Juan Cortes
LAAS-CNRS

Title:
WASCO: A Wasserstein-based statistical tool to compare conformational ensembles of intrinsically disordered proteins

Abstract: The structural investigation of intrinsically disordered proteins (IDPs) requires ensemble models describing the diversity of the conformational states of the molecule. Due to their probabilistic nature, there is a need for new paradigms that understand and treat IDPs from a purely statistical point of view, considering their conformational ensembles as well-defined probability distributions. In this work, we define a conformational ensemble as an ordered set of probability distributions and provide a suitable metric to detect differences between two given ensembles at the residue level, both locally and globally. The underlying geometry of the conformational space is properly integrated, one ensemble being characterized by a set of probability distributions supported on the three-dimensional Euclidean space (for global-scale comparisons) and on the two-dimensional flat torus (for local-scale comparisons). The inherent uncertainty of the data is also taken into account to provide finer estimations of the differences between ensembles. Additionally, an overall distance between ensembles is defined from the differences at the residue level. We illustrate the interest of the approach with several examples of applications for the comparison of conformational ensembles: (i) produced from molecular dynamics (MD) simulations using different force fields, and (ii) before and after refinement with experimental data. We also show the usefulness of the method to assess the convergence of MD simulations, and discuss other potential applications such as in machine-learning-based approaches. The numerical tool has been implemented in Python through easy-to-use Jupyter Notebooks available at https://gitlab.laas.fr/moma/WASCO.