1. PHD PROJECT DESCRIPTION (4000 characters max., including the aims and work plan)

Project title: Statistical Learning of Slow Collective Variables from Atomistic Simulations

1.1. Project goals

- Developing machine learning (ML) methodology to select slow collective variables used in the characterization of atomistic systems
- To test the methodology methods in atomistic model systems
- To solve selected biophysical problems involving ligand diffusion through protein and conformational changes in proteins
matrices

1.2. Outline

This Ph.D. student position will be an appointment for the project entitled "Statistical Learning of Slow Collective Variables from Atomistic Simulations," led by Dr. Jakub Rydzewski at the Institute of Physics, Nicolaus Copernicus University in Toruń, Poland. The project is financed by National Science Center (NCN Sonata).

Modeling the long-timescale dynamics of complex systems is a fundamental task in the physical sciences. Molecular dynamics (MD) simulations allow to probe the spatiotemporal details of molecular processes, but the so-called sampling problem severely limits their usefulness in practice.

One way to alleviate the sampling problem is to employ enhanced sampling simulations in which fluctuations of a few degrees of freedom, called collective variables (CVs), are boosted. Finding CVs that quantify the essential characteristics of a rare event may not be trivial. In this project, we consider crucial problems related to estimating CVs for complex physical systems: How to construct the CVs without resorting to system-specific expert knowledge? Is it possible to construct the slow CVs directly from

enhanced sampling simulations?
We will devise a tool that can learn slow CVs in a near-blind manner, making it accessible to many users without detailed knowledge about enhanced sampling theory. We expect the method will significantly impact the MD community's current state and apply to long-timescale processes in chemistry, physics, and biology.

1.3. Work plan

1. Mastering molecular dynamics, Monte Carlo and statistical physics.
2. Finding general schemes for ML based selection of slow modes.
3. Optimizing computational efficiency of the method.
4. 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

Rydzewski, J., Jakubowski, R., Nowak, W., & Grubmüller, H. (2018) J. Chem. Theory Comput., 14, 2843
Rydzewski, J., & Nowak, W. (2016) J. Chem. Theory Comput., 12, 2110
Rydzewski, J., & Nowak, W. (2017a) Sci. Rep., 7, 7736
Rydzewski, J., & Nowak, W. (2017b) Phys. Life Rev. 22, 58
Rydzewski, J., & Valsson, O. (2019) J. Chem. Phys. 150, 220901 (2019)
Tribello, G. A.,et al. . (2014) Comput. Phys. Commun., 185, 604
Valsson, O., & Parrinello, M. (2014) Phys. Rev. Lett., 113, 090601

1.5. Required initial knowledge and skills of the PhD candidate

- Basic knowledge about molecular dynamics simulations
- Programming (e.g., C++, Python), Linux operating systems, and data analysis will be an advantage.

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

- Better understanding of advanced modeling 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