

1. PHD PROJECT DESCRIPTION

Project title: **Machine learning in all-atom simulations**

1.1. Project goals

- To find out the most suitable machine learning (ML) methodology to select slow degrees of freedom used in characterization of atomistic systems
- To test selected ML methods in atomistic model systems
- To solve urgent biophysical problems using enhanced sampling and importance sampling in atomistic simulations

1.2. Outline

Computational physics offers tremendous progress in understanding nature at the atomistic level. Dynamical processes involving biomolecules determine our life and health. Detailed studies of processes governing, for example, metabolism in cells require efficient computational tools. Current methods of all-atom, or even coarse-grained simulations of biochemical reactions or large scale conformational transitions of macromolecules suffer from very short time scales accessible for simulations. The mismatch between computational and experimental regimes is as large as between nanoseconds and second, so spans through some 9-10 orders of magnitude. Fortunately, enhanced sampling methods allow for shrinking of this gap. In this project, we will seek for new methods for extraction of meaningful data from biased simulations of atomistic systems. Classical and quantum molecular dynamics, or Monte Carlo methods will be used.

Within this PhD project new machine learning (ML) methods will be applied to improve selection of collective variables reducing dimensionality of a given biophysical problem (Rydzewski et al., 2016, 2017, 2019). Advanced statistical and numerical methods will be used to implement open-source codes for the extraction of slow modes in atomistic dynamics. Variationally enhanced sampling (Valsson and Parrinello, 2014) method will be further expanded, tested, and optimized. New functionalities will be added to the widely used software PLUMED2 (Tribello et al., 2014). The research group has state-of-the art local computer facilities, supercomputers will be used when necessary. We plan to use new methods in studies of some current protein systems related to medical problems. These efforts will bring new, fast and improved tools that will allow for the effective usage of computational simulations in physics, chemistry and material science.

1.3. Work plan

- Mastering molecular dynamics, Monte Carlo and statistical physics.
- Finding general schemes for ML based selection of slow modes.
- Testing new ML approaches on given biological problems: finding ligand unbinding pathways in protein channels, reduction of noise in molecular dynamics trajectories, reaching millisecond timescales of large conformational transitions in allostery.

1.4. Literature

1. Rydzewski, J., Jakubowski, R., Nowak, W., & Grubmüller, H. (2018) *J. Chem. Theory Comput.*, 14, 2843
2. Rydzewski, J., & Nowak, W. (2016) *J. Chem. Theory Comput.*, 12, 2110
3. Rydzewski, J., & Nowak, W. (2017a) *Sci. Rep.*, 7, 7736
4. Rydzewski, J., & Nowak, W. (2017b) *Phys. Life Rev.* 22, 58
5. Rydzewski, J., & Valsson, O. (2019) *J. Chem. Phys.* **150**, 220901 (2019) <https://doi.org/10.1063/1.5092590>
6. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014) *Comput. Phys. Commun.*, 185, 604
7. Valsson, O., & Parrinello, M. (2014) *Phys. Rev. Lett.*, 113, 090601

1.5. Required initial knowledge and skills of the PhD candidate

- High motivation to pursue research and get the PhD degree
- Analytical thinking
- Programming skills
- Curiosity
- Basic understanding of physics, chemistry and mathematics

1.6. Expected development of the PhD candidate's knowledge and skills

- Better understanding of advanced modelling methods used in computer physics and computational biophysics
- Practical knowledge of machine learning methods
- Advanced programming skills (Unix, Python, C++)
- “Fluency” in work in international scientific settings