

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



Twenty conformations of a protein, showing a well defined core and disordered tails. Nobel lecture of K. Wütrich, 2002.

PHD THESIS PROPOSAL

GEOMETRIC ANALYSIS OF HIGH DIMENSIONAL POTENTIAL ENERGY SURFACES,  
WITH APPLICATIONS TO BIOMOLECULAR SIMULATION

**Keywords:** high-dimensional spaces, dimensionality reduction, importance sampling, molecular simulation.

**Context:** All biological functions (cognition, immune response, metabolism, etc) rely on biomolecules. The functions of these biomolecules depend on their structure but also dynamics, requiring the identification of meta-stable states (i.e. states stable on long time scales) together with transitions between these. As of today, except in rare cases where massive molecular dynamics or Monte Carlo simulations are used [1], these time scales remain out of reach.

The goal of this thesis will be to develop novel methods, ideally orders of magnitude faster than current ones, delivering accurate information on complex multi-scale mechanisms.

**Goals:** Since each atom has three Cartesian coordinates, the conformation of a molecule with  $n$  atoms is described by  $3n$  coordinates and  $d = 3n - 6$  degrees of freedom or dof. (One removes the dof for 3D translations and rotations, whence  $3n - 6$ .) All properties of the molecule are thus determined by an energy surface of dimension  $d$ , typically several thousands, called the potential energy surface (PES). Despite intensive research, the exploration and the characterization of such surfaces is currently an open problem.

From a computer science / applied mathematics standpoint, the problems faced as well posed, though. From the structural standpoint, stable structures correspond to local minima of the PES. From the thermodynamic viewpoint, meta-stability is characterized by occupancy probabilities (for Boltzmann's distribution) of selected basins of the PES. Finally, dynamics may be modeled by a Markov state model involving meta-stable states.

The derivation of such properties and models intrinsically requires a dimensionality reduction step, as average properties are best inferred using collective variables providing an abstraction of the aforementioned degrees of freedom. The goal of the thesis will be to develop novel methods delivering accurate models for biomolecules, by combining dimensionality reduction via diffusion maps [2, 3, 4, 5], importance sampling [6, 7], as well as geometric/topological techniques to explore and characterize PES [8, 9, 10].

Software developments will be integrated to the Structural Bioinformatics Library (<http://sbl.inria.fr>), a state-of-the-art environment providing both low level methods (in generic C++) and specific applications in molecular modeling.

**Background.** Master in theoretical computer science, or applied mathematics, or bioinformatics/biophysics.

## References

- [1] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.
- [2] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, and S.W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, 102(21):7426–7431, 2005.
- [3] M. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi. Determination of reaction coordinates via locally scaled diffusion map. *J. of Chemical Physics*, 134(12), 2011.

- [4] L. Nedialkova, M. Amat, I. Kevrekidis, and G. Hummer. Diffusion maps, clustering and fuzzy markov modeling in peptide folding transitions. *The Journal of chemical physics*, 141(11):09B611-1, 2014.
- [5] E. Chiavazzo, R. Covino, R. Coifman, C.W. Gear, A. Georgiou, G. Hummer, and I. Kevrekidis. Intrinsic map dynamics exploration for uncharted effective free-energy landscapes. *Proceedings of the National Academy of Sciences*, 114(28):E5494–E5503, 2017.
- [6] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. World Scientific, 2010.
- [7] A. Chevallier and F. Cazals. Wang-landau algorithm: an adapted random walk to boost convergence. *NA*, 2018.
- [8] 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.
- [9] J. Carr, D. Mazauric, F. Cazals, and D. J. Wales. Energy landscapes and persistent minima. *The Journal of Chemical Physics*, 144(5):4, 2016.
- [10] 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.