pKaMD: Development of a fast AI-based method that includes conformational sampling in the predictions of protein pK_a values

Place of work: BioISI-FCUL (C8, 8.5.50D)

Supervisors: Miguel Machuqueiro (machuque@ciencias.ulisboa.pt);
Pedro Reis, Bayer AG, Berlin, DE (pedro.reis@bayer.com)

The importance of pH in the structure and function of proteins is illustrated by the fact that about a quarter of amino acid residues contain ionizable side chains. The pH-dependent ionizations that arise can generate strong electrostatic interactions that inevitably influence the stability of proteins and their ability to bind to other molecules, for example to substrates. Many research groups have invested in the development and improvement of computational methods capable of modeling the protonation equilibrium and predicting the pK_a values of proteins. One of the greatest difficulties encountered is related to the great conformational variability of proteins, leading to a very strong coupling between protonation and conformation. To deal with this difficulty, constant-pH molecular dynamics (CpHMD) methods were developed [1-2]. Recently, with advances in computational power, this type of methodology has become relatively accessible and promises to become state-of-the-art in terms of pK_a predictions in proteins. Furthermore, machine learning methods have been shown to accelerate pK_a calculations by a factor of up to three orders of magnitude [3]. Nonetheless, the introduction of conformational sampling produces a significant computational penalty. The aim of the MSc. thesis work is the development of a new pK_a calculation method based on short simulations of the current AI-accelerated CpHMD method. Several tasks will be carried out during the period of the work plan:

1. Adapt the CpHMD method for short runs (optimizing its parameters) aiming at both efficiency and accuracy of the pK_a predictions;
2. Apply the method and benchmark with a set of proteins for which there is abundant experimental pK_a data [4].
3. Implement an online UI and merge it with the current infrastructure (pypka.org) and its backend servers to make this service available to the scientific community.
4. Compile the results into an MSc. thesis and write a scientific article to be published in an international journal.

[1] Machuqueiro, M., Baptista, A. M. (**2006**) *J. Phys. Chem. B*, 110, 2927.
[2] Vila-Viçosa, D., Reis, P.B.P.S., Baptista, A.M., Oostenbrink, C., Machuqueiro, M. (**2019**) *J. Chem. Theory Comput.*, 15, 3108-3116.
[3] Reis, P.B.P.S, Bertolini, M., Montanari, F., Rocchia, W., Machuqueiro, M., Clevert, DA. (**2022**) J. Chem. Theory Comput., 18, 8, 5068–5078
[4] Pahari, S., Sun, L., Alexov, E. (**2019**) *Database*, baz024.