# Accelerating binding free energy calculations by combining Monte Carlo simulations, enhanced sampling and Markov State Models

Joan F. Gilabert*, Victor Guallar*†
*Barcelona Supercomputing Center, Barcelona, Spain
†Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
E-mail: {joan.gilabert, victor.guallar}@bsc.es

***Keywords—AdaptivePELE, PELE, Markov State Models, free energy calculations.***

## I. Extended Abstract

Computational approaches to the estimation of binding affinities have received a great deal of attention for their potential impact in the drug discovery process. Despite notorious advances and positive results, the current methods still present essential limitations, mainly the difficulty of obtaining thermodynamic sampling in highly-dimensional systems such as protein-ligand energy landscapes.

Biomolecular simulations of atomistic resolution are typically separated into two groups, molecular dynamics (MD) and Monte Carlo (MC) algorithms. Both techniques rely on a force field, a molecular mechanics model of the biomolecule that takes into account many possible interactions, such as electrostatic, bonded or Van Der Waals forces. MD integrates Newton's equations of motion for each atom to obtain the time evolution of the system, while MC techniques apply random movements to sample the energy landscape. Theoretically, MC is expected to generate a higher variety in the obtained conformations, however, the difficulty in generating uncorrelated structures of proteins makes this theoretical expected advantage vanish.

In our group, we developed the Protein Energy Landscape Exploration (PELE)[1], a method that combines Monte Carlo sampling with protein structure prediction techniques to attempt to retain some sampling advantage. Nevertheless, such trajectories exhibit metastability, due to the ruggedness of the energy landscape. To overcome this limitation we have developed an enhanced sampling method, called AdaptivePELE[2]. This method performs several rounds of simulations, combined with clustering and reinforcement learning techniques.

The use of AdaptivePELE drastically improves the efficiency of our simulations, the same reason for this speed-up becomes a hindrance when estimating thermodynamic properties. Enhanced sampling methods introduce a bias and distort the sampled landscape, thus special care has to be taken when performing thermodynamic calculation. Several works have been introduced in this direction[3][4][5], but despite showing great promise, the routine application of such methods is still far from our reach.

Due to the elevated dimensionality of the systems under consideration, many analysis techniques used for molecular simulations are based on some kind of dimensionality reduction, among such methods the use of Markov State Models[6] (MSM) has seen a swift increase in recent years, due to important advances in both its theoretical formulation and its usability. MSMs allow for the extraction of equilibrium properties from simulations of moderate lengths, being particularly suited for the analysis of simulations run in parallel computing setups.

Our work is based on the combination of the three techniques presented here: AdaptivePELE, PELE and MSM. We start with an AdaptivePELE simulation, to quickly map the system landscape, we then cluster the results of the simulation to obtain a few representative structures (approximately 40) to run a longer standard PELE simulation, which will provide a more through sampling. From the PELE simulation we build an MSM, which will give us an estimation of the probability of each state. This probability is used to build a potential of mean force (PMF), $g$, according to

$$g_i = -k_b T \ln\left(\frac{\pi_i}{V_i}\right) \tag{1}$$

where $k_b$ is the Boltzmann constant, $T$ the temperature, $\pi_i$ is the probability of state $i$ and $V_i$ the volume of said state. From the PMF, the free energy is calculated as

$$\Delta G^o_{bind} = \Delta W - k_b T \ln\left(\frac{V_b}{V_o}\right) \tag{2}$$

where $\Delta G^o_{bind}$ is the binding free energy, $\Delta W$ is the depth of the PMF, $V_b$ is the binding volume and $V_o$ is the standard volume, defined in equations 3 and 4 respectively.

$$V_b = \sum_i V_i \exp^{-\frac{g_i}{k_b T}} \tag{3}$$

$$V_o = 1661 \text{ Å}^3 \tag{4}$$

The combination of the three methodologies allows us to obtain a quick estimation of absolute binding free energies for

different scaffolds, which will hopefully help accelerate the drug discovery process.

## REFERENCES


[1] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar, "PELE: Protein energy landscape exploration. A novel Monte Carlo based technique," *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1304–1311, 2005.

[2] D. Lecina, J. F. Gilabert, and V. Guallar, "Adaptive simulations, towards interactive protein-ligand modeling," *Scientific Reports*, vol. 7, no. 1, p. 8466, 2017. [Online]. Available: http://www.nature.com/articles/s41598-017-08445-5

[3] H. Wu, F. Paul, C. Wehmeyer, and F. Noé, "Multiensemble Markov models of molecular thermodynamics and kinetics," *Proceedings of the National Academy of Sciences*, vol. 113, no. 23, pp. E3221–E3230, 2016. [Online]. Available: http://www.pnas.org/lookup/doi/10.1073/pnas.1525092113

[4] N. V. Buchete and G. Hummer, "Peptide folding kinetics from replica exchange molecular dynamics," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 3, pp. 1–4, 2008.

[5] L. S. Stelzl, A. Kells, E. Rosta, and G. Hummer, "Dynamic Histogram Analysis To Determine Free Energies and Rates from Biased Simulations," *Journal of Chemical Theory and Computation*, p. acs.jctc.7b00373, 2017. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/acs.jctc.7b00373

[6] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," *Current Opinion in Structural Biology*, vol. 25, pp. 135–144, 2014.


**Joan F. Gilabert** received his BSc degree in Physics Engineering from the Universitat Politècnia de Catalunya (UPC), Barcelona in 2015. He completed his MSc degree in Bionformatics for Health Sciences from Universitat Pompeu Fabra (UPF), Barcelona in 2017. Currently, he is a PhD student of the Computational and Applied Physics of Universitat Politècnica de Catalunya with the Electronic and Atomic Protein Modelling group of Barcelona Supercomputing Center (BSC).