Combination of Molecular Dynamics and Machine Learning for the Prediction of Molecular Properties

Riniker, S.[1]

[1]Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

1. Abstract

From simple clustering techniques to sophisticated neural networks, the use of machine learning has become a valuable tool in many fields of chemistry in the past decades. Here, we describe the combination of machine learning (ML) and molecular dynamics (MD) simulations through MD fingerprints (MDFP). These descriptors encode the information in MD simulations such that it can be used as input to train ML models for the quantitative understanding of molecular systems. The concept of the MDFPs is highly versatile, depending on the property to be predicted, different systems can be simulated and different properties can be extracted from the MD simulations. We show the application for the prediction of solvation free energies, partition coefficients, activity coefficients, and vapour pressure, as well as the classification of substrates of the efflux transporter P-gp. We find that MDFP are information-rich descriptors, which contain orthogonal information of the molecules compared to topological fingerprints or physicochemical descriptors. This orthogonality can be exploited to further improve the prediction accuracy of ML models.