Inria Sophia Antipolis
Group: Algorithms-Biology-Structure
Frederic.Cazals@inria.fr
Web: `http://team.inria.fr/abs`

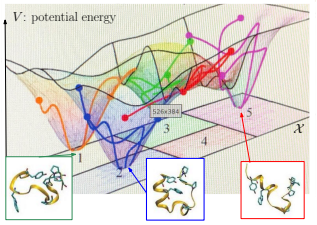

Figure 1: Exploring and mapping energy landscapes

Master internship proposal

Trekking and mapping high-dimensional landscapes

Keywords. High-dimensional spaces, exploration, random walks, probabilistic algorithms, energy landscapes, proteins.

Context. Macroscopic properties of biomolecules result from average values computed over ensembles of conformations [1, 2]. More precisely, consider the mapping associating a potential energy to each conformation, the so-called *potential energy landscape* (PEL) (Fig. 1). The PEL encodes all properties. The *structure* of a macromolecular system requires characterizing conformations associated with significant (deep and/or wide) basins of local minima. In assigning occupation probabilities to these basins by integrating Boltzmann's distribution, one treats *thermodynamics*. Finally, transitions between such basins correspond to *kinetics*. Since each atom has 3 Cartesian coordinates, a shear difficulty to study PELs is their huge dimensionality, namely $d = 3n (\gg 1000)$. Fortunately, PEL exhibit a remarkable hierarchical structure [3, 4].

Goals. Algorithms exploring PEL are probabilistic algorithms sampling conformations and connecting them to build a graph-based representation. We have shown that standard methods such as basin-hopping [5], or `T-RRT` [6] follow this *sample-select-connect* template [7]. Unfortunately, nothing is known on the ability of such algorithms to discover all significant basins, letting alone their complexity.

The goal of this master will be to investigate novel ideas to analyse these exploration algorithms, with feedback on their design. The starting point will be a set of novel analysis for sampling algorithms in geometric motion planning [8, 9], carried out in the framework of geometric random graphs [10], as well as novel insights on the multiscale structure of PEL.

Conditions. Internship with *gratification*. Possibility to follow-up with a PhD thesis.

References

[1] D. J. Wales. *Energy Landscapes*. Cambridge University Press, 2003.

[2] D.M. Zuckerman. *Statistical Physics of Biomolecules: An Introduction*. CRC Press, 2010.

[3] F. Cazals, T. Dreyfus, D. Mazauric, A. Roth, and C.H. Robert. Conformational ensembles and sampled energy landscapes: Analysis and comparison. *J. Comp. Chem.*, 36(16):1213–1231, 2015.

[4] J. Carr, D. Mazauric, F. Cazals, and D. J. Wales. Energy landscapes and persistent minima. *The Journal of Chemical Physics*, 144(5):4, 2016.

[5] Z. Li and H.A. Scheraga. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *PNAS*, 84(19):6611–6615, 1987.

[6] L. Jaillet, F.J. Corcho, J-J. Pérez, and J. Cortés. Randomized tree construction algorithm to explore energy landscapes. *J. Comp. Chem.*, 32(16):3464–3474, 2011.

[7] A. Roth, T. Dreyfus, C.H. Robert, and F. Cazals. Hybridizing rapidly growing random trees and basin hopping yields an improved exploration of energy landscapes. *J. Comp. Chem.*, 37(8):739–752, 2016.

[8] S. Karaman and E. Frazzoli. Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research*, 30(7):846–894, 2011. from Danny Halperin.

[9] Kiril Solovey and Michal Kleinbort. The critical radius in sampling-based motion planning. *The International Journal of Robotics Research*, 39(2-3):266–285, 2020.

[10] M. Penrose. Geometric random graphs, 2003.