A Neural Network Approach for the Identification of Pathways in Molecular Dynamics Simulations of Ligand Binding

Motta Stefano[1], Lara Callea[1], Laura Bonati[1] Alessandro Pandini[2,3]

1 Department of Earth and Environmental Sciences, University of Milano-Bicocca, Milan, Italy.

2 Department of Computer Science, Brunel University London, Uxbridge, United Kingdom.

3 The Thomas Young Centre for Theory and Simulation of Materials, London, United Kingdom.

Motivation: Understanding the process of ligand-protein recognition is important to unveil biological mechanisms and to inform drug design and discovery. Enhanced sampling molecular dynamics (MD) methods can be used to simulate the ligand-binding process, obtaining information regarding its thermodynamics and kinetics (Limongelli V, 2020). The extensive sampling of the binding/unbinding events, through several simulation replicas or with a single simulation that describes several re-crossing events, creates the need for tools able to analyze a huge amount of data and to provide a clear picture of the sampled pathways.
Methods: We propose that binding events can be analysed using Self-Organizing Maps (SOMs), a type of artificial neural network useful for effective identification of patterns in the data (Kohonen T 2013). SOMs have already been used successfully for the analysis of MD simulations (Motta S et al. 2021). We designed, implemented and tested a SOM-based tool to build and analyze models of ligand binding pathways sampled along a simulation or in different replicas. The tool is written in R and can be used to perform a geometric clustering of the trajectories under study to obtain an overview of the conformations sampled during the simulation; to trace the sampled pathways to recover the temporal evolution of the system on the SOM; to build a network model that represents the pathways in a clear way using a transition matrix built from the SOM neurons.
Results: We tested the tool on different systems to demonstrate its general applicability and good performance. For this reason, the study-cases differ in the choice of MD methods used to investigate the ligand binding, including steered and metadynamic simulations of the THS-020 ligand binding to the HIF-2α protein (Callea L et al. 2021) and the infrequent metadynamic simulations of the GC7 ligand unbinding from human deoxyhypusine synthase (D'Agostino M et al. 2020). For both systems the tool was able not only to identify the preferred unbinding pathway but also to detect the position of the energetic barrier for the transition.

References

- Callea L, Bonati L, Motta S. Metadynamics-Based Approaches for Modeling the Hypoxia-Inducible Factor 2α Ligand Binding Process. Journal of Chemical Theory and Computation. 2021; 17; 3841–3851. https://doi.org/10.1021/acs.jctc.1c00114.
- D'Agostino M, Motta S, Romagnoli A, Orlando P, Tiano L, La Teana A, Di Marino D. Insights into the Binding Mechanism of GC7 to Deoxyhypusine Synthase in Sulfolobus Solfataricus: A Thermophilic Model for the Design of New Hypusination Inhibitors. Frontiers in Chemistry. 2020; 8, 1–14. https://doi.org/10.3389/fchem.2020.609942.
- Kohonen T. Essentials of the Self-Organizing Map. Neural Networks. 2013; 37; 52–65. https://doi.org/10.1016/j.neunet.2012.09.018.
- Limongelli V. Ligand Binding Free Energy and Kinetics Calculation in 2020. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2020; 10; 1–32. https://doi.org/10.1002/wcms.1455.
- Motta S, Pandini A, Fornili A, Bonati L. Reconstruction of ARNT PAS-B Unfolding Pathways by Steered Molecular Dynamics and Artificial Neural Networks. Journal of Chemical Theory and Computation. 2021; 17 2080–2089. https://doi.org/10.1021/acs.jctc.0c01308.