Symposium Data-driven Methods in Molecular Simulations of Soft-Matter Systems (SYMS)

jointly organized by
the Chemical and Polymer Physics Division (CPP),
the Biological Physics Division (BP),
the Dynamics and Statistical Physics Division (DY), and
the Metal and Material Physics Division (MM)

Christoph Dellago
University of Vienna
Sensengasse 8/9
1090 Vienna
christoph.dellago@univie.ac.at

Daan Frenkel
University of Cambridge
Lensfield Road
Cambridge CB2 1EW, UK
df246@cam.ac.uk

Kurt Kremer
MPI for Polymer Research
Ackermannweg 10
55128 Mainz
kremer@mpip-mainz.mpg.de

Multiscale simulations offer an appealing framework to approach many soft-matter systems. Recently, the advancement of data-driven methods, pushed forward by the statistics and computer-science communities, has started making significant impact in various fields of condensed-matter physics, however mostly hard condensed matter. In contrast to hard matter, for soft matter intra- and intermolecular molecular entropy is as relevant as are interaction energies. This symposium will present and discuss early developments of data-driven methods applied to soft matter. It will provide a forum to help illustrate how these methods can help solve or ease problems in the field, e.g., force-field parametrization and adequate conformational sampling.

Overview of Invited Talks and Sessions

(Lecture room H 0105)

Invited Talks

SYMS 1.1	Mon	15:00–15:30	H 0105	**Stochastic numerical algorithms: from molecular dynamics to big data analytics** — •Benedict Leimkuhler
SYMS 1.2	Mon	15:30–16:00	H 0105	**A Generally-Applicable Machine-Learning Scheme for Materials and Molecules** — •Michele Ceriotti
SYMS 1.3	Mon	16:00–16:30	H 0105	**Girsanov reweighting for path ensembles and Markov state models** — •Bettina G. Keller, Luca Donati, Carsten Hartmann
SYMS 1.4	Mon	16:45–17:15	H 0105	**Liquid State Theory Meets Deep Learning and Molecular Informatics** — •Alpha Lee
SYMS 1.5	Mon	17:15–17:45	H 0105	**Computational high-throuhgput screening of drug-membrane thermodynamics** — •Tristan Bereau

Sessions

SYMS 1.1–1.5	Mon	15:00–17:45	H 0105	**Data-driven Methods in Molecular Simulations of Soft-Matter Systems**

SYMS 1: Data-driven Methods in Molecular Simulations of Soft-Matter Systems

Time: Monday 15:00–17:45 Location: H 0105

Invited Talk SYMS 1.1 Mon 15:00 H 0105

Stochastic numerical algorithms: from molecular dynamics to big data analytics — •Benedict Leimkuhler — University of Edinburgh, Edinburgh, UK

I will discuss the interplay between the methods of molecular/particle simulation and algorithms for statistical inference, as used for example for inference of model properties and parameter selection to describe a large data set. On the one hand, I will describe some of the increasingly sophisticated methods available for sampling molecular conformations, including constrained stochastic methods and other schemes which make use of variable temperature. I will show that many of these methods have analogous applications in data science, where they can help in the discovery of robust parameterisations or to address challenges due to problem structure. When the data set itself is generated by molecular dynamics trajectory simulation, the resulting procedures offer prospects for coarse-graining and enhanced sampling of highly complex systems.

Invited Talk SYMS 1.2 Mon 15:30 H 0105

A Generally-Applicable Machine-Learning Scheme for Materials and Molecules — •Michele Ceriotti — Institute of Materials, EPFL, Lausanne, Switzerland

Determining the stability of molecules and condensed phases is the cornerstone of atomistic modelling, underpinning our understanding of chemical and materials properties and transformations. I will show that a machine-learning model, based on a local description of chemical environments and Bayesian statistical learning, provides a unified framework to predict atomic-scale properties. It captures the quantum mechanical effects governing the complex surface reconstructions of silicon, predicts the stability of different classes of molecules with chemical accuracy, and distinguishes active and inactive protein ligands with more than 99% reliability. The universality and the systematic nature of this framework provides new insight into the potential energy surface of materials and molecules. I will also discuss how the method can be extended to yield a "symmetry-adapted" Gaussian process regression approach that is capable of learning tensorial properties without the need of defining explicitly a local reference frame.

Invited Talk SYMS 1.3 Mon 16:00 H 0105

Girsanov reweighting for path ensembles and Markov state models — •Bettina G. Keller[1], Luca Donati[1], and Carsten Hartmann[2] — [1]Freie Universität Berlin, Germany — [2]Brandenburgische Technische Universität Cottbus-Senftenberg, Germany

Enhanced sampling techniques, such as metadynamics or umbrella sampling, in which a biasing potential U(x) is added to the unbiased force field V(x) increase the sampling of rare events. However, the distortion of the timescales in the system due to the biasing potential is not uniform. The resulting biased trajectories can hence not be used to estimate models of the molecular dynamics, e.g. Markov state models.

I will present the Girsanov reweighting method with which one can estimate the the expected path ensemble average of an unbiased dynamics for a set of biased paths. The method is based on the concept of path probability measure and the Girsanov theorem, a result from stochastic analysis to estimate a change of measure of a path ensemble. Since Markov state models of molecular dynamics can be formulated as a combined phase-space and path ensemble average, the method can be extended to reweight these models by combining it with a reweighting of the Boltzmann distribution. Besides its use in enhanced sampling simulations, the Girsanov reweighting can also be used to test the response of the slow dynamic processes to perturbations of the potential energy surface.

15 min. break

Invited Talk SYMS 1.4 Mon 16:45 H 0105

Liquid State Theory Meets Deep Learning and Molecular Informatics — •Alpha Lee — Department of Physics, University of Cambridge, Cambridge, United Kingdom

A large class of problems in machine learning pertains to making sense of high dimensional and unlabelled data. The challenge lies in separating direct variable-variable interactions (e.g. cause and effect) and transitive correlations, as well as removing noise due to insufficient number of samples relative to the number of variables. In this talk, I will discuss an Ornstein-Zernike-like approach for data analysis that disentangles correlations in datasets using ideas from the theory of liquids. The Ornstein-Zernike closure is parameterised by deep learning, and a framework inspired by random matrix theory is used to remove finite sampling noise. I will illustrate this approach by applying it to problems such as ligand-based virtual screening and predicting protein function from sequence covariation.

Invited Talk SYMS 1.5 Mon 17:15 H 0105

Computational high-throuhgput screening of drug-membrane thermodynamics — •Tristan Bereau — Max Planck Institute for Polymer Research, Mainz, Germany

The partitioning of small molecules in cell membranes—a key parameter for pharmaceutical applications—typically relies on experimentally-available bulk partitioning coefficients. Computer simulations provide a structural resolution of the insertion thermodynamics via the potential of mean force, but require significant sampling at the atomistic level. Here, we introduce high-throughput coarse-grained molecular dynamics simulations to screen thermodynamic properties. This application of physics-based models in a large-scale study of small molecules establishes linear relationships between partitioning coefficients and key features of the potential of mean force. This allows us to predict the structure of the insertion from bulk experimental measurements for more than 450,000 compounds. The potential of mean force hereby becomes an easily accessible quantity—already recognized for its high predictability of certain properties, such as passive permeation. Further, we demonstrate how coarse graining helps reduce the size of chemical space, enabling a hierarchical approach to screening small molecules.