## 1. PHD PROJECT DESCRIPTION

**Project title:**

**Uncovering Hidden Collective Processes in Molecular Dynamics using Machine Learning**

### 1.1. Project goals

- To select a dynamical system (e.g., protein, DNA, material) characterised by the time scale separation responsible for slow collective processes.
- To study the selected system using advanced molecular dynamics (MD) methods.
- To select or construct the most suitable machine learning (ML) methodology to uncover hidden physical processes in dynamical systems (collective motion, ligand dissociation).

### 1.2. Outline

Computational physics, particularly molecular dynamics (MD), offers methods for tremendous progress in understanding nature at the atomistic level. Dynamical processes involving biomolecules determine our life and health. Detailed studies of physical processes governing, for example, metabolism in cells, require efficient computational tools. Current MD methods, for all-atom systems or even coarse-grained simulations, fo studying biochemical reactions or large scale conformational transitions of macromolecules, suffer frequently from very short time scales accessible for simulations (a problem known as time scale separation). 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, advanced MD methods allow to reduce this gap. In this project, we will seek new methods for extraction of meaningful data from such simulations of atomistic systems.

Within this PhD project machine learning (ML) methods will be applied to improve uncovering the hidden knowledge from and reducing dimensionality of a given biophysical problem (Rydzewski et al., 2016, 2017, 2019). An atomistic system exhibiting the time scale separation will be chosen based on the interests of student and his advisors. Advanced statistical and numerical methods will be used to implement open-source codes for the extraction of collective processes in atomistic dynamics. New functionalities will be added to the widely used software PLUMED2 (Tribello et al., 2014). Our research group has state-of-the art local computer facilities, supercomputers will be used when necessary. We will offer every means necessary for scientific growth and frequent collaborations with experts in the field of advanced simulation methods. Hopefully, 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

- Introduction to molecular dynamics, statistical physics, and scientific programming.
- Expanding knowledge about advanced simulation and machine learning methods.
- Testing ML on given biological problems: finding ligand unbinding pathways in protein channels, reaching millisecond time scales of large conformational transitions important in, e.g., allostery.
- Proposing a new method appropriate for studying physical process with time scale separation.

- Implementation in the PLUMED2 software.

### 1.4. Literature


1. Multiscale reweighted stochastic embedding: Deep learning of collective variables for enhanced sampling, J Rydzewski, O Valsson, The Journal of Physical Chemistry A 125 (28), 6286-6302
2. Rydzewski, J., Jakubowski, R., Nowak, W., & Grubmüller, H. (2018) J. Chem. Theory Comput., 14, 2843
3. Rydzewski, J., & Nowak, W. (2016) J. Chem. Theory Comput., 12, 2110
4. Rydzewski, J., & Nowak, W. (2017a) Sci. Rep., 7, 7736.
5. Rydzewski, J., & Nowak, W. (2017b) Phys. Life Rev. 22, 58.
6. Rydzewski, J., & Valsson, O. (2019) J. Chem. Phys. **150**, 220901 (2019).
7. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014) Comput. Phys. Commun., 185, 604


### 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
- Basic understanding of physics 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