

PhD Position in AI accelerated simulations of chemical reactivity at ChimieParisTech – PSL University

Application Deadline: May 20 at 17h00 (CET)

Start Date: Between September 1, 2025, and November 30, 2025

Supervisor: Junior Prof. Thijs Stuyver, fellow of PR[AI]RIE-PSAI institute

Host Institution: ChimieParisTech – PSL University

Funding: PR[AI]RIE-PSAI cluster

Research Topic

One of the main research interests within the Chemical Theory and Modelling (CTM) research group is the use of AI accelerated computational chemistry to explore chemical reaction spaces, e.g., with the aim to discover new performant catalysts, and analyze/engineer complex reaction networks, including plausible prebiotic (auto-)catalytic cycles that could shed light on the origin of life.

In recent years, we have contributed both application-driven and methodological advances in this area. For instance, we have developed a high-throughput screening protocol to identify bioorthogonal click reactions from a chemical space exceeding 10 million possibilities.¹ Additionally, we have created TS-tools, a software package for the automated generation of diverse reaction profiles for unknown reactions.²

However, a significant limitation in our current approaches is the lack of an accurate and efficient description of solvent effects, particularly for reactions occurring in polar environments. This hinders our ability to extend high-throughput reaction screening methods to many biologically and industrially relevant processes. **The goal of this PhD project is to leverage machine learning interatomic potentials (MLIPs) to enable large-scale reactivity exploration in solution.**

MLIPs are neural network-based models trained to predict energies – and atomic forces – based on molecular geometries.³ When trained on high-quality Density Functional Theory (DFT) data, these models can simulate (the dynamics associated with) a complete reaction path, including explicit solvation effects, at a fraction of the computational cost of full quantum chemical simulations. However, obtaining representative training data for reactive events remains a challenge, as conventional molecular dynamics (MD) simulations often fail to sample rare reaction events effectively.

To overcome this, most researchers employ enhanced sampling MD simulations at the DFT level, combined with active learning, to generate training data.⁴ While effective, these approaches require pre-defining reaction coordinates, inherently biasing the training data generated, and hence also the resulting MLIP. This PhD project will firstly aim to develop MLIPs for reactions in solution with minimal pre-imposed mechanistic assumptions, by training them on snapshots from many, diverse reaction pathways, generated by TS-tools, in combination with our in-house reaction pathway enumeration software (currently under development).

More specifically, the methodology will involve:

1. Generating diverse reaction pathways, with a range of intermediate geometries or snapshots along them, for a given molecular system using TS-tools and our pathway enumerator.
2. Solvating all generated intermediate geometries/snapshots in an automated manner with explicit solvent clusters.
3. Training an MLIP on these diverse solvated geometries, leveraging the approach pioneered by Fernanda Duarte and co-workers, who demonstrated that MLIPs trained on cluster models of water are transferable to simulations in bulk solution.⁵

Subsequently, we will also aim to integrate the TS-tools approach into the active learning-based refinement of the developed MLIPs, i.e., the sampling of additional snapshots in regions of the PES where the initial MLIP is uncertain about its predictions.

In a final part of the project, the impact of transfer learning, either by finetuning an existing general purpose MLIP, such as ANI,⁶ or by transferring an in-house developed MLIP from its original reactive system to a new one, on data-efficiency and generalizability, will be considered.

Overall, our aim is to rapidly develop generalizable, unbiased MLIPs capable of mechanistic discovery without imposing (strong) human preconceptions in the training data. The resulting models could potentially transform high-throughput reaction exploration in aqueous environments, with applications spanning catalysis, prebiotic chemistry, and beyond.

References

1. Stuyver, T.; Coley, C. W. Machine Learning-Guided Computational Screening of New Candidate Reactions with High Bioorthogonal Click Potential. *Chem.—Eur. J.* **2023**, *29*, e202300387.
2. Stuyver, T. TS-Tools: Rapid and Automated Localization of Transition States Based on a Textual Reaction SMILES Input. *J. Comput. Chem.* **2024**, *45*, 2308–2317.
3. Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
4. David, R.; de la Puente, M.; Gomez, A.; Anton, O.; Stirnemann, G.; Laage, D. ArcaNN: Automated Enhanced Sampling Generation of Training Sets for Chemically Reactive Machine Learning Interatomic Potentials. *Digit. Discov.* **2025**, *4*, 54–72.
5. Zhang, H.; Juraskova, V.; Duarte, F. Modelling Chemical Processes in Explicit Solvents with Machine Learning Potentials. *Nat. Commun.* **2024**, *15*, 6114.
6. Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.

Eligibility and Selection Criteria

Candidates will be evaluated based on:

- Academic excellence
- Relevance of their background to the research topic

The selection process follows an open, transparent, and merit-based (OTM) recruitment procedure.

Non-discrimination, openness, and transparency:

All PR[AI]RIE-PSAI partners are committed to supporting and promoting equality, diversity, and inclusion within their communities. We encourage applications from diverse backgrounds and ensure a fair selection process.

Application Requirements

Applicants must submit the following documents:

1. Curriculum Vitae (CV)
 2. Motivation Letter (max. 1 page) describing:
 - Your interest in the research topic
 - How your background aligns with the project
 3. Copy of your most recent diplomas
-

Application Procedure

- Deadline: May 20, 2025, at 17h00
- Applications should be submitted to: thijs.stuyver@chimieparistech.psl.eu
- The evaluation process consists of two phases:
 1. **Pre-selection** by the supervisor.
 2. **Final selection** by an expert committee, evaluating applications based on excellence and alignment with PR[AI]RIE-PSAI's scientific program.

Final results will be communicated by **June 15, 2025**.

For additional information, please visit <https://thijsstuyver.com> or contact thijs.stuyver@chimieparistech.psl.eu.