SORBONNE UNIVERSITÉ

PROGRAMME INTITUTS ET INITIATIVES

Appel à projet – campagne 2021

Proposition de projet de recherche doctoral (PRD)

ISCD-Institut des Sciences du calcul des Données

Intitulé du projet de recherche doctoral (PRD): PRebiOtic ChEmistry kinetics from Stochastic Sampling (PROCESS)

Directeur.rice de thèse porteur.euse du projet (titulaire d'une HDR) :

NOM : **SAITTA** Prénom : **A. Marco**
Titre : Professeur des Universités ou
e-mail : marco.saitta@sorbonne-université.fr
Adresse professionnelle : Campus PMC, 23-24/309
(site, adresse, bât., bureau)

Unité de Recherche :
Intitulé : IMPMC
Code *(ex. UMR xxxx)* : UMR 7590

École Doctorale de rattachement de l'équipe (future école doctorale du.de la doctorant.e) : **ED397-Physique Chimie des Matériaux**

Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) : 3 (2018, 2019, 2020), tous à 50%, quotité totale 1,5

Co-encadrant.e :

NOM : **VUILLEUMIER** Prénom : **Rodolphe**
Titre : Professeur des Universités ou HDR ☒
e-mail : rodolphe.vuilleumier@ens.fr

Unité de Recherche :
Intitulé : PASTEUR
Code *(ex. UMR xxxx)* : UMR 8640

École Doctorale de rattachement : **ED388-ChimiePhysiqueChimieAnalytique ParisCentre**
Ou si ED non Alliance SU :

Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) : 4 (2017, 2019, 2020, 2020), dont 3 à 50%, quotité totale 2,5

Co-encadrant.e :

NOM : **Pietrucci** Prénom : **Fabio**
Titre : Maître de Conférences des Universités ou HDR ☒

e-mail : fabio.pietrucci@sorbonne-universite.fr

Unité de Recherche :
Intitulé : Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie
Code *(ex. UMR xxxx)* : UMR 7590

École Doctorale de rattachement : **ED397-Physique Chimie des Matériaux**
Ou si ED non Alliance SU :

Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) :3 (2018, 2019, 2020), tous à 50%, quotité totale 1,5

Cotutelle internationale : ☒ Non ☐ Oui, précisez Pays et Université :

Selon vous, ce projet est-il susceptible d'intéresser une autre Initiative ou un autre Institut ?
☐ Non ☒ Oui, précisez ISim - Initiative Sces et ingénierie moléculaires

Description du projet de recherche doctoral *(en français ou en anglais) :*

Ce texte sera diffusé en ligne : il ne doit pas excéder 3 pages et est écrit en interligne simple.

Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.

Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le projet.
Préciser le profil d'étudiant(e) recherché.

Scientific context

Research in origins of life aims at finding answers to the formidably complex problem of the emergence of life from the modern versions of Charles Darwin's celebrated "primordial soup". One of the crucial questions that need to be addressed is the emergence of life chemical building-blocks, such as amino acids, sugars, nucleobases, and fatty acids. However, this is a formidable task as the environmental, atmospheric and geochemical conditions of the early Earth can only be inferred and speculated the relevant chemical reactions having possibly occurred in the presence of solvent, salts, minerals, variable pH, elevated temperatures and pressures, including out-of-equilibrium conditions.

In the last few years, new microscopic insights into "small-scale" fundamental phenomena having possibly occurred on the early Earth could be gained in computer simulations. Computer-based ab initio methods are nowadays a full, self-standing branch of prebiotic chemistry, both for their quantum-based accuracy, and for their multiscale extent, as we have shown in our group in the recent review invited by the prestigious journal Physics of Life: Reviews [A Pérez-Villa, F Pietrucci, A.M. Saitta, Prebiotic chemistry and origins of life research with atomistic computer simulations, Phys. Life Rev. 34-35, 105-135 (2020)]. A significant step forward towards quantitative results in computational prebiotic chemistry requires the determination of the free energy surface (FES) and kinetics of the key chemical reactions.

The major bottleneck, for state-of-the-art ab initio computational approaches, is that prebiotic chemical reactions in realistic environments, i.e. in the liquid phase, even at equilibrium conditions, form an extremely complex network of elementary steps, and brings a serious challenge in estimating entropic contributions, which is a major difference with respect to gas phase calculations. Establishing a quantitative connection between prebiotic reactivity and the various sources of chemical energy thus requires a change of paradigm in the definition of the so-called generalized reaction coordinates (RC). This is even more frustrating, as geochemistry only provides loose constraints on the thermodynamic, kinetic, and chemical characteristics of primordial environments from which life has emerged; this limitation is even more problematic since it is nowadays widely accepted that out-of-equilibrium conditions (pressure and/or

temperature gradients, day/night cycles, seasonality, alternance of wet and dry) have certainly been crucial in the emergence of life building blocks. In other words, prebiotically scenarios are meaningful as long as they are compatible with the rates of the corresponding chemical reactions.

Theoretical and computational problem

In order to take into account these realistic prebiotic environments, we have recently developed a novel topological approach, which maps the chemical space of a given reaction onto a matrix space, which then allows to describe variations of the chemical network along a well-chosen but putative RC [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, PNAS 112, 15030 (2015)], at the same time i) calculating the FES of the whole explored space, and ii) taking into account the active participation and the multifaceted role of the explicit solvent.

However, if the free energy barriers of relevant prebiotic chemical reactions can now be accurately reconstructed by means of our method, these calculations, fully ab initio, are computationally very expensive. This problem can partly be solved by using our own neural network potentials, that we are developing right now, which allow faster calculations. On the other hand, the quantitative prediction of kinetic properties, like transition rates, is even more difficult, and it often rests on approximations - sometimes inadequate - like those of transition state theory, classical nucleation theory, Markovian dynamics, etc. Algorithms addressing the kinetics or rare events are comparatively less abundant and developed than those for free-energy calculations, and they typically require the production of very extensive simulation data sets.

Central idea

From a theoretical viewpoint, the framework of free energy landscapes with barriers separating metastable states, customarily invoked to interpret rare events, corresponds to the analysis of equilibrium probability distributions as a function of a small number of collective variables (CVs). The kinetics of the reaction, and its mechanism, is however encoded in the transition path from one metastable basin (the reactants) to another (the products). Transition paths are dynamical trajectories connecting directly the reactants to the products. In some limit the transition paths in the space of CVs can be concentrated along the minimum free energy path (MFEP) and this is the reason why free energy landscapes are often used also for this purpose. [Maragliano, L et al. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. J. Chem. Phys. 2006, 125 (2), 024106.] However, this is not ensured, in particular for chemical reactions in solutions. We plan here to extend on our previous work to perform transition path sampling and to map them on the CVs that we have designed for describing the free energy landscape of many different chemical reactions.

From the analysis of the density of transition paths or of a related quantity, the committor probability, the first aim of the work is to parametrize a RC to describe the reaction. In a certain sense, the committor probability, i.e. the probability that, starting from a point in phase space, the system evolves forward to the product state and backward from the reactant state, was suggested as the best RC [Vanden-Eijnden, E. Towards a Theory of Transition Paths. J Stat Phys 2006, 123 (3), 503–523; Vanden-Eijnden, E.; Tal, F. A. Transition State Theory: Variational Formulation, Dynamical Corrections, and Error Estimates. J. Chem. Phys. 2005, 123 (18), 184103]. An alternative approach to achieve the same aim, will be to parametrize the RC as a function of the CVs to best reproduce the measured committor, for example from a maximum likelihood approach. This will allow us to extend the set of CVs that are used and a good candidate set of CVs are the descriptors used to describe atomic environments in machine learning kernels for potential energy surfaces.

The second aim of this project is then to study the kinetics of the chemical reactions by inspection of the dynamics of the RC. Transition Path Sampling, a modern approach to Transition State

Theory, identifies the reaction rate as the flux along the RC. However, this fails if the projected dynamics along the CV has a strong memory. Here, we plan to exploit the current developments carried out in the MAESTRO project, which clearly justifies the pertinence of our application to ISCD.

Building on our ongoing work on Langevin equation in the context of ionic solutions and the nucleation of Solid-Electrolyte Interphases (SEI) within batteries, we wish to extend it to the field of chemical reactions in the presence of explicit solvents, which is indeed more likely to display memory effects. In practice, we envisage to use actual simulated dynamics of model reaction steps in the synthesis of glycine and ribose, already obtained, to extract a memory kernel for the projected dynamics and the RC. The knowledge of both the RC and of the projected dynamics will provide us with all information about the reaction mechanism and kinetics [Peters, B. Reaction coordinates and mechanistic hypothesis tests. Ann. Rev. Phys. Chem. 67, 669-690 (2016); Banushkina, P.V.; Krivov, S.V.. Optimal reaction coordinates. Comp. Mol. Sci. 6.6, 748-763 (2016); Pietrucci, F.. Strategies for the exploration of free energy landscapes: unity in diversity and challenges ahead. Rev. Phys. 2, 32 (2017); Camilloni C. and Pietrucci, F. Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. Adv. Phys. X 3, 1477531 (2018).]

Objectives

The broad long-term aim of this project will be to provide a new understanding in one of the most compelling problems of the chemical origins of life. Was the abiotic synthesis of the RNA nucleoside, or of amino acids, driven by thermodynamics or kinetics? Are we able to quantitatively predict the kinetics and rates of those complex reactions?

The main conceptual objective of this collaborative effort is thus to gain insights into the capability of atomistic simulations to go beyond the (already complex) free-energy picture and provide i) a methodological framework to assess the quality of reaction coordinates; ii) the tools to obtain information on kinetics. In practice, we will dispose of a battery of ab initio molecular dynamics trajectories, generated by our third year PhD student, to study the glycine synthesis in solution, starting from small precursors, and following the so-called Strecker mechanism [T. Magrino, F. Pietrucci, A.M. Saitta, Step by Step Strecker Amino Acid Synthesis from ab Initio Prebiotic Chemistry, J. Phys. Chem. Lett., to appear (2021)].

The role of the hired PhD student will be to exploit those ab initio trajectories of this complex reaction path, and to build up, in collaboration with the MAESTRO postdocs, the memory kernels required to establish the quality of the CV, and thus to improve the choice of the best reaction coordinates. Once this preliminary (but long!) analysis is achieved, he/she will carry out the necessary developments in order to identify the best RC for each reaction step of the glycine synthesis, and then try to infer the kinetics information of those steps, to be compared to experimental data.

Interest of the collaboration, complementarity of the team

The project builds on the complementary expertise of the two groups. On the one hand, the IMPMC partner has made significant recent advances on the topological description and the ab initio computational modeling of realistic condensed-phase chemical reactions, explicitly taking into account key parameters such as temperature, pressure, pH and addressing inhomogeneous and/or discontinuous systems. On the other hand, the PASTEUR partner has expertise on the analysis of projected dynamics along effective coordinates as well as determination of memory kernels. The candidate will also benefit from the environment of the Theory group at PASTEUR, which has a focus on chemical reactions and hosts J.T. Hynes, as emeritus member, who has pioneered the field of the computation of reaction rates in solution beyond the Kramers

approximation [Grosjean, B. ; Bocquet, M.-L. ; Vuilleumier, R. Versatile Electrification of Two-Dimensional Nanomaterials in Water. Nature Comm. (2019) 10, 1656; Lesnicki, D., R. Vuilleumier, A. Carof, B. Rotenberg. Molecular Hydrodynamics from Memory Kernels. Phys. Rev. Lett. 116: 147804 (2016); Carof, A., Vuilleumier, R. and Rotenberg, B. Two algorithms to compute projected correlation functions in molecular dynamics simulations. J. Chem. Phys. 140, 124103 (2014).]

Merci d'enregistrer votre fichier au format PDF et de le nommer :
«ACRONYME de l'initiative/institut – AAP 2021 – NOM Porteur.euse Projet »

Fichier envoyer simultanément par e-mail à l'ED de rattachement et au programme :
cd_instituts_et_initiatives@listes.upmc.fr avant le 20 février.