Frédéric Cazals
Centre Inria at Université Côte d'Azur
Sophia-Antipolis
http://team.inria.fr/abs
Frederic.Cazals@inria.fr

Tony Lelièvre
CERMICS
Ecole des Ponts et Chaussées, Marne la Vallée
http://cermics.enpc.fr/~lelievre/
Tony.Lelievre@enpc.fr

Master internship proposal

On the design of Markov kernels/move sets for free energy calculations

Keywords: biomolecules, conformational sampling, thermodynamics, collective variables, Monte Carlo Markov Chains.

Context. In studying biomolecules, one distinguishes three complementary aspects. The *structure* of a molecular system requires the characterization of (meta-)stable conformations and important transition states; by assigning occupation probabilities to these conformations through integration of the Boltzmann's distribution, one treats *thermodynamics*; finally, transitions between the states, modeled, say, by a master equation (a continuous-time Markov process), correspond to *kinetics*. `AlphaFold` [1], the structure prediction program developed by Deepmind, made a very significant step for structural aspects of proteins with little/no disordered regions [2], and its two leaders were co-awarded the 2024 Nobel prize in chemistry. But the question of fast and accurate predictions of thermodynamics and kinetics remains open [3].

Goals. The goal of this internship is to make a stride towards more effective algorithms for thermodynamics, using two main ingredients. The first one is a recently developed algorithm sampling metastable systems using so-called collective variables and Jarzynski-Crooks paths [4]. This algorithm – referred to as the JC algorithm thereafter, is in particular able to provide an estimate of a potential of mean force (PMF). The second one is a sampling strategy exploiting inverse problems using internal coordinates (torsion angles) to generate large amplitude conformational changes of flexible regions in proteins [5, 6], orders of magnitude faster than move sets provided by molecular dynamics.

In a nutshell, the strategy will consist in developing Markov kernels–also called *move sets* in molecular modeling–for the JC algorithm, using inverse problems and torsion angles. The overall algorithm will be implemented in C++. The ability to handle polypeptide chains up to one hundred residues would be a very significant achievement, with far reaching applications perspectives in biology and medicine.

Conditions. Internship with *gratification*.

Location: Warmup at CERMICS, then Sophia-Antipolis with visits at CERMICS.

This internship may be followed by a PhD thesis.

References

[1] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[2] F. Cazals and E. Sarti. Alphafold predictions on whole genomes at a glance. *Submitted*, 2024. https://biorxiv.org/cgi/content/short/2024.11.16.623929v1.

[3] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[4] C. Schönle, m. Gabrié, T. Lelièvre, and G. Stoltz. Sampling metastable systems using collective variables jarzynski–crooks paths. *preprint*, 2024.

[5] T. O'Donnell, V. Agashe, and F. Cazals. Geometric constraints within tripeptides and the existence of tripeptide reconstructions. *J. Comp. Chem.*, 44(13):1236–1249, 2023.

[6] T. O'Donnell and F. Cazals. Enhanced conformational exploration of protein loops using a global parameterization of the backbone geometry. *J. Comp. Chem.*, 44(11):1094–1104, 2023.