INTERNSHIP PROPOSAL

Laboratory name: Sorbonne Université - IMPMC
CNRS identification code: UMR 7590
Internship director'surname: Pietrucci
e-mail: fabio.pietrucci@sorbonne-universite.fr Phone number: 01 44 27 52 30
Web page: https://sites.google.com/site/fabiopietrucci/home
Internship location: Sorbonne Université, 4 Place Jussieu, 75005 Paris
Thesis possibility after internship: YES
Funding: YES (pending a doctoral school exam) type of funding: doctoral school ED397

New approaches to biomolecular kinetics based on machine-learned reaction coordinates

Computer simulations coupled with theoretical methods from statistical mechanics are a key tool to predict and explain experimental results. In the last 20 years, theory made important steps towards a deeper understanding of how matter transforms: piece by piece, **a general framework is emerging**, able to explain phase transitions, chemical reactions, and the fascinating operations of biomolecular machines. The basic ingredients are **molecular dynamics** (MD) simulations and **stochastic processes** (Langevin and Fokker-Planck equations) coupled with **out-of-equilibrium thermodynamics**.

Many **transformation processes of complex systems** can be studied with realistic MD simulations – using models of ~10^5 atoms – that provide a data set to train stochastic low-dimensional models (Langevin equations), based on diffusion on a free-energy landscape. Our research group at Sorbonne University recently invented new efficient computational approaches based on rigorous mathematical results, completing several pilot projects (see https://sites.google.com/site/fabiopietrucci/Home/papers).

The internship will develop **new ways to access important physical information**: free-energy barriers and kinetic rates will be estimated using non-conventional methods, and new types of loss function – the key quantity in machine learning optimizations – will be explored, based on physical variational principles, to identify optimal reaction coordinates based on scientific criteria and not solely on data science. We will address these two main tasks:

1) Estimate accurate kinetic rates for the dissociation of the barnase-barstar protein complex in water, based on a theoretical correction of metadynamics (where kinetics is perturbed by biasing forces, to accelerate rare transitions). We will start from the method in J. Phys. Chem. Lett. 13, 7490 (2022), to predict processes that take days from microsecond simulations.

2) Machine-learn optimal reaction coordinates starting from hundreds of microsecond-long molecular dynamics trajectories in a high dimensional space (generated by a previous PhD student, see https://doi.org/10.26434/chemrxiv-2025-cc60b-v2): in the case of the barnase-barstar complex, this will allow to clarify the role of electrostatic, hydrophobic and water-mediated interactions between proteins. Differently to common methods in the literature, we will use an original loss-function based on Langevin kinetics: this has crucial advantages, providing a rigorous mathematical foundation to the machine-learning tools.

The internship could be the first step of a PhD thesis, in collaboration with world leaders in the field at the universities of Amsterdam and Vienna, and at SISSA in Trieste.

Please, indicate which speciality(ies) seem(s) to be more adapted to the subject:

Condensed Matter Physics:	YES	Soft Matter and Biological Physics:	YES
Quantum Physics:	NO	Theoretical Physics:	YES