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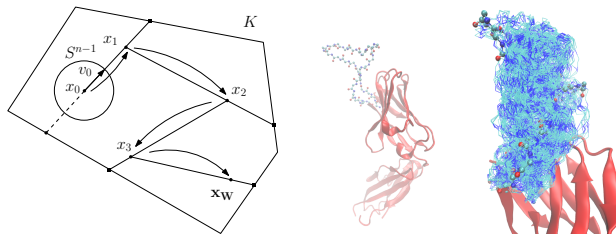Web: http://team.inria.fr/abs

MASTER INTERNSHIP PROPOSAL

Figure 1: (Left) Hit-and-Run in a 2D polygon [1]. (Middle, right) Conformations of a long flexible protein loop [2, 3].

Sampling algorithms in high-dimensional spaces coding the geometry of proteins

Keywords: sampling algorithms, Monte Carlo Markov Chain, volume calculations, statistical physics, conformational spaces, protein structure analysis.

Context. A well studied problem in geometry is the calculation of the volume of a high-dimensional polytope. While this problem is NP-hard, probabilistic algorithms based on Monte Carlo Markov Chain techniques and random walks of the Hit-and-Run type have been developed [1]. Currently, state-of-the-art random walks compute volumes in hundreds of dimensions within minutes on a laptop [2, 3].

In structural biology, a central problem is the prediction of the function of a protein, which requires understanding its dynamics–see e.g. AI, molecular design and the Covid19. A key achievement in protein structure analysis is the `AlphaFold` by Deepmind, which predicts a plausible structure from the amino acid sequence [4]. However, a single structure falls short from providing insights in the function of a protein. (For a metaphor, consider getting a single picture for an entire movie.) The generation of exhaustive molecular conformations is therefore a major challenge, and we recently contributed state-of-the-art algorithms for flexible loops, based on Hit-and-Run like algorithms working in high dimensional curved spaces coding the geometry of proteins [5, 6].

Goals. The work envisioned encompasses the analysis of our sampling algorithms, and their extension to compute provably correct averages in the so-called NVT ensemble. The analysis are likely to suggest modifications and improvements, which will also be implemented in C++ in the scope of the Structural Bioinformatics Library, an advanced software environment providing both low level algorithmic classes, and applications for end-users.

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

References

[1] B. Cousins and S. Vempala. A practical volume algorithm. *Mathematical Programming Computation*, 8(2):133–160, 2016.

[2] A. Chevallier, S. Pion, and F. Cazals. Improved polytope volume calculations based on Hamiltonian Monte Carlo with boundary reflections and sweet arithmetics. *J. of Computational Geometry*, 13(1):55–88, 2022.

[3] A. Chevallier, F. Cazals, and P. Fearnhead. Efficient computation of the the volume of a polytope in high-dimensions using piecewise deterministic markov processes. In *AISTATS*, 2022.

[4] 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.

[5] T. O'Donnell, V. Agashe, and F. Cazals. Geometric constraints within tripeptides and the existence of tripeptide reconstructions. Technical report, 2022.

[6] T. O'Donnell and F. Cazals. Enhanced conformational exploration of protein loops using a global parameterization of the backbone geometry. Technical report, 2022.