MELD-DNA: Accurate Structural Prediction of Protein-DNA Interactions

Reza Esmaeeli[1], Antonio Bauza Riera[2], Alberto Perez[1]

[1] Chemistry, University of Florida, Gainesville, FL, USA.
[2] Chemist, Universitat de les Illes Balears, Palma de Mallorca, Spain

Protein-DNA interaction is one of the most important aspects of biology. It is a main step of gene regulation in all organisms and is a potential target for therapeutic and biotech applications. Protein-DNA interactions are distinct from protein-protein interactions due to several factors: 1) DNA sequence and specific interactions (base readout); 2) DNA shape and its deformability (shape readout); 3) Exploration and searching on the DNA for alternative binding states. Such factors dictate affinity and specificity of transcription factors. Traditionally, these interactions were only predicted at the sequence level (e.g. using position weight matrices) and more recently using docking and machine learning models. Here we present MELD-DNA, a too that combines cheap information from experiment (which can be noisy, ambiguous and sparse) with physics-based molecular mechanics force fields using bayesian inference. MELD will accelerate sampling of the binding process by temperature and hamiltonian replica exchanging that is guided by experimental data. We have tested the tool on a diverse set of 15 systems from various TF families and we observed native-like bound structures most of them with room for more improvements. We also compare and benchmark with the state-of-the-art in docking (HADDOCK) and ML (RosettaFold2NA). In this study we discuss several issues in the available methods and techniques that need to be addressed for more accurate predictions. We believe MELD-DNA can be an essential tool in the community from drug design, regulation and fundamental research.