Resolution transformation in molecular dynamics: Boosting backmapping via knowledge-driven machine learning

Christian Pfaendner[1,2], Kristyna Pluhackova[1,2], Benjamin Unger[1,2]

[1]) University of Stuttgart, Stuttgart Center for Simulation Science, Germany
[2]) University of Stuttgart, Artificial Intelligence Software Academy, Germany
christian.pfaendner@simtech.uni-stuttgart.de

Molecular dynamics (MD) simulations provide valuable insights into biological macromolecules with temporal and spatial resolution not currently available to experimental methods. Thereby, MD has helped to uncover biochemical mechanisms important for understanding the development, progression, and treatment of serious diseases such as cancer. Despite advances in high-performance computing, system sizes and simulation times for atomistic simulations are still limited to the nanometer and microsecond scales. To overcome these limitations, sequential multiscale MD simulations switch between different levels of resolution of the molecular representation, allowing to study processes over longer timescales and simultaneously recover atomistic details. While coarse-graining is straightforward, the reverse transformation from low to high resolution, also termed backmapping, is a non-trivial task. This is mainly due to the loss of information about the atomistic structure underlying the coarse-grained representation, which must be reintroduced (see Figure 1). Available state-of-the-art methods neglect the Boltzmann distribution, get trapped in local minima, might fail to recover the correct stereochemistry, and often require extensive expert knowledge from the user.

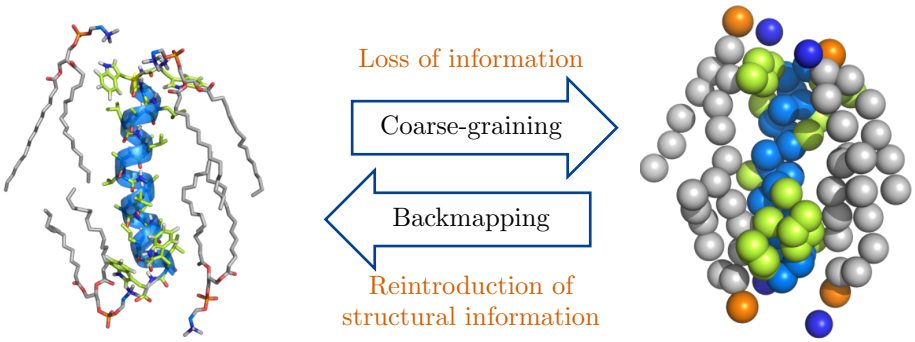

Figure 1: The same molecule is shown on the left in atomistic and on the right in coarse-grained representation. Switching between the different resolutions is possible via coarse-graining or backmapping, respectively.

We develop a novel chemistry-informed machine learning framework for the re-

verse transformation of coarse-grained MD data to atomistic resolution. Our algorithm is based on the Boltzmann distributions of structural features, thus avoiding the drawbacks of current methodologies, and shows how hybrid first-principle and certified artificial intelligence-based models can boost MD. Moreover, we provide standardized evaluation criteria to directly compare the performance of different backmapping algorithms, allowing users to choose an appropriate method for a given use case.