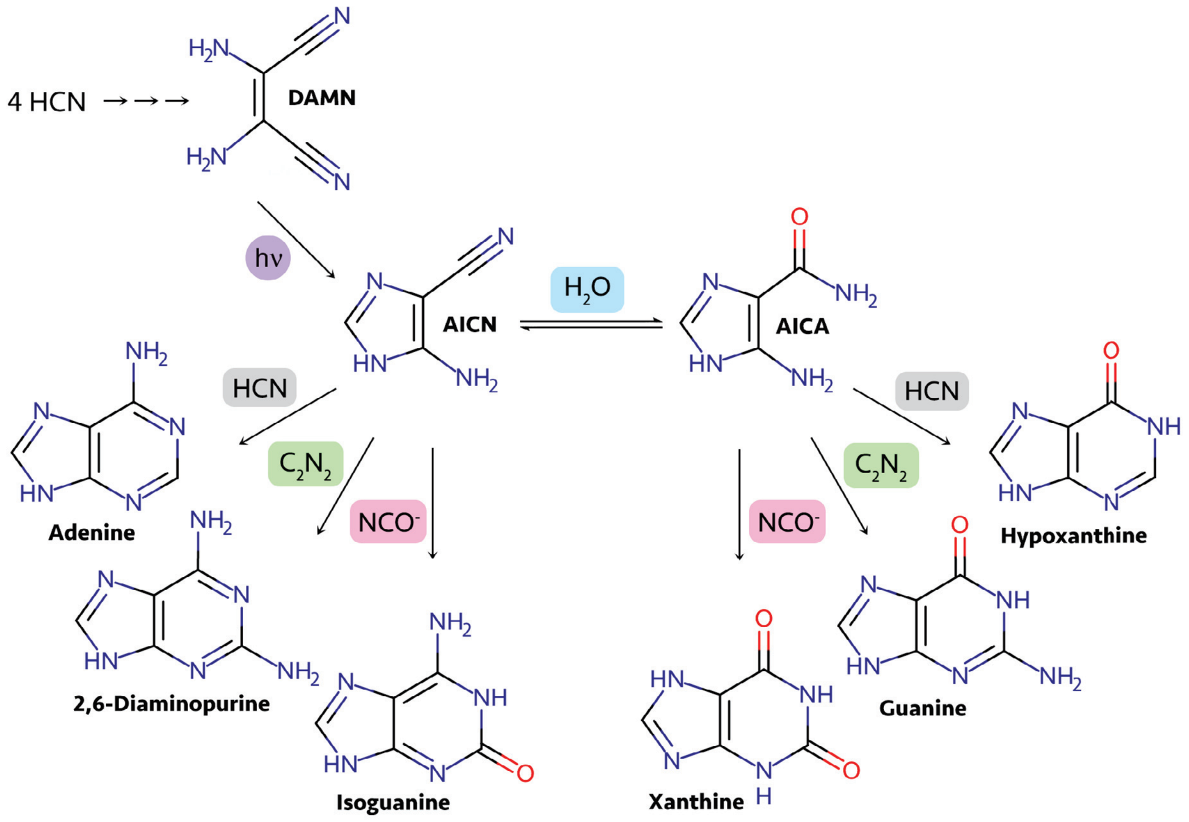

Figure 4: Photochemical reaction to form 4-aminoimidazole-5-carbonitrile (AICN), a purine precursor, starting from HCN tetramer, *cis*-diaminomaleonitrile (*cis*-DAMN) [142, 143].

nonradiative deactivation, phototautomerization, or photodissociation of a hydrogen atom. Ring-opening pathways were also unveiled by QC simulations showing a ring-puckering process that could occur after electronic excitation. However, this type of reactions are less likely since they involve a higher-lying excited singlet state and are slower than hydrogen-atom dynamics.

3.3. Mineral surfaces

A crucial question in prebiotic chemistry concerns the role, likely important, of mineral inorganic surfaces entering in contact with prebiotic substrates. In the late 40's John D. Bernal [148] pointed to mineral particles such as clays as a favorable environment for the synthesis of small polypeptides. Further works by Orgel and collaborators [149] observed the catalytic efficiency of hydroxyapatites in the peptide bond formation. Studies by Ferris and co-workers support the importance of clay minerals, such as montmorillonite, in prebiotic chemistry [15, 150]. This mineral has been one of the most widely studied due to its catalytic properties towards several prebiotic organic reactions. Montmorillonite is formed by the weathering of volcanic ashes, hence its relevance due to the presumable high volcanic activity in the primitive Earth. Furthermore, previous works by Smith and Parson have showed that silica-rich zeolites and feldspars (aluminosilicates of K, Na or Ca) might be responsible for the generation of replicating biopolymers due to the presence of acidic sites and/or hydroxylated species localized in molecular-sized cavities [151, 152, 153].

Mineral surfaces can provide an essential role from the catalytic point of view. First, they can favor the formation of biopolymers since they can selectively adsorb substrates from the solution and concentrate them. Moreover, the adsorbed organic material can be stabilized compared to the aqueous phase. Another important role of catalytic surfaces is the enhancement of oligomerization rates, since for certain types of monomers it might difficultly to spontaneously polymerize in bulk aqueous solution due to competition with the hydrolysis reaction.

An important challenge consists in rationalizing the polymerization of amino acids. Ugliengo and collaborators have explored the possible role of anorthite ($CaAl_2Si_2O_8$) and sanidine ($KAlSi_3O_8$), two of the most abundant feldspar minerals on Earth, as plausible catalysts for the formations of small peptides [154]. The importance of such mineral surfaces lays in their capability of providing Lewis and Brønsted sites that enhance the formation of complexes and catalyze peptide bond formation. *Ab initio* calculations (hybrid DFT, B3LYP) assessed the formation of glycine dipeptide. The final product remains in a complex with the catalytic surface in the Lewis acid site and can subsequently be attacked by another monomer (glycine molecule) in order to elongate the polypeptide chain, suggesting that the inorganic interface can play a role as a template for subsequent synthesis of biologically relevant enzymes.

Further works based on DFT simulations have addressed the adsorption of amino acids on silica surfaces to analyze a potential competitive effect between the formation of peptide bonds and their hydrolysis. Free-energies associated and the mechanistic details of both processes were elucidated, providing a quantitative insight and the effect of aluminosilicate surfaces in stabilizing oligopeptides [155]. From this framework, it was clearly observed that peptide bond formation is more likely with respect to its hydrolysis, suggesting the role of mineral surfaces as suitable promoters for biomolecular polymerisation.

Additionally, adsorption of RNA/DNA nucleobases on clay surfaces such as Na^+- montmorillonite has been studied by DFT (GGA functionals) calculations in absence of explicit solvent [156], with the aim of obtaining insights about intrinsic nucleobase and montmorillonite contacts. Works employing a similar approach investigated the interaction of DNA nucleobases on acidic montmorillonite [157], observing the different atomic configurations and the type of intermolecular interactions as a function of the nucleobase proximity with the surface. More recently, AIMD simulations by Mignon and Sodupe explored the structural details of an aqueous solution of cytosine inside a montmorillonite interlayer [158], showing that the nucleobase is stabilized on the surface by dispersion forces.

In the late 80's, Wächtershäuser introduced the so-called Iron Sulfur World (ISW) hypothesis [159, 13], claiming the important role of pyrites (NiS and FeS) as relevant prebiotic minerals. Pyrite-water interfaces are typically found in hydrothermal sources, such as the black smokers, and have the potential to catalyze amino acid polymerization. Further experimental assays in the late 90's showed the synthesis of acetic acid upon carbon fixation on FeS surfaces [160], as well as peptide formation under simulated hydrothermal conditions in presence of iron-nickel sulfides [161]. Following the same ISW setup, *ab initio* Car-Parrinello MD simulations coupled with metadynamics where performed to address glycine adsorption and desorption at water-pyrite interfaces [162, 163]. Energy barriers associated to the adsorption between the neutral and the zwitterionic forms of glycine show a significantly higher value for the neutral species. The reactivity of iron and sulfur atoms was observed to be higher in the defective regions (commonly found on naturally occurring pyrite) of the surface, displaying a more favorable interaction with glycine and stabilizing its adsorption at the pyrite-water interface under typical ISW thermodynamic conditions.

Further works by Schreiner and collaborators addressed peptide synthesis under plausible ISW conditions [164]. Car-Parrinello MD simulations in combination with free-energy methods modeled pyrite surfaces (ideal and defective structures) in presence of an aqueous solution of glycine. Following this computational approach, it is possible to estimate the free energy barriers for glycine dimerization at finite temperature and pressure at a fully solvated mineral interface, as well as to unveil the chemical mechanism and provide all atomistic details for the interactions. A very important feature extracted from the computer simulations was the observed catalytic effect of pyrite, since it significantly decreases the barriers associated to the dimerization reaction. Simulations adopting similar protocols further investigated pyrite-water interface properties under hydrothermal conditions in the context of the ISW [165], including catalytic effects on prebiotic ammonia synthesis from NO_3^- and NO_2^- precursors [166].

One of the most puzzling issues in the chemistry of RNA is linked to the instability of ribose and its difficult synthesis, making challenging its further accumulation in prebiotic scenarios [167]. In this regard, experimental works by Benner and collaborators demonstrated the effect of borates in the chemistry and stability of pentoses [168]. Further works by Cossetti et al addressed the role of borates in the stability of RNA molecules [169], due to their natural abundance on the Earth and prebiotic relevance in the RNA world context. From the theoretical point of view, QC studies by Šponer and collaborators investigated the structure and stability of aldopentoses (ribose, arabinose, lyxose and xylose) in complex with borates by DFT (B3LYP functional) and continuum solvent approaches [170]. It was observed that for the aldopentose-borate complexes, the electrostatic field of the borate is strong enough to change the orientation of the nearby hydroxyl groups compared to non-complexed aldopentoses. Moreover, it was shown that the ribose-borate complex is the most stable one, due to the strong hydrogen bond between the ribose 3-OH and one of the negatively charged borate oxygen atoms, and the favorable contact between the aqueous medium and the 5-CH_2OH group.

3.4. Bulk solutions and hydrothermal "extreme" conditions

Among the many relevant scenarios put forward to explain the emergence of life, the terrestrial ones constituted essentially by hot surface pools and deep-oceans hydrothermal sources are of crucial importance, as even space-delivered small organic molecules (up to amino acids and nucleobases) would have to assemble on Earth in order to form more and more complex (bio)chemical species [12]. From the computational point of view, the description of this class of prebiotic reactions implies the explicit inclusion of the solvent, and possibly of impurities and/or mineral surfaces. As discussed in previous sections, these conditions require a significant increase of the minimal number of atoms necessary to satisfactorily describe these chemical systems, and a direct inclusion of the temperature through MD. As a consequence, the corresponding computational cost increases dramatically, which explains why this major class of prebiotic problems has only been studied so far at the DFT level of electronic structure.

In this subsection we describe three classes of prebiotic reactions occurring when the environmental conditions are sufficiently "extreme" to provide the chemical energy necessary, with the addition of the thermal one, to overcome large reaction barriers. In the computational chemistry framework, this typically translates into reactions occurring "spontaneously", *i.e.* within the relatively short timescale of *ab initio* molecular dynamics trajectories, of the order of tens of *ps*.

The first class concerns the simulation of shock wave impacts possibly occurring on early Earth's surface pools. The study of shock waves has a special interest in prebiotic chemistry, since they can reach very extreme pressures and temperatures, triggering reactions that are typically thermodynamic- and kinetically difficult to overcome. On the primitive Earth, shock waves were very likely to happen and were normally generated by meteorite, comet or other icy-body impacts on the planet surface. During their impact, the material of those entities is highly compressed and display complex reactive pathways. Moreover, in many ISM icy-bodies, there is a significant presence of carbon species that react to produce different activated molecules upon shock wave impacts on the primitive Earth [171]. Previous works have addressed this phenomena, exploiting AIMD-based simulations. In particular, the modeling of shock-compressed interstellar icy-mixture have shown amino acid and hydrocarbon synthesis as well as the formation of organic complexes [172, 173, 171] (see also Figure 5). AIMD framework becomes very attractive to treat this kind of phenomena, since the common timescales associated with the rise time of shock waves are in the order of picoseconds [174, 175], and it is a powerful tool to unveil the chemical mechanisms upon show wave impacts. To address shock wave events, AIMD simulations are generally coupled with multiscale shock-compression simulation technique (MSST) [176, 177], able to reproduce with accuracy shock-wave profiles. MSST approach is based on atomistic MD and Navier–Stokes equations for compressible flow, allowing the description the hydrodynamic effects in the cometary impacts.

Different ranges of shock wave pressures and temperatures can be reached depending on the strength of impact. This implies a wide window of chemical conditions and as well as a varied reactivity. For instance, works by Goldman et al. [178] have shown how moderated impacts of CO_2-rich icy mixtures can drive the formation of nitrogen-heterocycles, that subsequently dissociate to form aromatic hydrocarbons after the expansion and cooling under ambient conditions. However, when considering those systems under higher shock wave conditions long carbon-chain molecules, CH_4 and formaldehyde are more likely to be formed, indicating the rich variety of chemical events under these extreme conditions. The extreme impact conditions, however, pose a important question about the stability and survival of the organic molecules, since they might follow undesirable degradation processes. In this regard, previous works have demonstrated that oblique impacts can potentially lower the shock wave temperatures and pressures, without affecting the synthesized prebiotic molecules. Additionally, since the impact events are happening at very narrow time scales, some organic material can survive the extreme temperature and pressure, allowing their successful further delivery on the Earth surface [178]. It is worth mentioning a study [179]

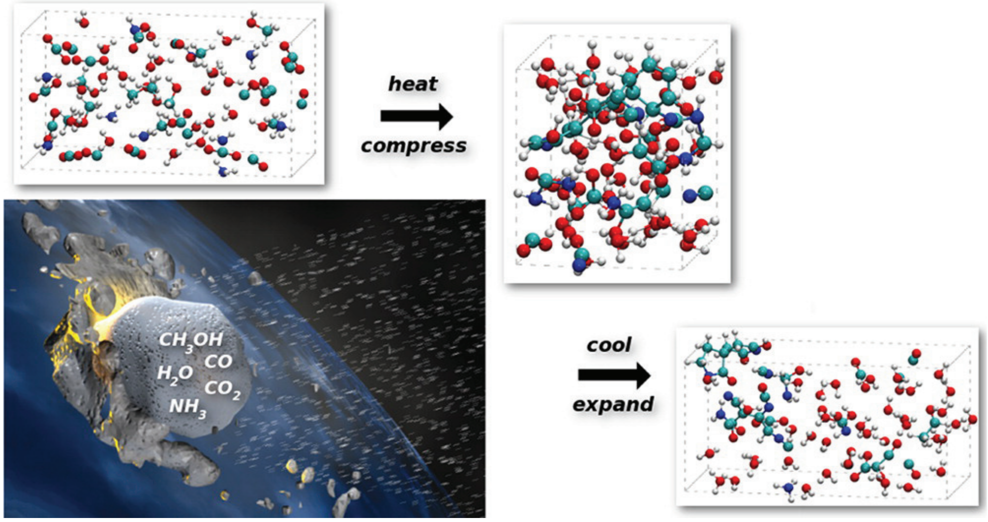

Figure 5: Scheme depicting shock compression simulations upon cometary impact on Earth. Shock wave involves an initial shock compression (due to impact on Earth crust), followed by a subsequent adiabatic expansion and the final cooling and equilibration of the chemical material. In this specific example, it is showed how shock wave impact triggers the synthesis of long carbon-chain molecules, CH_4, and formaldehyde, starting from very simple precursors from cometary bodies. Figure adapted from Goldman et al [178].

in which simple molecules, cold-compressed at extremely high pressures, become extremely reactive, yielding a large variety of chemical entities, including some of prebiotic interest. This approach, however, is more of a proof of principle of "virtual high-pressure prebiotic chemistry", as its conditions are not realistic and results are not quantitative.

As previously discussed in subsection 3.2, DAMN has been characterized as a key chemical precursor for nucleobase synthesis, however, when investigating the prebiotic chemistry of formamide ($HCONH_2$), similar conclusions can be drawn. Typically, formamide is obtained from the hydrolysis of HCN, and it is found in different plausible prebiotic scenarios, such as the ISM. Adenine can be obtained from formamide solutions upon heating, or by the reaction between formamidine and AICN [142], suggesting that the chemical channels towards nucleobases from DAMN and formamide are connected. In this regard, DFT-level simulations performed by Jeilani and co-workers have addressed the formation of nucleobases such as uracil, adenine and cytosine [180], studying radical reactions involving AICN (synthesized from DAMN, as mentioned in subsection 3.2), as well as the reaction between formamide and a radical amino-imidazole intermediate.

Further works by Ferus et al. investigated the formamide-based synthesis of nucleobases with a combination of experiments and QC calculations [181]. Formamide ice was exposed to laser-induced dielectric breakdown (LIDB) in the presence of FeNi meteorite material, modeling in this way the high energy impact of an extraterrestrial body on the early Earth atmosphere. The reaction mechanisms and free-energies were obtained by hybrid DFT (B3LYP) and CCSD(T) simulations. Vibrational wavenumbers at MP2 level were calculated to interpret experimentally measured spectra. The formation of 2-amino-2-hydroxy-acetonitrile (AHAN) and 2-amino-2-hydroxy-malononitrile (AHMN) was observed, and interpreted in terms of a reaction between formamide and CN· radicals, on a possible route to DAMN. Subsequent similar experiments and calculations indicated, remarkably, that liquid formamide exposed to high-power LIDB in absence of catalyzers yields the four RNA nucleobases adenine, guanine, cytosine, and uracil [182].

A second class of "extreme conditions" prebiotic chemistry is the one initiated by the celebrated Miller-Urey experiment in 1953 [2]. In this famous experiment, the author reported in fact the surprising results he had achieved by the application of an electric discharge on a simple mixture of gases that simulated the atmosphere of the primordial Earth, based on the Oparin-Haldane hypothesis. This way, he had observed the spontaneous formation of amino acids. By all means, this finding gave birth to a new multidisciplinary research field: prebiotic chemistry. Hence, the role played by electric fields in increasing the reactivity of atoms in matter had been already ascertained more than 60 years ago. However, the chemical reactions leading to the formation of amino acids, as those occurring in Miller's experiments, have never been studied at the atomic level before the advent of the extremely powerful computing techniques that have become available in more recent times. Again, *ab initio* methods have proven to be particularly suited for this purpose: in 2014 the first *in silico* Miller experiments have been reported [27], showing for the first time the active role of molecules such as formic acid and formamide as direct precursors of the simplest amino acid glycine.

A key aspect of the historical Miller experiments was the formation of hydrogen cyanide (HCN), aldehydes and ketones, and amino acids, suggesting that the experimentally-observed products were obtained through a Strecker-Cyanohydrin reaction [183].

In the theoretical *ab initio* study [27], the authors chose to decompose the study of the full Strecker chemical reactions into more elementary steps, including at each one of them the products of the previous one. On this basis, Saitta and Saija set up three corresponding simulation boxes. In particular, the end products of the reaction were glycine (NH_2CH_2COOH) and NH_3 in equal proportions, while the intermediate reactants were H_2O, NH_3, formaldehyde (H_2CO), and HCN, in equal proportions. The authors thus built up a starting Miller's sample box containing a mixture of molecules similar to the most recent experiments performed by Miller and coworkers [184], and bearing the same atomic species ratios as the two other boxes, that is 8 H_2O, 8 NH_3, 8 CH_4, 10 CO, and 5 N_2 molecules. Similarly, the intermediate Miller-Strecker box contained 9 H_2O, 9 NH_3, 9 H_2CO, and 9 HCN, while the end Miller-Strecker box contained 9 glycines, and 9 NH_3. This way, each box containing exactly the same number of atoms of the same chemical species, one could compare in a judicious manner the potential energies relative to the end products. In the absence of any field the only reactions were observed in the intermediate box, where formamide ($HCONH_2$) would occasionally form.

Once an electric field of about 0.35 V/Å was switched on the starting "Miller" box, the authors observed the formation of formic acid and formamide molecules, rather than formaldehyde and hydrogen cyanide ones, as suggested by Strecker mechanism. As a consequence, they chose to study a different chemical pathway to glycine synthesis, namely through those HCOOH and $HCONH_2$ precursors. A new "Miller" box was set, containing water, ammonia, carbon monoxide, formic acid and formamide, and put under intense fields of ~ 0.5 V/Å. In those conditions, the most interesting reactions occurred. In particular, all along the trajectory, the formation of formamide was observed through the combination of ammonia and carbon monoxide. Formamide then either formed larger and more complex molecules, or broke down into water and hydrogen cyanide, possibly after converting into its formimidic acid tautomer. Typically, a carbon monoxide then combined with a formamide molecule, a complex which dissociated into a hydrogen cyanide and a formate anion, which combined with a just-formed formamide-proton cation to yield a α-hydroxyglycine, as shown in the following reaction:

$$[\mathrm{NH_2CHOH}]^+ + \mathrm{COOH}^- \rightharpoonup \mathrm{NH_2}...{-}\mathrm{CHOH}...{-}\mathrm{COOH}. \tag{11}$$

evolving then into dehydroglycine ($NH...{=}CH...{-}COOH$), and finally, and successfully, glycine. One of the main aspects emerged from this study is the revelation of a chemical pathway leading to glycine which is different from the standard interpretation of Miller experiments. Moreover, it proved that the electric field acts as an order-maker, promoting the spontaneous assembling of simple Miller molecules into more complex ones of biochemical interest.

As a consequence, the potential role of the electric field in prebiotic chemistry has been studied in the more complex context of the synthesis of sugars. In fact, one of the crucial aspects related to the origins of life and thus to the onset of informational polymers on early Earth – such as RNA – is represented by the formation of the specific sugar composing this latter macromolecule (*i.e.*, ribose). However, the formation of the first C-C bonds from very simple molecules such as formaldehyde represents the rate-limiting step of the so-called formose reaction (*i.e.*, the autocatalytic reactions leading to the synthesis of sugars from aldehydes). Only very recently, the free-energy surface associated with such a process has been determined in condensed phase [185]. In fact, by means of AIMD and MetD techniques, the free-energy landscape underlying the synthesis of glycolaldehyde from a formaldehyde aqueous solution has been reported [185]. The latter work quantitatively explained why the very first step of the formose reaction is kinetically inhibited.

Moreover, in the same work [185], electric fields proved again to be dramatically important in the complexification of the investigated systems. In fact, numerical samples of formaldehyde (both neat and in water solution) and of glycolaldehyde (both neat and in aqueous solution) were exposed to intense electric fields. The application of electrostatic gradients strongly prevented the formaldehyde umpolung (*i.e.* polarity inversion) and catalyzed the formation of C-O-bonded polymers in formaldehyde-containing samples. Notwithstanding this latter finding, when the field was applied on glycolaldehyde aqueous solutions, new C-C bonds have been formed and (D)-erythrose has been synthesized. This way, a numerical Miller-like experiment led to the formation of a prebiotically relevant (D)-tetrose – *i.e.*, a direct precursor to ribose – from ubiquitarious molecules such as glycolaldehyde and water [185].

Other *ab initio* studies [186] investigated the thermodynamics of the synthesis of RNA nucleosides by DFT calculations in implicit solvent (conductor polarized continuum model CPCM), exploring three different routes: (*i*) direct combination of nucleobase and ribose, (*ii*) the multi-step synthesis proposed by Powner et al [146] through 2-amino-oxazole, and (*iii*) combination between nucleobase and RPP (rybosyl-pyrophosphate). This theoretical approach estimated the energetics of the different pathways and unveiled the potential role of phosphates in the formation of glycosidic bonds within nucleoside synthesis.

A third class of condensed-phase prebiotic chemistry within the reach of AIMD is constituted by reactions occurring in solution at the interface with a mineral, thus simulating both Earth surface and Earth depth geochemical phenomena. In this regard, conditions at the interface might be considered "extreme" because of the confinement and concentration of reactants, and the possible catalytic role of the surface. The most prominent work in this category is the study of the dimerization of glycine at the surface of prebiotically-relevant iron sulfide [164]. It is noteworthy that, besides this work, quite surprisingly this class of studies has not been significantly pursued, despite its important in prebiotic chemistry. A likely explanation is that those systems are still too large and computationally expensive to allow for accurate and quantitative *ab initio* studies. Within the broad category of wet interface chemistry, it must be noted that even apolar interfaces provide a strong electric field, which decays rapidly inside the solution, but whose impact on the chemical reactivity within the solution is still quite sizeable [187]. Prebiotic chemistry-oriented simulations in those conditions will be discussed in the next section, as they have been carried out within free-energy approaches.

4. New strategies for the direct comparison of thermodynamics and kinetics of reactive channels in prebiotic scenarios

Understanding the thermodynamics (free energy differences) and kinetics implied in reactions in solution is a crucial issue in prebiotic chemistry, as it is key to assess the likelihood of the different primordial scenarios put forward in the literature, and mentioned in the previous section.

However, our knowledge of reaction dynamics in condensed phases is far from complete [71, 188], despite the fact that water is a polyvalent molecule, known to participate to reactions under different roles, including as a stabilizer through hydrogen bonds, as an efficient acid-base bifunctional catalyst, and as a co-reactant [189]. Additional effects should also be considered, including vibrational energy dissipation upon birth of exothermic products, or solute trapping into finite-lifetime cages affecting its diffusion and reactivity [188, 190, 191]. The large number of possible configurations (already including few water molecules [189]) together with the strong anharmonicity of liquids naturally calls for methods like MD that include from the start the finite-temperature dynamics. Particularly in a prebiotic perspective, it is necessary a comparative understanding of reaction networks in different environments (gas or condensed phase, with different solvents and also interfaces with minerals) and at different conditions (T, P, irradiation, shock waves, etc.), eventually embracing also non-equilibrium scenarios, for their role in the emergence of life [192, 193].

So far, the widespread exploitation of MD simulations in the study of chemical reactions has been slowed down by the lack of general-purpose formulations of reaction coordinates. In particular, it is challenging to design coordinates that fully include the important role of the solvent degrees of freedom [117, 194, 195] and that are general enough to be applied to a range of diverse reaction mechanisms. Ref. [28] proposed a new way to tackle this problem. The method deals in the same formal way (namely, in the same space of coordinates) with gas phase and solutions, allowing a direct comparison of the two environments and an easier assessment of the effects of a given solvent on reaction pathways. The simulations described in the following typically exploit DFT-based Born-Oppenheimer MD simulations, with models including between a few (gas phase) and ~ 500 atoms (solutions and/or mineral surfaces, represented through periodically repeated supercells).

The basic idea is to combine together two successful tools. The first tool is represented by path collective variables s and z [196]: given the structure of the system at a given time $\mathbf{R}(t)$ and a set of n reference structures providing a (discretized) putative pathway between an initial $\mathbf{R}_1$ and a final $\mathbf{R}_n$ state, the variables are generally defined as

$$s(t) = \frac{\sum_{k=1}^{n} k e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_k)}}{\sum_{k'=1}^{n} e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_{k'})}} \tag{12}$$

$$z(t) = -\frac{1}{\lambda} \log \left(\sum_{k=1}^{n} e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_k)} \right) \tag{13}$$

with s quantifying the progress of the transformation and z the distance from the putative path. Both are important: s allows to direct biasing forces to accomplish the transformation, whereas z discriminates between different pathways and allows tracking transitions to states different from the target. The crucial ingredient is the metric D, as discussed below. λ is a parameter that can be conveniently set to be of the order of the inverse distance between neighboring reference configurations, assuming they are equally spaced. A much larger λ would produce very irregular and discontinuous pathways, while a much smaller one would hamper the resolution of different molecular structures. Customarily, path collective variables are employed to study conformational changes of biomolecules and chemical reactions in gas phase, using, e.g., the root mean square deviation (RMSD) of Cartesian positions as a distance metric between configurations.

Umbrella sampling simulations on methyl phosphate hydrolysis [197], as well as metadynamics simulations on CO_2 hydration [198] and on transformations of carbon nanostructures [199], clearly showed that path coordinates are a general and effective tool to reconstruct gas-phase *ab initio* free-energy landscapes even for concerted reaction mechanisms, a case where simpler coordinates easily fail. In the latter works, however, the simple RMSD metric was employed to define the path coordinates: this choice requires to identify beforehand the role that each individual atom will play in the reaction mechanism. Moreover, multiple reference structures (a putative path) were employed to define s and z. Clearly, such demanding prerequisites are suitable only for relatively simple gas-phase reactions with a mechanism (including intermediates) that is not completely unknown.

In general, it would be desirable to let the system "choose" which atom, among identical ones, will participate in a given reaction step (particularly, but not only, in solutions), and also which pathway is the most convenient one. As demonstrated by numerous applications, coordination numbers (formulated through Fermi-Dirac-like continuous functions of interatomic distances) are a powerful collective variable to describe chemical reactions, particularly in solution [200, 201, 194, 202, 164, 166]. However, customarily only one or two coordination numbers are employed as landscape coordinates in free energy calculations, to avoid the difficulty of sampling a high-dimensional space [92]. This is a severe limitation in general cases, where reactants, transition states and products display different values of several coordination numbers. The solution proposed in Ref. [28] consists in employing information about multiple (e.g., a dozen or more) coordination numbers within the following distance D between the atomic configuration at time t and the k^{th} path-CV reference structure:

$$D\left(\mathbf{R}(t), \mathbf{R}_k\right) = \sum_{IS} \left(C_{IS}(t) - C_{IS}^{k}\right)^2 \tag{14}$$

where C_{IS} is the coordination number between individual atom I of species (element) S' and all atoms J of species S, defined by a sum of smooth switching functions decaying from one to zero for increasing distance. In practice, the topology of the system is summarized into tables of *coordination patterns*, whose entries are the sum over sets of entries of the all-to-all adjacency matrix.

The metric takes into account variations in a complex pattern of many coordination numbers during a reaction mechanism (see Fig. 6), whereas its use within path coordinates reduces to only two the dimensionality of the free energy landscape to be reconstructed. As demonstrated by the following examples, this combination proved very effective for the simulation of different type of reactions, even in the case where only two reference patterns (the reactant and the putative product) are employed to define the coordinates. In this case, remarkably, the system is typically able to discover a suitable reaction pathway, including the possible appearance of locally-stable intermediate states.

The same definition of the metric D takes into account gas-phase situations as well as crowded environments like solutions and solids (or their combination, at a liquid/solid interface). As an example, for reactions involving the participation of one solute molecule in water, I in Eq. 14 runs over several individual atoms of the molecule (often excluding hydrogens, which can be exchanged rather easily), whereas solute and solvent atoms are contained in the groups S. Another convenient feature of the present formulation is that the detailed geometry of the reference structure needs not be explicitly given: instead, the matrix of C_{IS}^{k} entries can be directly constructed starting simply from structural formulas (Fig. 6). In practice, employing the average C_{IS}^{k} values observed in a short MD simulation

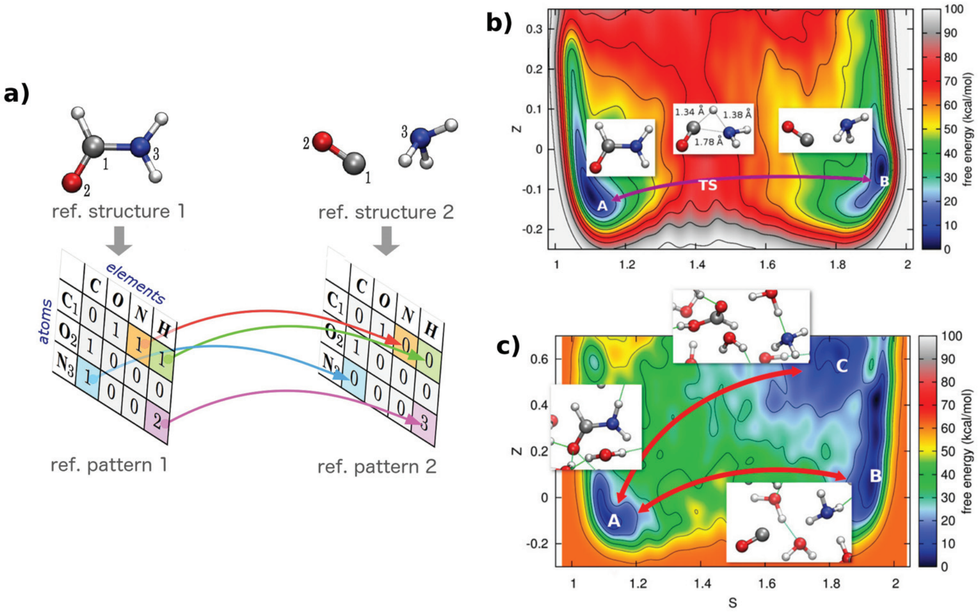

Figure 6: a) Construction principle of path collective variables based on coordination patterns. The connectivity patterns of reactants and products are represented by tables having individual, non-hydrogen atoms on rows and atomic species (the set of all atoms of a given element) on columns. Arrows indicate changes of coordination numbers, however all other matrix elements are free to change as well thanks to the flexibility of path collective variables, a useful feature for the discovery of transition mechanisms and unforeseen intermediate states. b) Gas-phase free-energy landscape for the formation of formamide from CO and NH_3, compared with b) the landscape in water solution [28].

of each end-state can improve the resolution of intermediate and transition states in the free energy landscape. The reason is that smoothly-defined coordination numbers deviate from perfect integers.

The coordinates can be employed in combination with efficient enhanced sampling techniques to explore reaction mechanisms and reconstruct free energy landscapes, and are implemented in a freely-available modified version of the plugin Plumed 1.3 [203], compatible with DFT codes such as CPMD, CP2K and Quantum Espresso. Newer versions Plumed 2.x [204] can be employed without modification, by exploiting suitable input structures. See Ref. [205] for code downloading and input examples.

5. Case studies: from small prebiotic precursors to amino acids

5.1. Small prebiotic precursors: formamide and formic acid

As laid out in Section 4, novel free energy calculation approaches, able to address in a general way a wide range of chemical reaction mechanisms in solution, are now available, allowing to unveil reaction networks of remarkable complexity. As an example, this has been applied to the study of the formamide decomposition channels in aqueous solutions, using DFT-based AIMD [28]. In this study, the authors investigated, in a first set of simulations, the emblematic decarbonylation reactions and their inverse in formamide:

$$\mathrm{HCONH_2} \rightleftharpoons \mathrm{CO} + \mathrm{NH_3}\,. \tag{15}$$

First, the gas-phase reaction was addressed. In a single trajectory of 170 ps, accelerated with metadynamics, the system performed 7 forward and 7 backward reactions despite the lack of educated guess about the transition state. The latter (identified by committor analysis, section 2.3), as well as the corresponding free energy barrier ($\Delta F^* = 79$ kcal/mol), were very similar to previously reported results obtained from higher-level QC approaches [206].

Compared to gas phase, the free-energy landscape in aqueous solution resulted significantly modified in both quantitative and qualitative features (Fig. 6, bottom panel). In particular, besides the reference states used to build the collective variables (formamide and CO+NH_3), the simulation explored a third basin featuring formic acid and ammonia, demonstrating at once the predictive power of the approach, able to freely discover unexpected pathways and products. The three states corresponding to aqueous formamide, CO+NH_3 and HCOOH+NH_3 showed a similar stability within the statistical uncertainty, with free-energy barriers of 35 kcal/mol for the reaction $HCONH_2 \rightarrow HCOOH+NH_3$ and of 40 kcal/mol for the reaction $HCONH_2 \rightarrow CO+NH_3$.

The fact that the two decomposition channels, decarbonylation and hydrolysis, have quite similar barriers, confirm the effectiveness of topological path collective variables in discovering competitive pathways. Additionally, ionic forms NH_4^+ and $HCOO^-$ were also explored, as expected from the basicity of ammonia and acidity of formic acid. An advantage of this MD approach is to have built-in anharmonic temperature effects, so that one can directly break down ΔF into energetic and entropic contributions: ΔE can be directly obtained as the average energy observed in a relatively long equilibrium (unbiased) MD simulation performed in each metastable state, whereas $T\Delta S = \Delta E - \Delta F$.

More in detail, all observed reactions began with water donating a proton to the amino group of formamide as the first step (leaving an OH^- species in solution), overcoming a barrier $\Delta F^* \approx 35$ kcal/mol. The resulting cation, that was identified as a relevant intermediate for this same reaction in Ref. [27], is locally stable only within a time scale of less than 1 ps, as verified with hundreds of unbiased trajectories started from 10 different configurations featuring this species. Experimentally, formamide is a weak base, with the protonated species having a standard free energy almost equivalent (only 0.1 kcal/mol higher) to the neutral species [207], however the carbonyl is expected to be the most probable protonation site [208], in agreement with the transient nature of the protonated nitrogen in these simulations. Due to the transient nature and very short lifetime of the $HCONH_3^+$ intermediate, the mechanism could also be approximately considered as one-step. From $HCONH_3^+$ the system can evolve either towards carbon monoxide and ammonia or towards formic acid and ammonia.

The new enhanced sampling approach was applied to the emblematic case of formamide, a centerpiece of many prebiotic scenarios recently put forward to explain the chemical origins of life on Earth, and observed in the *ab initio* simulation of Miller experiment under electric field [27]. On one hand, this method is capable of quantitatively reproducing existing gas-phase results. Much more importantly, liquid phase calculations of all three fundamental reaction channels, i.e. dehydration, decarbonylation, and hydrolysis, provided novel results of prebiotic significance and implications.

5.2. Decomposition and synthesis of amino acids in hydrothermal conditions

One of the central questions in prebiotic chemistry concerns the origin of amino acids, the building blocks of proteins. Assessing the likelihood of competing scenarios implies a quantitative understanding of the reaction network connecting amino acids with simpler molecules (synthesis and decomposition pathways), as well as of the effect of environmental parameters (temperature, pH, mineral surfaces, etc.) on the network itself. Moreover, to explain the observed abundances of different amino acids, amines, carboxylic acids and other organics found on meteorites (including the observed enantiomeric excesses) it is necessary to understand the synthetic relationship between all of these molecules [209].

Despite the Strecker mechanism being customarily invoked as a source of amino acids, starting from water solutions of aldehydes, ammonia and hydrogen cyanide [2, 183, 209], even in simple water solutions limited quantitative experimental and computational results available, hence a lack of satisfying understanding of the detailed reactions steps [210, 211, 212]. At the same time, alternative pathways based on different precursors have been also put forward [27, 213].

A recent investigation [215] addressed the simulation of amino acid decomposition mechanisms in hydrothermal solution, employing ab initio MD (for a total of more than 3 nanoseconds) and enhanced sampling. The models included an amino acid molecule solvated in 79 water molecules, at a temperature of 373 K compatible with the estimated conditions of aqueous alteration on parent bodies of chondrite meteorites [216]. Both glycine and isovaline were simulated: they are commonly found on carbonaceous chondrites, in the case of isovaline with L-enantiomeric excess [209]. Decarboxylation and deamination

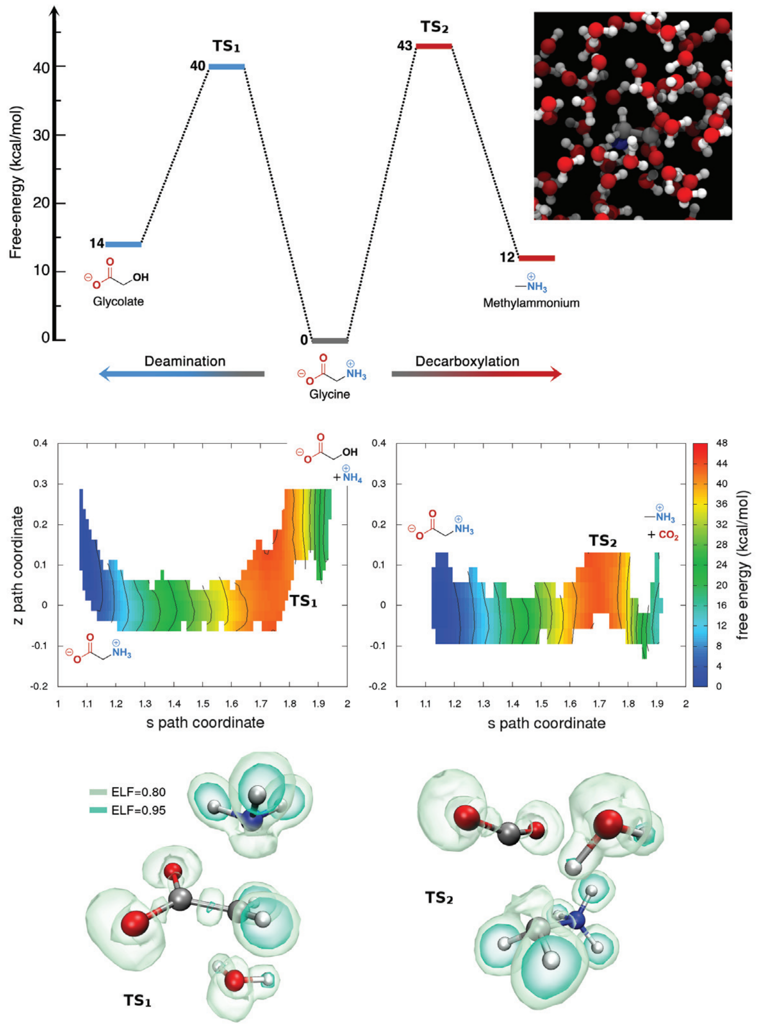

Figure 7: Glycine hydrothermal decomposition. (Top) Relative free-energy values for glycine decarboxylation and deamination. (Middle) Free-energy landscapes associated to the glycine decomposition on pathCV space. (Bottom) Electron locatization function (ELF) [214] analysis on transition state configurations for glycine decarboxylation and deamination reactions.

reactions have been simulated exploiting a robust protocol composed by i) metadynamics simulations for the discovery of reaction pathways without a priori assumptions, ii) committor analysis to extract transition state structures and unbiased transition trajectories, and iii) weighted histogram analysis of umbrella sampling simulations to reconstruct precise free energy landscapes [92, 215].

Being based on the general-purpose coordination pattern-based path variables of Ref. [28], the protocol could be applied without extensive fine-tuning to five different reactions: as shown in Fig. 7 in the case of glycine, detailed quantitative information could be obtained, including free-energy barriers in agreement with existing experimental data (a significantly improved result compared to a previous work based on less flexible reaction coordinates [217]). Transition states were characterized in terms of geometry as well as electronic structure, with deamination proceeding through a carbocation and deamination through a carbanion, as pinpointed by the electron localization function [214] (Fig. 7 bottom panel). From this point of view, the tools of conceptual density functional theory can play an important role in elucidating the reactivity and mechanisms of prebiotic systems [218].

The main findings of Ref. [28] consist in the rationalization of the contradictory literature on glycine decomposition [219], showing that deamination and decarboxylation are competitive channels whose prevalence can likely be inverted depending on experimental conditions, whereas isovaline deamination features a barrier higher than decarboxylation but leads to the most stable product, sec-butylamine. Note that, read in reverse order, decomposition channels could also suggest novel synthetic pathways alternative to Strecker. Mechanistic insight therefore compounds with thermodynamic and kinetic predictions, suggesting that this simulation protocol could represent a viable general computational tool highly complementary to experiments in prebiotic chemistry, most notably for reactions solutions.

5.3. Abiotic synthesis of ribonucleotides

The formation of RNA and its further accumulation on the primitive Earth is a crucial issue concerning the RNA World hypothesis. For this reason during the last years, many efforts have been focused to explore the RNA synthesis under different plausible prebiotic conditions, comprising the catalytic function of mineral surfaces, the exposure to drying/wetting cycles, the role of salts or lipid compounds, and so forth [146, 220, 221, 222]. In particular, one of the main challenges is to address the formation of ribonucleotides, that subsequently polymerize to constitute larger RNA oligomers with potential functionality [223].

In this context, a recent study assessed the formation of ribonucleotides under hydrothermal prebiotic conditions following a similar computational framework as discussed in sections 5.1 and 5.2: *(i)* exploratory ab initio metadynamics, followed by *(ii)* commitor analysis in the transition state region, and *(iii)* further umbrella sampling to reconstruct the free-energy profile of the reaction. In this study, purine or pyrimidine nucleobases are combined with a molecule of phosphorybosyl-pyrophosphate (PRPP), a metabolite involved in the current biological synthesis of nucleotides, under mild hydrothermal temperatures (400 K) and protonation states compatible with neutral or slightly alkaline pH [33].

As previously discussed in section 2.3, a key ingredient when performing enhanced sampling simulations is the choice of a set of generalized coordinates able to efficiently distinguish among two different metastable states. In the case of ribonucleotide synthesis, similarly to several other reactions in bulk solution, the implementation of path collective variables based on patterns of coordination numbers (as introduced in section 4) demonstrated to be an effective and versatile tool to track the changes of the chemical bonds.

The atomistic details provided by the ab initio MD simulations show a simple S_N2 mechanism that leads to the formation of both purine and pyrimidine nucleotides. Moreover, the comparison between this abiotic reaction mechanism with the enzymatic ones suggests that cellular nucleotide synthesis could have evolved through an optimization of the kinetics of a corresponding abiotic pathway [33]. This hypothesis addresses one of the major open questions in the RNA world scenario. Mass spectroscopy experiments on AMP are compatible with the theoretical pathway (despite low yields that could possibly be enhanced by mineral catalysts [224]).

6. Outlook

In the last decades, different experimental and theoretical strategies have been applied, with the principal aim of providing insights about the origins of the first living forms and its precursors in the primitive Earth. Investigations from different research fields are needed to converge to viable hypotheses.

The preceding sections have shown that the most advanced methods in computational physics and chemistry can play a particularly useful role in the modeling and simulation of chemical reactions. The level of accuracy has increased significantly and the implementation of highly demanding algorithms is becoming more feasible, allowing the investigation of very specific and complex problems in physics and chemistry. Simulations are notably very effective in the determination of free energy landscapes based on reaction coordinates built from the structural topological properties of the reactants and the products. However, the need to describe, at the quantum level, the forces acting between atoms and the complexity of the calculations therein have until now only permitted the processing of reactions of very small systems in solutions. Consequently, the work undertaken until now in the field of prebiotic chemistry and the molecular origins of life is still in its preliminary stages.

The new challenges that computational physics and chemistry will have to face, in

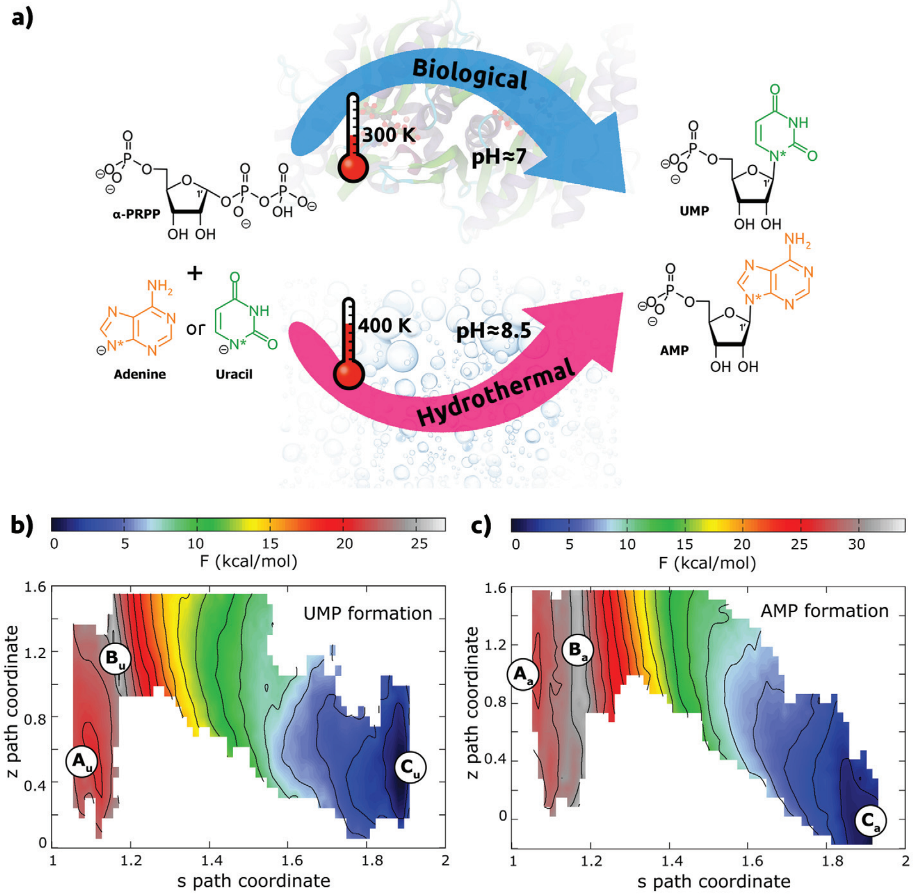

Figure 8: Panel **a)** shows a pictorial representation of nucleotide formation from PRPP and purine/pyrimidine bases through the enzymatic pathway and through a putative prebiotic one. The bottom panels show free energy profiles as a function of the path collective variables s and z for nucleotide formation, for **b)** UMP and **c)** AMP synthesis, respectively. Adapted from ref [33]

the years to come, include the detailed and complete description of truly prebiotic reactions, such as the genesis of RNA and DNA nucleotides and polymers described in the introduction. Just restricting to the RNA world, what mechanisms led to the emergence of nucleotides, combining three distinct chemical entities? Why, amongst the multitude of sugars, did nature select a single one - ribose - to constitute the supporting structure of nucleic acids? Moreover, given the low hydrothermal stability of ribose, what are the most favorable conditions for its formation and further processing in the reaction networks leading to nucleotides? Why is phosphate so central, despite being the salt of a mineral relatively insoluble in water, and despite lacking a major abiotic source on modern Earth? How did the five nitrogenous bases present in DNA and RNA, and essential to the transmission of genetic information, emerge from the primordial soup? Why specifically those five? In the larger picture, how did proteins emerge from the inorganic world? Why do only (or that many) 20 proteinogenic amino-acids exist, and why precisely those? Why was left-handed chirality chemically or biologically selected? Many other crucial questions on the synthesis of vesicles, on compartmentelisation, on metabolism, can be formulated from a molecular/prebiotic chemistry point of view, in order to understand how and when all those entities emerged, when did their chemical interactions developed into a biological one.

These points touch the core of man's big questions inherent in the origin and nature itself of life, which is just a metastable state from the thermodynamic point of view. Interpreting the fundamental prebiotic chemical reactions in terms of thermodynamics and kinetics is thus a crucial step in understanding the emergence of the living from the "primordial soup" imagined almost 150 years ago by Charles Darwin. Computational physics has since provided prebiotic chemistry with tools indispensable for pursuing the ambition and tackling the challenge of answering to these questions.

7. Acknowledgement

This work was supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02, within the framework of the cluster of excellence MATériaux Interfaces Surfaces Environnement (MATISSE) led by Sorbonne Université, and the DIM ACAV+.

[1] C. R. Darwin, Letter to J. D. Hooker, Darwin Project, 1871. DCP-LETT-7471.

[2] S. L. Miller, A production of amino acids under possible primitive earth conditions, Science 117 (1953) 528–529.

[3] J. Oró, Synthesis of adenine from ammonium cyanide, Biochem. Biophys. Res. Commun. 2 (1960) 407 – 412.

[4] W. Napier, J. Wickramasinghe, N. Wickramasinghe, The origin of life in comets, Int. J. Astrobiol 6 (2007) 321–323.

[5] A. Bar-Nun, N. Bar-Nun, S. H. Bauer, C. Sagan, Shock synthesis of amino acids in simulated primitive environments, Science 168 (1970) 470–472.

[6] A. S. Burton, J. C. Stern, J. E. Elsila, D. P. Glavin, J. P. Dworkin, Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites, Chem. Soc. Rev. 41 (2012) 5459–5472.

[7] P. de Marcellus, C. Meinert, I. Myrgorodska, L. Nahon, T. Buhse, L. L. S. d'Hendecourt, U. J. Meierhenrich, Aldehydes and sugars from evolved precometary ice analogs: Importance of ices in astrochemical and prebiotic evolution, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 965–970.

[8] S. Raffaele, B. Giorgia, D. Michela, D. Ernesto, Meteorites as catalysts for prebiotic chemistry, Chem. Eur. J. 19 (2013) 16916–16922.

[9] L. Rotelli, J. M. Trigo-Rodríguez, C. E. Moyano-Cambero, E. Carota, L. Botta, E. Di Mauro, R. Saladino, The key role of meteorites in the formation of relevant prebiotic molecules in a formamide/water environment, Sci. Rep. 6 (2016) 38888.

[10] A. I. Oparin, The origin of life, Mosckovskii Rabochii, 1924. Moscow.

[11] J. B. S. Haldane, The origin of life, Rationalist Annual 148 (1929) 3–10.

[12] W. Martin, J. Baross, D. Kelley, M. J. Russell, Hydrothermal vents and the origin of life, Nat. Rev. Micro 6 (2008) 805–814. 10.1038/nrmicro1991.

[13] G. Wächtershäuser, Groundworks for an evolutionary biochemistry: The iron-sulphur world, Prog. Biophys. Mol. Biol. 58 (1992) 85–201.

[14] A. Brack, Chapter 7.4 Clay minerals and the origin of life, in: F. Bergaya, B. K. Theng, G. Lagaly (Eds.), Handbook of Clay Science, volume 1 of *Developments in Clay Science*, Elsevier, 2006, pp. 379 – 391.

[15] J. P. Ferris, Mineral catalysis and prebiotic synthesis: Montmorillonite-catalyzed formation of RNA, Elements 1 (2005) 145.

[16] J. W. Boclair, P. S. Braterman, B. D. Brister, J. Jiang, S. Lou, Z. Wang, F. Yarberry, Cyanide self-addition, controlled adsorption, and other processes at layered double hydroxides, Orig. Life Evol. Biosph. 31 (2001) 53–69.

[17] R. M. Hazen, Genesis: Rocks, minerals, and the geochemical origin of life, Elements 1 (2005) 135.

[18] R. Saladino, G. Botta, B. M. Bizzarri, E. Di Mauro, J. M. Garcia Ruiz, A global scale scenario for prebiotic chemistry: Silica-based self-assembled mineral structures and formamide, Biochemistry 55 (2016) 2806–2811.

[19] J. E. Šponer, R. Szabla, R. W. Gora, A. M. Saitta, F. Pietrucci, F. Saija, E. Di Mauro, R. Saladino, M. Ferus, S. Civis, J. Šponer, Prebiotic synthesis of nucleic acids and their building blocks at the atomic level - merging models and mechanisms from advanced computations and experiments, Phys. Chem. Chem. Phys. 18 (2016) 20047–20066.

[20] A. Szabo, N. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Dover Publications, 2nd edition, 1989.

[21] R. G. Parr, W. Yang, Density-functional theory of the electronic structure of molecules, Annu. Rev. Phys. Chem. 46 (1995) 701–728.

[22] A. Warshel, R. M. Weiss, An empirical valence bond approach for comparing reactions in solutions and in enzymes, J. Am. Chem. Soc. 102 (1980) 6218–6226.

[23] G. Torrie, J. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, J. Comput. Phys. 23 (1977) 187 – 199.

[24] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12562–12566.

[25] A. Laio, F. L. Gervasio, Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, Rep. Prog. Phys. 71 (2008) 126601.

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

[27] A. M. Saitta, F. Saija, Miller experiments in atomistic computer simulations, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 13768–13773.

[28] F. Pietrucci, A. M. Saitta, Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 15030–15035.

[29] A. Rimola, M. Sodupe, P. Ugliengo, Deep-space glycine formation via Strecker-type reactions activated by ice water dust mantles. A computational approach, Phys. Chem. Chem. Phys. 12 (2010) 5285–5294.

[30] N. N. Nair, E. Schreiner, D. Marx, Peptide synthesis in aqueous environments: The role of extreme conditions on amino acid activation, J. Am. Chem. Soc. 130 (2008) 14148–14160.

[31] D. Muñoz-Santiburcio, C. Wittekindt, D. Marx, Nanoconfinement effects on hydrated excess protons in layered materials, Nat. Comm. 4 (2013) 2349.

[32] M. P. Kroonblawd, F. Pietrucci, A. M. Saitta, N. Goldman, Generating converged accurate free energy surfaces for chemical reactions with a force-matched semi-empirical model, J. Chem. Theory Comput. 14 (2018) 2207–2218.

[33] A. Pérez-Villa, A. M. Saitta, T. Georgelin, J.-F. Lambert, F. Guyot, M.-C. Maurel, F. Pietrucci, Synthesis of rna nucleotides in plausible prebiotic conditions from ab initio computer simulations, The Journal of Physical Chemistry Letters 9 (2018) 4981–4987.

[34] P. Dirac, The Principles of Quantum Mechanics, International series of monographs on physics, Clarendon Press, 1981.

[35] C. D. Sherrill, H. F. Schaefer, The configuration interaction method: Advances in highly correlated approaches, in: P.-O. Löwdin, J. R. Sabin, M. C. Zerner, E. Brändas (Eds.), Advances in Quantum Chemistry, volume 34, Academic Press, 1999, pp. 143 – 269.

[36] M. L. Leininger, W. D. Allen, H. F. S. III, C. D. Sherrill, Is Møller–Plesset perturbation theory a convergent ab initio method?, J. Chem. Phys. 112 (2000) 9213–9222.

[37] R. J. Bartlett, M. Musiał, Coupled-cluster theory in quantum chemistry, Rev. Mod. Phys. 79 (2007) 291–352.

[38] A. Sengupta, R. O. Ramabhadran, K. Raghavachari, Breaking a bottleneck: Accurate extrapolation to "gold standard" CCSD(T) energies for large open shell organic radicals at reduced computational cost, J. Comput. Chem. 37 (2015) 286–295.

[39] J. J. Eriksen, D. A. Matthews, P. Jørgensen, J. Gauss, Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. I. Triples expansions, J. Chem. Phys. 144 (2016) 194102.

[40] T. J. Lee, G. E. Scuseria, Achieving chemical accuracy with coupled-cluster theory, in: S. R. Langhoff (Ed.), Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy, Springer Netherlands, Dordrecht, (1995), pp. 47–108.

[41] C. Attaccalite, S. Sorella, Stable liquid hydrogen at high pressure by a novel ab initio molecular-dynamics calculation, Phys. Rev. Lett. 100 (2008) 114501.

[42] M. Dagrada, M. Casula, A. M. Saitta, S. Sorella, F. Mauri, Quantum Monte Carlo study of the protonated water dimer, J. Chem. Theory Comput. 10 (2014) 1980–1993.

[43] F. Mouhat, S. Sorella, R. Vuilleumier, A. M. Saitta, M. Casula, Fully quantum description of the zundel ion: Combining variational quantum Monte Carlo with path integral Langevin dynamics, J. Chem. Theory Comput. 13 (2017) 2400–2417.

[44] W. Kohn, A. D. Becke, R. G. Parr, Density functional theory of electronic structure, J. Phys. Chem. 100 (1996) 12974–12980.

[45] A. D. Becke, Perspective: Fifty years of density-functional theory in chemical physics, J. Chem. Phys. 140 (2014) 18A301.

[46] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871.

[47] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138.

[48] S. H. Vosko, L. Wilk, M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, Can. J. Phys 58 (1980) 1200–1211.

[49] J. P. Perdew, Y. Wang, Accurate and simple analytic representation of the electron-gas correlation energy, Phys. Rev. B 45 (1992) 13244–13249.

[50] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868.

[51] K. Burke, J. P. Perdew, Y. Wang, Derivation of a Generalized Gradient Approximation: The PW91 Density Functional, Springer US, Boston, MA, pp. 81–111.

[52] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A 38 (1988) 3098–3100.

[53] C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785–789.

[54] C. Adamo, V. Barone, Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models, J. Chem. Phys. 108 (1998) 664–675.

[55] A. D. Becke, Density Functional Thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.

[56] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields, J. Phys. Chem. 98 (1994) 11623–11627.

[57] B. G. Johnson, P. M. W. Gill, J. A. Pople, The performance of a family of density functional methods, J. Chem. Phys. 98 (1993) 5612–5626.

[58] C. Adamo, V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys. 110 (1999) 6158–6170.

[59] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, Climbing the density functional ladder: Nonempirical meta˘generalized gradient approximation designed for molecules and solids, Phys. Rev. Lett. 91 (2003) 146401.

[60] Y. Zhao, D. G. Truhlar, Applications and validations of the Minnesota density functionals, Chem. Phys. Lett. 502 (2011) 1 – 13.

[61] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu, J. Chem. Phys 132 (2010) 154104.

[62] E. Caldeweyher, C. Bannwarth, S. Grimme, Extension of the d3 dispersion coefficient model, J. Chem. Phys 147 (2017) 034112.

[63] A. Tkatchenko, M. Scheffler, Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data, Phys. Rev. Lett. 102 (2009) 073005.

[64] J. Heyd, G. E. Scuseria, M. Ernzerhof, Hybrid functionals based on a screened coulomb potential, J. Chem. Phys 118 (2003) 8207–8215.

[65] J. Heyd, G. E. Scuseria, M. Ernzerhof, Erratum: "hybrid functionals based on a screened coulomb potential" [j. chem. phys. 118, 8207 (2003)], J. Chem. Phys 124 (2006) 219906.

[66] O. A. Vydrov, J. Heyd, A. V. Krukau, G. E. Scuseria, Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals, J. Chem. Phys 125 (2006) 074106.

[67] R. O. Jones, Density functional theory: Its origins, rise to prominence, and future, Rev. Mod. Phys. 87 (2015) 897–923.

[68] G. A. Voth, R. M. Hochstrasser, Transition state dynamics and relaxation processes in solutions: a frontier of physical chemistry, J. Phys. Chem. 100 (1996) 13034–13049.

[69] S. Miertus, E. Scrocco, J. Tomasi, Electrostatic interaction of a solute with a continuum. a direct utilizaion of ab initio molecular potentials for the prevision of solvent effects, Chem. Phys. 55 (1981) 117 – 129.

[70] B. Mennucci, J. Tomasi, R. Cammi, J. R. Cheeseman, M. J. Frisch, F. J. Devlin, S. Gabriel, P. J. Stephens, Polarizable continuum model (PCM) calculations of solvent effects on optical rotations of chiral molecules, J. Phys. Chem. A 106 (2002) 6102–6113.

[71] A. Nitzan, Chemical dynamics in condensed phases: relaxation, transfer and reactions in condensed molecular systems, Oxford University Press, 2006.

[72] J. T. Hynes, Molecules in motion: chemical reaction and allied dynamics in solution and elsewhere, Annu. Rev. Phys. Chem. 66 (2015) 1–20.

[73] A. J. Orr-Ewing, Taking the plunge: chemical reaction dynamics in liquids, Chem. Soc. Rev. 46 (2017) 7597–7614.

[74] M. Born, R. Oppenheimer, Zur quantentheorie der molekeln, Ann. Phys. 389 (1927) 457–484.

[75] L. Skjaerven, S. M. Hollup, N. Reuter, Normal mode analysis for proteins, J. Mol. Struct. 898 (2009) 42–48.

[76] E. Paquet, H. L. Viktor, Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: A computational review, BioMed Res. Int. 2015 (2015) 1–18.

[77] G. Bussi, M. Parrinello, Accurate sampling using Langevin dynamics, Phys. Rev. E 75 (2007) 056707.

[78] D. Frenkel, B. Smit, in: Understanding Molecular Simulation, Academic Press, San Diego, 2nd edition edition, 2002.

[79] J. Åqvist, A. Warshel, Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches, Chem. Rev. 93 (1993) 2523–2544.

[80] U. W. Schmitt, G. A. Voth, Multistate empirical valence bond model for proton transport in water, J. Phys. Chem. B 102 (1998) 5547–5551.

[81] S. C. L. Kamerlin, A. Warshel, The empirical valence bond model: theory and applications, Wiley Interdiscip Rev. Comput. Mol. Sci. 1 (2011) 30–45.

[82] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, R. Kaschner, Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon, Phys. Rev. B 51 (1995) 12947.

[83] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, Phys. Rev. B 58 (1998) 7260.

[84] P. Koskinen, V. Mäkinen, Density-functional tight-binding for beginners, Comput. Mater. Sci. 47 (2009) 237–253.

[85] D. Marx, J. Hutter, Ab Initio Molecular Dynamics, Cambridge University Press, 2009.

[86] M. E. Tuckerman, Ab initio molecular dynamics: basic concepts, current trends and novel applications, J. Phys.: Condens. Matt. 14 (2002) R1297.

[87] A. A. Hassanali, J. Cuny, V. Verdolino, M. Parrinello, Aqueous solutions: state of the art in ab initio molecular dynamics, Phil. Trans. R. Soc. A 372 (2014) 20120482.

[88] J. C. Phillips, Energy-band interpolation scheme based on a pseudopotential, Phys. Rev. 112 (1958) 685–695.

[89] J. C. Phillips, L. Kleinman, New method for calculating wave functions in crystals and molecules, Phys. Rev. 116 (1959) 287–294.

[90] R. Car, M. Parrinello, Unified approach for molecular dynamics and density-functional theory, Phys. Rev. Lett. 55 (1985) 2471–2474.

[91] L.-P. Wang, BIOS 203 lecture 4: Ab initio molecular dynamics, https://www.slideshare.net/bios203/bios203-lect4, Accessed: April 2018.

[92] F. Pietrucci, Strategies for the exploration of free energy landscapes: unity in diversity and challenges ahead, Rev. Phys. 2 (2017) 32–45.

[93] A. F. Voter, A method for accelerating the molecular dynamics simulation of infrequent events, J. Chem. Phys. 106 (1997) 4665–4677.

[94] D. Hamelberg, J. Mongan, J. A. McCammon, Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, J. Chem. Phys. 120 (2004) 11919–11929.

[95] D. R. Glowacki, E. Paci, D. V. Shalashilin, Boxed molecular dynamics: A simple and general technique for accelerating rare event kinetics and mapping free energy in large molecular systems, J. Phys. Chem. B 113 (2009) 16603–16611.

[96] T. Huber, A. E. Torda, W. F. van Gunsteren, Local elevation: a method for improving the searching properties of molecular dynamics simulation, J. Comput. Aid. Mol. Des. 8 (1994) 695–708.

[97] H. Grubmüller, Predicting slow structural transitions in macromolecular systems: Conformational flooding, Phys. Rev. E 52 (1995) 2893.

[98] S. Marsili, A. Barducci, R. Chelli, P. Procacci, V. Schettino, Self-healing umbrella sampling: a non-equilibrium approach for quantitative free energy calculations, J. Phys. Chem. B 110 (2006) 14011–14013.

[99] E. Darve, D. Rodríguez-Gómez, A. Pohorille, Adaptive biasing force method for scalar and vector free energy calculations, J. Chem. Phys. 128 (2008) 144120.

[100] J. Henin, G. Fiorin, C. Chipot, M. Klein, Exploring multidimensional free energy landscapes using time-dependent biases on collective variables, J. Chem. Theory Comput. 6 (2010) 35–47.

[101] J. Schlitter, M. Engels, P. Krueger, Targeted molecular dynamics: A new approach for searching pathways of conformational transitions, J. Mol. Graph. 12 (1994) 84 – 89.

[102] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, K. Schulten, Steered molecular dynamics, in: Computational Molecular Dynamics: Challenges, Methods, Ideas: Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21–24, 1997, Springer, Berlin, (1999), pp. 39–65.

[103] M. Marchi, P. Ballone, Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems, J. Chem. Phys. 110 (1999) 3697–3702.

[104] E. Paci, M. Karplus, Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations1, J. Mol. Biol. 288 (1999) 441 – 459.

[105] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, J. Comput. Chem. 13 (1992) 1011–1021.

[106] B. Roux, The calculation of the potential of mean force using computer simulations, Comput. Phys. Commun. 91 (1995) 275 – 282.

[107] F. Zhu, G. Hummer, Convergence and error estimation in free energy calculations using the weighted histogram analysis method, J. Comput. Chem. 33 (2012) 453–465.

[108] E. Rosta, G. Hummer, Free energies from dynamic weighted histogram analysis using unbiased Markov state model, J. Chem. Theory Comput. 11 (2014) 276–285.

[109] H. Wu, F. Paul, C. Wehmeyer, F. Noé, Multiensemble Markov models of molecular thermodynamics and kinetics, Proc. Natl. Acad. Sci. U.S.A. (2016) 201525092.

[110] E. Thiede, B. Van Koten, J. Weare, A. R. Dinner, Eigenvector method for umbrella sampling enables error analysis, arXiv:1603.04505 (2016).

[111] A. Barducci, M. Bonomi, M. Parrinello, Metadynamics, WIREs Comput. Mol. Sci. 1 (2011) 826–843.

[112] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, M. Parrinello, Assessing the accuracy of metadynamics, J. Phys. Chem. B 109 (2005) 6714–6721.

[113] G. Bussi, A. Laio, M. Parrinello, Equilibrium free energies from nonequilibrium metadynamics, Phys. Rev. Lett. 96 (2006) 090601.

[114] Y. Crespo, F. Marinelli, F. Pietrucci, A. Laio, Metadynamics convergence law in a multidimensional system, Phys. Rev. E 81 (2010) 055701.

[115] J. F. Dama, M. Parrinello, G. A. Voth, Well-tempered metadynamics converges asymptotically, Phys. Rev. Lett. 112 (2014) 240602.

[116] P. Tiwary, M. Parrinello, A time-independent free energy estimator for metadynamics, J. Phys. Chem. B 119 (2015) 736–742.

[117] P. L. Geissler, C. Dellago, D. Chandler, Kinetic pathways of ion pair dissociation in water, J. Phys. Chem. B 103 (1999) 3706–3710.

[118] A. Berezhkovskii, A. Szabo, One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions, J. Chem. Phys. 122 (2005) 014503.

[119] R. B. Best, G. Hummer, Reaction coordinates and rates from transition paths, Proc. Natl. Acad. Sci. USA of the United States of America 102 (2005) 6732–6737.

[120] M. A. Rohrdanz, W. Zheng, C. Clementi, Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions, Annu. Rev. Phys. Chem. 64 (2013) 295–316.

[121] B. Peters, Reaction coordinates and mechanistic hypothesis tests, Annu. Rev. Phys. Chem. 67 (2016) 669–690.

[122] K. Fukui, Formulation of the reaction coordinate, J. Phys. Chem. 74 (1970) 4161–4163.

[123] K. Ishida, K. Morokuma, A. Komornicki, The intrinsic reaction coordinate, J. Chem. Phys. 66 (1977) 2153–2156.

[124] E. Herbst, Chemistry in the interstellar medium, Annu. Rev. Phys. Chem. 46 (1995) 27–54.

[125] V. S. Airapetian, A. Glocer, G. Gronoff, E. Hébrard, W. Danchi, Prebiotic chemistry and atmospheric warming of early earth by an active young sun, Nat. Geosci. 9 (2016) 452–455.

[126] R. Saladino, E. Carota, G. Botta, M. Kapralov, G. N. Timoshenko, A. Y. Rozanov, E. Krasavin, E. Di Mauro, Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) E2746–E2755.

[127] R. Saladino, B. M. Bizzarri, L. Botta, J. Šponer, J. E. Šponer, T. Georgelin, M. Jaber, B. Rigaud, M. Kapralov, G. N. Timoshenko, A. Rozanov, E. Krasavin, A. M. Timperio, E. Di Mauro, Proton irradiation: a key to the challenge of n-glycosidic bond formation in a prebiotic context, Sci. Rep. 7 (2017) 14709.

[128] N. Balucani, C. Ceccarelli, V. Taquet, Formation of complex organic molecules in cold objects: the role of gas-phase reactions, Mon. Notices Royal Astron. Soc. 449 (2015) L16–L20.

[129] V. Barone, C. Latouche, D. Skouteris, F. Vazart, N. Balucani, C. Ceccarelli, B. Lefloch, Gas-phase formation of the prebiotic molecule formamide: insights from new quantum computations, Mon. Notices Royal Astron. Soc. 453 (2015) L31–L35.

[130] R. Spezia, Y. Jeanvoine, W. L. Hase, K. Song, A. Largo, Synthesis of formamide and related organic species in the interstellar medium via chemical dynamics simulations, The Astrophysical Journal 826 (2016) 107.

[131] J. L. Snow, G. Orlova, V. Blagojevic, D. K. Bohme, Gas-phase ionic syntheses of amino acids: β versus α, J. Am. Chem. Soc. 129 (2007) 9910–9917.

[132] Y. Jeanvoine, A. Largo, W. L. Hase, R. Spezia, Gas phase synthesis of protonated glycine by chemical dynamics simulations, J. Phys. Chem. A 122 (2018) 869–877.

[133] A. Rimola, M. Sodupe, P. Ugliengo, Computational study of interstellar glycine formation occurring at radical surfaces of water-ice dust particles, Astrophys. J. 754 (2012) 24.

[134] A. Rimola, P. Ugliengo, The role of defective silica surfaces in exogenous delivery of prebiotic compounds: clues from first principles calculations, Phys. Chem. Chem. Phys. 11 (2009) 2497–2506.

[135] L. Song, J. Kastner, Formation of the prebiotic molecule NH_2CHO on astronomical amorphous solid water surfaces: accurate tunneling rate calculations, Phys. Chem. Chem. Phys. 18 (2016) 29278–29285.

[136] J. O. Richardson, Microcanonical and thermal instanton rate theory for chemical reactions at all temperatures, Faraday Discuss. 195 (2016) 49–67.

[137] H. J. Cleaves, S. L. Miller, Oceanic protection of prebiotic organic compounds from UV radiation, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 7260–7263.

[138] C. Meinert, P. de Marcellus, L. L. S. d'Hendecourt, L. Nahon, N. C. Jones, S. V. Hoffmann, J. H. Bredehöft, U. J. Meierhenrich, Photochirogenesis: Photochemical models on the absolute asymmetric formation of amino acids in interstellar space, Phys. Life Rev 8 (2011) 307 – 330.

[139] J. Finley, P. Åke Malmqvist, B. O. Roos, L. Serrano-Andrés, The multi-state CASPT2 method, Chem. Phys. Lett. 288 (1998) 299 – 306.

[140] P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, R. Shepard, Multiconfiguration self-consistent field and multireference configuration interaction methods and applications, Chem. Rev. 112 (2012) 108–181.

[141] C. Hättig, Structure optimizations for excited states with correlated second-order methods: CC2 and ADC(2), in: H. J. Å. Jensen (Ed.), Response Theory and Molecular Properties (A Tribute to Jan Linderberg and Poul Jørgensen), volume 50 of *Adv. Quantum Chem.*, Academic Press, 2005, pp. 37 – 60.

[142] J. P. Ferris, L. E. Orgel, An unusual photochemical rearrangement in the synthesis of adenine from hydrogen cyanide, J. Am. Chem. Soc. 88 (1966) 1074–1074.

[143] R. Sanchez, J. Ferris, L. E. Orgel, Conditions for purine synthesis: Did prebiotic synthesis occur at low temperatures?, Science 153 (1966) 72–73.

[144] E. Boulanger, A. Anoop, D. Nachtigallova, W. Thiel, M. Barbatti, Photochemical steps in the prebiotic synthesis of purine precursors from HCN, Angew. Chem. Int. Ed 52 (2013) 8000–8003.

[145] R. Szabla, J. E. Šponer, J. Šponer, A. L. Sobolewski, R. W. Gora, Solvent effects on the photochemistry of 4-aminoimidazole-5-carbonitrile, a prebiotically plausible precursor of purines, Phys. Chem. Chem. Phys. 16 (2014) 17617–17626.

[146] M. W. Powner, B. Gerland, J. D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions, Nature 459 (2009) 239–242.

[147] R. Szabla, D. Tuna, R. W. Góra, J. Šponer, A. L. Sobolewski, W. Domcke, Photochemistry of 2-aminooxazole, a hypothetical prebiotic precursor of RNA nucleotides, J. Phys. Chem. Lett. 4 (2013) 2785–2788.

[148] J. D. Bernal, The physical basis of life, Proc. Phys. Soc. A 62 (1949) 537.

[149] D. Gibbs, R. Lohrmann, L. E. Orgel, Template-directed synthesis and selective adsorption of oligoadenylates on hydroxyapatite, J. Mol. Evol. 15 (1980) 347–354.

[150] J. P. Ferris, Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life, Phil. Trans. R. Soc. B 361 (2006) 1777–1786.

[151] J. V. Smith, Biochemical evolution. i. polymerization on internal, organophilic silica surfaces of dealuminated zeolites and feldspars, Proc. Natl. Acad. Sci. USA 95 (1998) 3370–3375.

[152] I. Parsons, M. R. Lee, J. V. Smith, Biochemical evolution ii: Origin of life in tubular microstructures on weathered feldspar surfaces, Proc. Natl. Acad. Sci. USA 95 (1998) 15173–15176.

[153] J. V. Smith, F. P. Arnold, I. Parsons, M. R. Lee, Biochemical evolution iii: Polymerization on organophilic silica-rich surfaces, crystal–chemical modeling, formation of first cells, and geological clues, Proc. Natl. Acad. Sci. USA 96 (1999) 3479–3485.

[154] A. Rimola, M. Sodupe, P. Ugliengo, Aluminosilicate surfaces as promoters for peptide bond formation: an assessment of Bernal's hypothesis by ab initio methods, J. Am. Chem. Soc. 129 (2007) 8333–8344.

[155] A. Rimola, P. Ugliengo, M. Sodupe, Formation versus Hydrolysis of the Peptide Bond from a Quantum-mechanical Viewpoint: The Role of Mineral Surfaces and Implications for the Origin of Life, Int. J. Mol. Sci. 10 (2009) 746–760.

[156] P. Mignon, P. Ugliengo, M. Sodupe, Theoretical study of the adsorption of RNA/DNA bases on the external surfaces of Na+-montmorillonite, J. Phys. Chem. C 113 (2009) 13741–13749.

[157] P. Mignon, M. Sodupe, Theoretical study of the adsorption of DNA bases on the acidic external surface of montmorillonite, Phys. Chem. Chem. Phys. 14 (2012) 945–954.

[158] P. Mignon, M. Sodupe, Structural behaviors of cytosine into the hydrated interlayer of Na+-montmorillonite clay. An ab initio molecular dynamics study, J. Phys. Chem. C 117 (2013) 26179–26189.

[159] G. Wächtershäuser, Before enzymes and templates: theory of surface metabolism., Microbiol. Rev. 52 (1988) 452–484.

[160] C. Huber, G. Wächtershäuser, Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions, Science 276 (1997) 245–247.

[161] C. Huber, G. Wächtershäuser, Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: Implications for the origin of life, Science 281 (1998) 670–672.

[162] R. Pollet, C. Boehme, D. Marx, Ab initio simulations of desorption and reactivity of glycine at a water-pyrite interface at “iron-sulfur world” prebiotic conditions, Orig. Life Evol. Biosph. 36 (2006) 363–379.

[163] N. N. Nair, E. Schreiner, D. Marx, Glycine at the pyrite—water interface: the role of surface defects, J. Am. Chem. Soc. 128 (2006) 13815–13826.

[164] E. Schreiner, N. N. Nair, C. Wittekindt, D. Marx, Peptide synthesis in aqueous environments: The role of extreme conditions and pyrite mineral surfaces on formation and hydrolysis of peptides, J. Am. Chem. Soc. 133 (2011) 8216–8226.

[165] A. Stirling, T. Rozgonyi, M. Krack, M. Bernasconi, Pyrite in contact with supercritical water: the desolation of steam, Phys. Chem. Chem. Phys. 17 (2015) 17375–17379.

[166] A. Stirling, T. Rozgonyi, M. Krack, M. Bernasconi, Prebiotic NH_3 formation: Insights from simulations, Inorg. Chem. 55 (2016) 1934–1939.

[167] O. L. E., Prebiotic chemistry and the origin of the rna world, Crit. Rev. Biochem. Mol. Biol. 39 (2004) 99–123.

[168] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, Borate minerals stabilize ribose, Science 303 (2004) 196–196.

[169] C. Cossetti, C. Crestini, R. Saladino, E. D. Mauro, Borate minerals and rna stability, Polymers 2 (2010) 211–228.

[170] J. Šponer, B. Sumpter, J. Leszczynski, J. Šponer, M. Fuentes-Cabrera, Theoretical study on the factors controlling the stability of the borate complexes of ribose, arabinose, lyxose, and xylose, Chem. Eur. J. 14 (2008) 9990–9998.

[171] L. Koziol, N. Goldman, Prebiotic hydrocarbon synthesis in impacting reduced astrophysical icy mixtures, Astrophys. J. 803 (2015) 91.

[172] N. Goldman, E. J. Reed, L. E. Fried, I.-F. William, Kuo, A. Maiti, Synthesis of glycine-containing complexes in impacts of comets on early Earth, Nat. Chem. 2 (2010) 949.

[173] Z. Martins, M. C. Price, N. Goldman, M. A. Sephton, M. J. Burchell, Shock synthesis of amino acids from impacting cometary and icy planet surface analogues, Nat. Geosci. 6 (2013) 1045.

[174] D. H. Robertson, D. W. Brenner, C. T. White, Split shock waves from molecular dynamics, Phys. Rev. Lett. 67 (1991) 3132–3135.

[175] K. Kadau, T. C. Germann, P. S. Lomdahl, B. L. Holian, Microscopic view of structural phase transitions induced by shock waves, Science 296 (2002) 1681–1684.

[176] E. J. Reed, L. E. Fried, J. D. Joannopoulos, A method for tractable dynamical studies of single and double shock compression, Phys. Rev. Lett. 90 (2003) 235503.

[177] E. J. Reed, L. E. Fried, W. D. Henshaw, C. M. Tarver, Analysis of simulation technique for steady shock waves in materials with analytical equations of state, Phys. Rev. E 74 (2006) 056706.

[178] N. Goldman, I. Tamblyn, Prebiotic chemistry within a simple impacting icy mixture, J. Phys. Chem. A 117 (2013) 5124–5131.

[179] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. S. Pande, T. J. Martínez, Discovering chemistry with an ab initio nanoreactor, Nat. Chem. 6 (2014) 1044.

[180] Y. A. Jeilani, P. N. Williams, S. Walton, M. T. Nguyen, Unified reaction pathways for the prebiotic formation of RNA and DNA nucleobases, Phys. Chem. Chem. Phys. 18 (2016) 20177–20188.

[181] M. Ferus, S. Civiš, A. Mládek, J. Šponer, L. Juha, J. E. Šponer, On the road from formamide ices to nucleobases: Ir-spectroscopic observation of a direct reaction between cyano radicals and formamide in a high-energy impact event, J. Am. Chem. Soc. 134 (2012) 20788–20796.

[182] M. Ferus, D. Nesvorný, J. Šponer, P. Kubelík, R. Michalčíková, V. Shestivská, J. E. Šponer, S. Civiš, High-energy chemistry of formamide: A unified mechanism of nucleobase formation, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 657–662.

[183] J. L. Bada, How life began on Earth: a status report, Earth Planet. Sci. Lett. 226 (2004) 1–15.

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

[185] G. Cassone, J. Šponer, J. E. Šponer, F. Pietrucci, A. M. Saitta, F. Saija, Synthesis of (d)-erythrose from glycolaldehyde aqueous solutions under electric field, Chem. Commun. 54 (2018) 3211–3214.

[186] J. E. Šponer, J. Šponer, M. Fuentes-Cabrera, Prebiotic routes to nucleosides: A quantum chemical insight into the energetics of the multistep reaction pathways, Chem. Eur. J. 17 (2010) 847–854.

[187] S. Laporte, F. Finocchi, L. Paulatto, M. Blanchard, E. Balan, F. Guyot, A. M. Saitta, Strong electric fields at a prototypical oxide/water interface probed by ab initio molecular dynamics: MgO(001), Phys. Chem. Chem. Phys. 17 (2015) 20382–20390.

[188] A. J. Orr-Ewing, Perspective: Bimolecular chemical reaction dynamics in liquids, J. Chem. Phys. 140 (2014) 090901.

[189] V. S. Nguyen, T. M. Orlando, J. Leszczynski, M. T. Nguyen, Theoretical study of the decomposition of formamide in the presence of water molecules, J. Phys. Chem. A 117 (2013) 2543–2555.

[190] C. A. Guido, F. Pietrucci, G. A. Gallet, W. Andreoni, The fate of a Zwitterion in water from ab initio molecular dynamics: monoethanolamine (MEA)-CO_2, J. Chem. Theory Comput. 9 (2012) 28–32.

[191] C. Ma, F. Pietrucci, W. Andreoni, Capturing CO_2 in monoethanolamine (MEA) aqueous solutions: fingerprints of carbamate formation assessed with first-principles simulations, J. Phys. Chem. Lett. 5 (2014) 1672–1677.

[192] V. Sharma, A. Annila, Natural process–natural selection, Biophys. Chem. 127 (2007) 123–128.

[193] R. Pascal, A. Pross, J. D. Sutherland, Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics, Open Biol. 3 (2013) 130156.

[194] B. Ensing, M. De Vivo, Z. Liu, P. Moore, M. L. Klein, Metadynamics as a tool for exploring free energy landscapes of chemical reactions, Acc. Chem. Res. 39 (2006) 73–81.

[195] B. R. Prasad, N. V. Plotnikov, A. Warshel, Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping, J. Phys. Chem. B 117 (2013) 153–163.

[196] D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space, J. Chem. Phys. 126 (2007) 054103.

[197] D. Branduardi, M. De Vivo, N. Rega, V. Barone, A. Cavalli, Methyl phosphate dianion hydrolysis in solution characterized by path collective variables coupled with DFT-based enhanced sampling simulations, J. Chem. Theory Comput. 7 (2011) 539–543.

[198] G. A. Gallet, F. Pietrucci, W. Andreoni, Bridging static and dynamical descriptions of chemical reactions: An ab initio study of CO_2 interacting with water molecules, J. Chem. Theory Comput. 8 (2012) 4029–4039.

[199] F. Pietrucci, W. Andreoni, Fate of a graphene flake: A new route toward fullerenes disclosed with ab initio simulations, J. Chem. Theory Comput. 10 (2014) 913–917.

[200] M. Sprik, Computation of the pK of liquid water using coordination constraints, Chem. Phys. 258 (2000) 139–150.

[201] M. Iannuzzi, A. Laio, M. Parrinello, Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics, Phys. Rev. Lett. 90 (2003) 238302.

[202] J. M. Park, A. Laio, M. Iannuzzi, M. Parrinello, Dissociation mechanism of acetic acid in water, J. Am. Chem. Soc. 128 (2006) 11318–11319.

[203] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, M. Parrinello, PLUMED: a portable plugin for free-energy calculations with molecular dynamics, Comput. Phys. Commun. 180 (2009) 1961–1972.

[204] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, Plumed 2: New feathers for an old bird, Comput. Phys. Commun. 185 (2014) 604 – 613.

[205] F. Pietrucci, http://sites.google.com/site/fabiopietrucci/home/download, 2018.

[206] V. S. Nguyen, H. L. Abbott, M. M. Dawley, T. M. Orlando, J. Leszczynski, M. T. Nguyen, Theoretical study of formamide decomposition pathways, J. Phys. Chem. A 115 (2011) 841–851.

[207] E. Iglesias, L. Montenegro, Basicity of amides in water and in aqueous micellar solutions of SDS. Their influence on micellar structure, J. Chem. Soc. Faraday T. 92 (1996) 1205–1212.

[208] C. Chaudhuri, J. C. Jiang, C.-C. Wu, X. Wang, H.-C. Chang, Characterization of protonated formamide-containing clusters by infrared spectroscopy and ab initio calculations. ii. hydration of formamide in the gas phase, J. Phys. Chem. A 105 (2001) 8906–8915.

[209] J. E. Elsila, J. C. Aponte, D. G. Blackmond, A. S. Burton, J. P. Dworkin, D. P. Glavin, Meteoritic amino acids: diversity in compositions reflects parent body histories, ACS Centr. Sci. 2 (2016) 370–379.

[210] S. L. Miller, J. E. Van Trump, The Strecker synthesis in the primitive ocean, in: Origin of life, Springer, 1981, pp. 135–141.

[211] E. Peltzer, J. Bada, G. Schlesinger, S. Miller, The chemical conditions on the parent body of the Murchison meteorite: some conclusions based on amino, hydroxy and dicarboxylic acids, Adv. Space Res. 4 (1984) 69–74.

[212] V. Riffet, G. Frison, G. Bouchoux, Quantum-chemical modeling of the first steps of the Strecker synthesis: From the gas-phase to water solvation, J. Phys. Chem. A (2018).

[213] T. Koga, H. Naraoka, A new family of extraterrestrial amino acids in the Murchison meteorite, Sci. Rep. 7 (2017) 636.

[214] A. Savin, R. Nesper, S. Wengert, T. F. Fässler, ELF: The electron localization function, Angew. Chem. Int. Ed. 36 (1997) 1808–1832.

[215] F. Pietrucci, J. Aponte, R. D. Starr, A. Pérez-Villa, J. E. Elsila, J. P. Dworkin, A. M. Saitta, Hydrothermal decomposition of amino acids and origins of prebiotic meteoritic organics, ACS Earth Space Chem. 2 (2018) 588–598.

[216] W. Guo, J. M. Eiler, Temperatures of aqueous alteration and evidence for methane generation on the parent bodies of the CM chondrites, Geochim. Cosmochim. Acta 71 (2007) 5565–5575.

[217] A. N. Alexandrova, W. L. Jorgensen, On the mechanism and rate of spontaneous decomposition of amino acids, J. Phys. Chem. B 115 (2011) 13624–13632.

[218] P. Geerlings, S. Fias, Z. Boisdenghien, F. De Proft, Conceptual DFT: chemistry from the linear response function, Chem. Soc. Rev. 43 (2014) 4989–5008.

[219] J. S. Cox, T. M. Seward, The reaction kinetics of alanine and glycine under hydrothermal conditions, Geochim. Cosmochim. Acta 71 (2007) 2264–2284.

[220] L. Da Silva, M.-C. Maurel, D. Deamer, Salt-promoted synthesis of rna-like molecules in simulated hydrothermal conditions, J. Mol. Evol. 80 (2015) 86–97.

[221] J. G. Forsythe, S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernández, N. V. Hud, Ester-mediated amide bond formation driven by wet–dry cycles: A possible path to polypeptides on the prebiotic earth, Angew. Chem. Int. Ed. 54 (2015) 9871–9875.

[222] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis, Nat. Commun. 9 (2018) 163.

[223] B. J. Cafferty, N. V. Hud, Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology, Curr. Opin. Chem. Biol. 22 (2014) 146–157.

[224] M. Akouche, M. Jaber, M.-C. Maurel, J.-F. Lambert, T. Georgelin, Phosphoribosyl pyrophosphate: A molecular vestige of the origin of life on minerals, Angew. Chem. 129 (2017) 8028–8031.