4th SIMPLAIX Workshop on ”Machine Learning for Multiscale Molecular Modeling”

6-8 May 2026
Studio Villa Bosch Heidelberg, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg

Abstract Book

Scientific Organizing Committee: Alice Allen (MPIP Mainz), Marcus Elstner (KIT), Pascal Friederich (KIT), Jan Stühmer (HITS), Rebecca Wade (HITS), Frederic Teich (KIT), Maike Vogelbacher (KIT)

Local Organizing Committee: Darija Stein (HITS), Christina Blach (HITS), Angela Michel (HITS)

Organizers:

HITS
Heidelberg Institute for Theoretical Studies

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

KIT
Karlsruher Institut für Technologie

MAX-PLANCK-INSTITUT
FÜR POLYMERFORSCHUNG

Sponsors:

Klaus Tschira Stiftung
gemeinnützige GmbH

KTS

GRK
2450

Program

Wednesday, 6 May 2026

13:00 Registration

14:00 Opening & Welcome
Rebecca Wade (Heidelberg Institute for Theoretical Studies)

Session 1
Chair: Tristan Bereau (Heidelberg University)

14:15 Alexandre Tkatchenko (University of Luxembourg)
Machine-Learned Force Fields for Molecular Simulations Beyond AlphaFold and Empirical Potentials (talk)

14:55 Torben Berndt (Heidelberg Institute for Theoretical Studies)
Approximate Equivariance via Projection-Based Regularisation (short talk)

15:15 Van-Quan Vuong (Karlsruhe Institute of Technology)
Complete-Basis-Set Extrapolation: What Scheme Is the Best? (short talk)

15:45 Coffee break

Session 2
Chair: Rebecca Wade (Heidelberg Institute for Theoretical Studies)

16:30 Bingqing Cheng (University of California, Berkley)
Energy and forces are all you need (talk)

17:10 Qiang Cui (Boston University)
Recent developments and applications of QM/MM-ML free energy methods (talk)

17:50 Cesar Mendoza Martinez (University of Dundee)
Generation of open states for ion channels using AlphaFold subsampling and intrinsic protein properties (short talk)

18:15 Group photo (Speakers)

Thursday, 7 May 2026

Session 3
Chair: Pascal Friederich (Karlsruhe Institute of Technology)

09:00 Fernanda Duarte (University of Oxford)
Modelling condensed-phase reactivity with MLIPs (talk)

09:40 Luca Ghiringhelli (Karlsruhe Institute of Technology)
Learning parsimonious models for materials characterization and design (talk)

10:20 Niccolo Di Eugenio (Polytechnic University of Turin)
Machine-Learned Interatomic Potentials for Radiation Damage Analysis in High-Temperature Superconductors
(short talk)

10:40 Coffee Break

Session 4
Chair: Frauke Gräter (Max Planck Institute for Polymer Research)

11:20 Patrick Bryant (Stockholm University SciLifeLab)
Beyond the Canonical Alphabet: De Novo Design of Functional Peptides and GPCR Agonists (talk)

12:00 Chris John (University of Pisa)
Machine-Learning Framework for Excitation Energies of Chromophores in Polarizable Environments (short talk)

12:20 Discussion session
Chair: Fred Hamprecht (Heidelberg University)
"Agentic AI: How will it transform our life as researchers?"

13:00 Lunch

14:20 Group picture

Session 5
Chair: Jan Stühmer (Heidelberg Institute for Theoretical Studies)

14:30 Alexander Tong (University of Dundee)
Controlling Generative Models (talk)

15:10 Ben Murrell (University of Dundee)
Branching Flows: Discrete, Continuous, and Manifold Flow Matching with Splits and Deletions (talk)

15:50 Riccardo Beccaria (Heidelberg Institute for Theoretical Studies)
Physics Informed Machine Learning Potential for Brownian Dynamics Simulations (short talk)

16:10 Coffee break

Session 6

Chair: Alice Allen (Max Planck Institute for Polymer Research)

16:40 Yair Litman (Max Planck Institute for Polymer Research)
Molecular Strategies for Confinement-Controlled Water Structure and Chemistry (talk)

17:20 Roman Remme (Heidelberg University)
Surrogate Functionals for Machine-Learned Orbital-Free Density Functional Theory (short talk)

17:45 Poster session

19:00 Scientist evening session with finger food + poster session (Studio)

Friday, 8 May 2026

Session 7
Chair: Fred Hamprecht (Heidelberg University)

09:00 Ilyes Batatia (University of Cambridge)
Foundational Machine Learning Force Fields for Atomistic Chemistry (talk)

09:40 Pratyush Tiwary (University of Maryland)
Efficient, explainable and extrapolative AI for biophysics *(talk)*

10:20 Tobias Kaczun (Heidelberg University)
Taking the next Step with Machine Learning Orbital Free DFT (short talk)

10:40 Coffee Break

Session 8
Chair: Alice Allen (Max Planck Institute for Polymer Research)

11:10 Martin Stöhr (Stanford University)
Differentiable Quantum Chemistry: Hybrid Machine Learning Models and More (talk)

11:50 Henrik Christiansen (NEC Laboratories Europe)
Hyperspatial Sampling *(short talk)*

12:10 Round up

13:00 Lunch and end of workshop

Talks

Machine-Learned Force Fields for Molecular Simulations Beyond AlphaFold and Empirical Potentials

Alexandre Tkatchenko (alexandre.tkatchenko@uni.lu)
University of Luxembourg, Esch-sur-Alzette, Luxembourg

The convergence between accurate quantum-mechanical (QM) models (and codes) with efficient machine learning (ML) methods seem to promise a paradigm shift in all-atom simulations. Many challenging applications are now being tackled by increasingly powerful QM/ML methodologies[1,2]. These include modeling covalent materials, molecules, molecular crystals, surfaces, and even whole proteins under physiological conditions [3,4]. In this talk, I attempt to provide a reality check on these recent advances and on the developments required to enable fully predictive dynamics of complex functional (bio)molecular and material systems. Multiple challenges are highlighted -- in particular transferability in chemical space and interatomic interactions -- that should enable this field to grow for the foreseeable future.

---

[1] Unke, O. et al., Chem. Rev. 2021, 121, 16, 10142–10186.
[2] Keith, J. et al., Chem. Rev. 2021, 121, 16, 9816–9872.
[3] Unke, O. et al., Sci. Adv. 2024, 10, 14.
[4] Kabylda, A. et al., J. Am. Chem. Soc. 2025, 147, 37, 33723–33734.

Approximate Equivariance via Projection-Based Regularisation

Torben Berndt (torben.berndt@h-its.org)
Heidelberg Institute for Theoretical Studies, Germany

Equivariance is a useful inductive bias for neural networks, improving generalisation and physical consistency, but exact symmetries are often too constraining and fully equivariant models can be costly. This talk presents a projection-based regulariser for approximate equivariance that decomposes linear layers into equivariant and non-equivariant components and penalises non-equivariance at the operator level, avoiding sample-based augmentation. The framework enables exact and efficient computation of the penalty in both spatial and spectral domains, including for continuous groups such as SO(3).

Complete-Basis-Set Extrapolation: What Scheme Is the Best?

Van-Quan Vuong (vanquan.vuong@kit.edu)
Karlsruhe Institute of Technology, Germany

Complete-basis-set (CBS) extrapolation is a standard method for correcting basis-set errors in ab initio calculations to reach the basis-set limit. While numerous extrapolation schemes exist, identifying the best approach remains challenging. Recent benchmarks report near-identical performance across all schemes. In this work, we demonstrate that this apparent agreement is a mathematical artifact: at a fixed basis-set pair, single-parameter least-squares fits collapse all two-point formulae onto a single scheme-free scalar. To break this degeneracy and rigorously rank the methods, we introduce a strategy that fits a global parameter across multiple basis-set pairs simultaneously against a 183-species dataset. This approach identifies the constant-factor family (Feller-1992) as the optimal scheme for Hartree-Fock and Bakowies-2007 as the best for CCSD(T). Finally, we provide the reoptimized parameters and simplified per-pair scalars for these winning schemes, allowing straightforward application in practical workflows.

Energy and Forces are all you need

Bingqing Cheng (bingqingcheng@berkeley.edu)
University of California, Berkeley, United States

Standard machine learning interatomic potentials (MLIPs) often rely on short-range approximations, limiting their applicability to systems with significant electrostatics. We recently introduced the Latent Ewald Summation (LES) method, which learns long-range electrostatics from *just energy and force data*. We show that LES can effectively infer physical partial charges, polarization and Born effective charge (BEC) tensors, as well as achieve better accuracy compared to methods that explicitly learn charges. As demonstrations, we predict the infrared spectra of bulk water under zero or finite external electric fields, ionic conductivities of high-pressure superionic ice, and the phase transition and hysteresis in ferroelectric PbTiO3 perovskite.

Recent Developments and Applications of QM/MM-ML Free Energy Methods

Qiang Cui (qiangcui@bu.edu)
Boston University, Boston, United States

I'll discuss recent developments and applications of QM/MM methods enhanced with machine learning (ML) in our group. The first topic concerns the integration of Delta-learning in the QM/MM context. We will explore the use of many-body expansion to enhance the efficiency and transferability of Delta learning. In addition, we will discuss how Delta learning can improve the stability and efficiency of multiple time step based multi-level QM/MM molecular dynamics simulations. Next, we discuss progress towards improving the accuracy of semi-empirical QM/MM free energy simulations by learning how a semi-empirical QM/MM potential function differs from a high-level QM/MM model. In particular, we discuss strategies that minimize the amount of sampling at the high-level.

Generation of Open States for Ion Channels using AlphaFold Subsampling and Intrinsic Protein Properties

Cesar Mendoza Martinez (cmendozamartinez001@dundee.ac.uk)
Contributing: Ulrich Zachariae

University of Dundee, Scotland

Ion channels are major drug targets, but structure-based discovery is limited by the scarcity of experimentally resolved open, ion-conductive states. Because these conformations are transient and underrepresented in structural databases, most available ion channel structures correspond to closed states. To address this limitation, we developed a computational workflow that focuses on the pore domain to generate and prioritise candidate open-like conformations. Our approach uses AlphaFold2 subsampling with reduced multiple sequence alignment depth to increase conformational diversity within the pore-forming region. This produces ensembles of alternative pore conformations that may capture transitions toward more conductive states. Candidate models are then filtered using intrinsic properties derived from AlphaFold2 and the protein structure itself, including confidence-related metrics and structural features linked to pore architecture and steric plausibility. Selected conformations are subsequently subjected to molecular dynamics simulations to track their structural behaviour over time and assess whether predicted open-like states remain stable or relax to alternative conformations. We applied this strategy to human TRPC, TRPM, and TRPV families, using experimentally determined closed-state structures as references. The workflow generated structurally diverse pore conformations with varying degrees of expansion relative to closed templates. This provides a scalable framework for exploring ion channel gating and supports future structure-based design of modulators targeting conductive states.

Also Stand №6

Modelling Condensed-phase Reactivity with MLIPs

Fernanda Duarte (fernanda.duartegonzalez@chem.ox.ac.uk)
University of Oxford, United Kingdom

Accurately modelling chemical reactions in solution is key to optimising synthetic procedures and discovering new molecules, and remains one of the grand challenges in computational chemistry. Machine learning interatomic potentials (MLIPs) have become an effective alternative to traditional methods, especially with the advent of general "fundational" models that offer stable, out-of-the-box dynamics. However, their accuracy and efficiency in condensed-phase reactivity remain relatively underexplored. In this talk, I will discuss our efforts to develop efficient training strategies for reactive MLIPs. Using the (M)ACE framework with active learning and enhanced sampling (https://github.com/duartegroup/mlp-train), we have developed MLIPs requiring minimal training sets. Ongoing work on systematically fine-tuning general models has examined how data, model architecture, and sampling influence accuracy and generalisability, highlighting the need for data tailored to each model's requirements. Applications across organic and organometallic systems will be showcased, illustrating how these approaches enable simulations of complex processes in solution.

Learning parsimonious Models for Materials Characterization and Design

Luca Ghiringhelli (luca.ghiringhelli@kit.edu)
Karlsruhe Institute of Technology, Germany

The modeling of macroscopic properties of materials often require to accurately evaluate physical quantities at several time and length scales. Here we show how symbolic inference, i.e., the machine learning of simple analytical expressions that explain and generalize the available data, can effectively bridge physical scales. The focus is on learning models that are as simple as possible (but not simpler!), with as few as possible data points. I will demonstrate the application of the methods to the modeling of catalytic properties of materials, thermal conductivity, and more.

Machine-Learned Interatomic Potentials for Radiation Damage Analysis in High-Temperature Superconductors

Niccolo Di Eugenio (niccolo.dieugenio@polito.it)

Contributing: Ashley Dickson, Flyura Djurabekova, Francesco Laviano, Federico Ledda, Daniele Torsello, Erik Gallo, Mark R. Gilbert, Duc Nguyen-Manh, Antonio Trotta, Samuel T. Murphy, and Davide Gambino

Politecnico di Torino, Turin, Italy
Istituto Nazionale di Fisica Nucleare, Turin, Italy
Max Planck Institute for Polymer Research, Mainz, Germany

High-Temperature Superconductors (HTS) are essential for next-generation fusion reactors, where $YBa_2Cu_3O_{7-}\delta$ (YBCO) forms the active layers of HTS magnets. Because superconducting properties depend strongly on oxygen stoichiometry and are sensitive to defects, understanding radiation-induced damage is critical for predicting performance under neutron irradiation. Atomistic simulations offer insight into microscopic damage processes, but Density Functional Theory (DFT) is limited to small scales, while existing empirical potentials struggle with oxygen stoichiometry and highly disordered environments. Here, we develop and benchmark machine-learned interatomic potentials (MLPs) like ACE, MACE, GAP and tabGAP, trained on a DFT database spanning equilibrium and irradiation-damaged configurations across diverse compositions. Both models achieve near-DFT accuracy for energies and forces, and reproduce key properties including equations of state, elastic constants, bulk modulus, defect formation energies, lattice parameters versus oxygen concentration, activation barriers, and the orthorhombic–tetragonal transition. We further evaluate ACE and tabGAP for radiation-damage modelling via 5–300 keV molecular dynamics collision cascades, threshold displacement energies and non-stoichiometric collision cascades.

Beyond the Canonical Alphabet: De Novo Design of Functional Peptides and GPCR Agonists

Patrick Bryant (patrick.bryant@scilifelab.se)
Stockholm University SciLifeLab, Sweden

Deep learning has revolutionized protein structure prediction and de novo protein design. However, the vast majority of current generative models are restricted to the 20 canonical amino acids. This limitation severely constrains the accessible chemical space, limiting the development of molecules with enhanced proteolytic stability, novel structural geometries, and extended functional capabilities. In this talk, I will present our recent advancements in overcoming these limitations through the development of RareFold. RareFold extends state-of-the-art structure prediction and generative design methodologies to natively support Non-Canonical Amino Acids (NCAAs). By adapting the underlying neural network architectures to process expanded chemical vocabularies, we enable the highly accurate modeling and generation of chemically diverse macromolecules. Furthermore, I will discuss the transition from designing static binders to engineering dynamic, functional modulators. Using RareFoldGPCR, an extension of RareFold, we have successfully designed de novo cyclic peptide agonists targeting complex G-protein coupled receptors (GPCRs). Designing true agonists requires molecular models capable of capturing the subtle, multi-state conformational shifts inherent to receptor activation, moving beyond simple shape-complementarity. I will detail the machine learning strategies required to model these complex multiscale functional interactions, demonstrating how our framework opens new avenues for precision therapeutics.

Machine-Learning Framework for Excitation Energies of Chromophores in Polarizable Environments

Chris John (chris.john@unipi.it)
University of Pisa, Italy

Multiscale methods integrating quantum chemical descriptions with molecular mechanics (MM) represent the standard modeling strategy employed to investigate properties and processes of molecules embedded in complex biological matrices. Statistically meaningful calculations however must require performing quantum mechanics/MM (QM/MM) calculations over a sufficiently large ensemble of configurations, a task that remains computationally demanding. This challenge becomes particularly critical when investigating multichromophoric assemblies, such as light-harvesting complexes, employed by photosynthetic organisms to capture and funnel solar energy to the reaction centers. The protein matrix and surrounding solvent provide the specific electrostatic and polarization environment that tunes the excitation energies of the individual chromophores, thereby determining the efficiency of light harvesting. The large number of polarizable QM/MM calculations needed can represent a significant computational bottleneck. Machine-learning (ML) techniques offer a way out of this computational challenge. Herein, we present an ML framework capable of predicting the different environmental contributions to excitation energies of chromophores in a polarizable embedding. Our ML models are built in a hierarchical structure to capture both the effect of ground-state polarization and the response of the polarizable environment to the electronic transition. With the use of the right descriptors, the models trained on electrostatic-embedding QM/MM calculations are able to successfully predict the effects of a polarizable environment on excitation energies. The ML models are applied to three prototypical photosynthetic chromophores (chlorophyll a, chlorophyll b, and lutein) and are used to reproduce the excitonic structure of a multichromophoric system unseen in the training set to a level of accuracy offered by a polarizable QM/MM calculation, while taking a fraction of the time.

Controlling Generative Models

Alexander Tong (atong@aithyra.at)
AITHYRA, Vienna, Austria

Generative modeling excels at mimicking data distributions, but scientific discovery often requires sampling from a modified distribution. In this talk, I will motivate and present a framework for controlling flow models to control sampling on a distribution level. I will demonstrate how this approach can be adapted to two distinct challenges in the natural sciences: (1) guiding generation to control for multiple functional properties for design, and (2) steering dynamics to sample from unnormalized energy densities for equilibrium sampling. Together, these methods illustrate the versatility of flows for both finding new, optimized structures and exploring the physical landscapes of existing ones.

Branching Flows: Discrete, Continuous, and Manifold Flow Matching with Splits and Deletions

Ben Murrell (benjamin.murrell@ki.se)
Karolinska Institute, Stockholm, Sweden

Diffusion and flow matching approaches to generative modeling have shown promise in domains where the number of elements in a state is fixed in advance (e.g. images), but require ad hoc solutions when, for example, the length of a response from a large language model, the number of atoms in a molecule, or the number of amino acids in a protein chain is not known a priori. Here we propose Branching Flows, a generative modeling framework that, like diffusion and flow matching approaches, transports a simple distribution to the data distribution, but in Branching Flows, the elements in the state evolve over a forest of binary trees, branching and dying stochastically with rates that are learned by the model. This allows the model to control, during generation, the number of elements in the sequence. We show that Branching Flows can compose with any flow matching base process on discrete sets, continuous Euclidean spaces, Riemannian manifolds, and `multimodal' product spaces that mix these components, and we demonstrate distribution matching on small molecules and antibody sequences, and that this scales to complicated domains such as protein structures.

Physics Informed Machine Learning Potential for Brownian Dynamics Simulations

Riccardo Beccaria (riccardo.beccaria@h-its.org)

Heidelberg Institute for Theoretical Studies, Germany

Brownian Dynamics (BD) simulations provide an efficient alternative to all-atom Molecular Dynamics (MD) by enabling the study of biomolecular interactions on larger spatial and temporal scales while maintaining the atomistic description of macromolecules. However, their predictive power critically depends on the quality of the underlying force field, which must remain computationally efficient to allow simulations of up to thousands of molecules. In this project, we develop a coarse-grained (CG), physics-informed machine learning (ML) force field designed to better capture the short-range residue-specific interactions of proteins. The model is designed to preserve physical principles to ensure interpretability, while introducing flexibility in its parameters to better capture the interactions. It is trained using energy matching to reproduce potentials of mean force (PMFs) derived from extensive all-atom MD simulations of all natural amino acid pairs, ensuring consistency with the underlying molecular physics. The resulting force field is implemented in SDA (a Brownian Dynamics software) and applied to many-body systems, including protein aggregation of the Villin Headpiece and diffusion of macromolecules in crowded environments. This establishes a transferable framework for incorporating data-driven interactions into efficient coarse-grained BD simulations.

Molecular Strategies for Confinement-Controlled Water Structure and Chemistry

Yair Litman (litmany@mpip-mainz.mpg.de)
Max Planck Institute for Polymer Research, Mainz, Germany

The behavior of aqueous electrolyte interfaces with carbon materials underpins the performance of many nanofluidic and electrochemical devices. In this talk, I will present recent advances combining surface-specific vibrational spectroscopy with machine-learning-assisted molecular simulations to reveal water's structure and reactivity at interfaces and in nanoconfinement. I will show how ions crowd at graphene surfaces without disrupting the interfacial hydrogen-bond network[1], how graphene's polarizability masks yet reshapes local water orientation[2], and how confinement, interfacial chemistry, and intense electric fields can dramatically transform the thermodynamics of the water autodissociation reaction[3,4]. Together, these findings outline fundamental principles for controlling interfacial water, opening new strategies for confinement-controlled chemistry in nanofluidic, electrochemical, and catalytic systems.

---

[1] X. R. Advincula, K. D. Fong, Y. Wang, C. Schran, M. Bonn, A. Michaelides, Y. Litman J. Am. Chem. Soc. In Press (2026).

[2] Y. Wang, Y. Litman, M. Cho, S. J. Cox, M. Bonn, Chem In Press (2026) (preprint https://arxiv.org/abs/2511.04930).

[3] Y. Litman, A. Michaelides, J. Am. Chem. Soc. , 147, 49, 44885 (2025).

[4] X. R. Advincula, Y. Litman, K. D. Fong, W. C. Witt, C. Schran, A. Michaelides (under review in Sci. Adv. preprint arXiv:2508.13034).

Surrogate Functionals for Machine-Learned Orbital-Free Density Functional Theory

Roman Remme (roman.remme@iwr.uni-heidelberg.de)
Heidelberg University, Germany

We introduce surrogate functionals: machine-learned energy functionals for orbital-free density functional theory (OF-DFT) that are designed not for universal accuracy, but to yield the correct ground-state electron density after optimization. This shifts the learning objective from faithfully reproducing a reference functional to enforcing correctness of the optimization outcome. As a result, training requires only ground-state labels, and an adaptive sampling scheme focuses learning along density optimization trajectories. The resulting functionals converge robustly across neural network architectures and eliminate the need for the $O(N^3)$ orthonormalization step required in prior approaches. They improve electron density predictions by an order of magnitude over previous machine-learned OF-DFT methods. We further introduce “strong” surrogates, which are additionally trained to reproduce ground-state energies, achieving accuracy competitive with state-of-the-art machine-learned interatomic potentials.

Foundational Machine Learning Force Fields for Atomistic Chemistry

Ilyes Batatia (ilyes319@hotmail.fr)
University of Cambridge, United Kingdom

Machine learning force fields are transforming computational materials science, offering accurate and efficient alternatives to electronic structure methods for extended simulations. The latest architectures utilise many-body equivariant message-passing networks such as MACE. Our recent results demonstrate a paradigm shift in the development of machine learning force fields, moving the field from system-specific models to widely transferable foundation models. Our latest foundation models are capable of performing stable and accurate molecular dynamics at ambient conditions across a wide range of systems, including crystals, liquids, molecules, and surfaces. The impressive generalisation of such foundation models paves the way for creating truly universal force fields with quantum-level accuracy. These foundation models are already transforming computational materials science, as with very little domain-specific data, one can fine-tune them to achieve quantitatively accurate results, significantly lowering the barrier to entry into the field.

Efficient, Explainable and Extrapolative AI for Biophysics

Pratyush Tiwary (ptiwary@umd.edu)
University of Maryland, USA

AI is now everywhere in chemistry, from structure prediction to molecule generation to automated synthesis. The excitement is real, but so is the unease about what is genuinely predictive and what is closer to impressive memorization. In this talk I will take a statistical physicist's perspective and use examples from my group's work to argue for cautious, but clear, enthusiasm for AI in chemistry and allied fields. I will show how we combine generative AI with statistical mechanics to learn Boltzmann weighted ensembles from limited training data, and then extrapolate across temperature, pressure, and other thermodynamic conditions reducing the need for explicit, expensive simulations or experiments. I will highlight the breadth of these methods through applications that include prediction of protein and RNA structural ensembles, and conformation selective drug discovery efforts aimed at Alzheimer's disease and hypertension.

Taking the Next Step with Machine Learning Orbital Free DFT

Tobias Kaczun (tobias.kaczun@iwr.uni-heidelberg.de)
Heidelberg University, Germany

Orbital-free density functional theory (OF-DFT) holds the promise of computing ground state molecular properties at low computational cost. However, despite decades of research, no kinetic energy functional has achieved sufficient accuracy for general use in computational chemistry. Recent advances in machine-learned kinetic and exchange–correlation functionals have shown great promise, reaching Kohn–Sham accuracy at OF-DFT cost and enabling extrapolation from training on small organic molecules to significantly larger systems. To date, however, these results have been limited to equilibrium geometries. Beyond single-point calculations, geometry optimizations and molecular dynamics simulations are essential components of many computational chemistry workflows. By deriving and implementing analytical forces within our OF-DFT ansatz and incorporating non-equilibrium geometries into the training data, we, for the first time, enable geometry optimizations and molecular dynamics simulations in machine-learned OF-DFT. We present first results for drug-like organic molecules with up to ∼60 atoms. This work represents the next step toward the general applicability of this method to molecular systems.

Differentiable Quantum Chemistry: Hybrid Machine Learning Models and More

Martin Stöhr (martin.stoehr@stanford.edu)
Stanford University, USA

Machine learning interatomic potentials (MLIPs) have transformed the methodological landscape of computational chemistry, by now challenging first-principles approaches as workhorse methods for ground-state simulations. However, for excited-state calculations and especially photochemistry simulations existing MLIPs often perform rather poorly due to the independent/inconsistent treatment of the involved electronic states. This can be addressed through hybrid approaches of ML and quantum mechanical (QM) model Hamiltonians, combining qualitative fidelity from the QM model with quantitative accuracy thanks to a flexible ML-based parametrization. Directly encoding non-local QM features, the parameters of the model Hamiltonian typically show a rather local dependence on the system's configuration, making them an ideal target for dynamic prediction by ML. A crucial step for establishing such hybrid QM-ML approaches is the development of differentiable quantum chemistry implementations to enable the efficient generation and propagation of property gradients as needed for training the embedded ML model. In this presentation, we will explore our recent GPU-accelerated differentiable platform for semiempirical multi-reference methods and its application in hybrid QM-ML frameworks. The final hybrid approach provides a consistent description across electronic states including their intersections and critical access to electronic properties, paving the way to efficient and robust photochemistry simulations.

Hyperspatial Sampling

Henrik Christiansen (henrik.christiansen@neclab.eu)
NEC Laboratories Europe, Heidelberg, Germany

The efficient simulation of systems with rugged free-energy landscapes remains a central challenge in computational science, ranging from protein folding to materials design. We introduce hyperspatial replica exchange, an enhanced sampling method in which the physical system is artificially extended by additional spatial dimensions. In higher dimensions, free-energy barriers that confine the system in three dimensions can be circumvented through paths that are geometrically inaccessible in the original space. An appropriate penalty confines atoms near the physical subspace; by simulating replicas at a series of different penalties and exchanging configurations between them, the system is capable of efficiently exploring the extended state space at small penalties, and correctly sampling the target canonical ensemble at large penalties. Restricting the penalty to solute atoms only, analogous to replica exchange solute tempering, dramatically reduces the number of replicas required for solvated systems compared to standard temperature replica exchange. We demonstrate the method on a double-well model system and on alanine dipeptide in explicit water, where hyperspatial replica exchange achieves enhanced conformational sampling of the slow modes of the molecule, without explicitly defining collective variables.

Posters

Identifying Dynamical Fingerprints of Rare-Event Transitions

Ali Sharifian (ali.sharifian@kit.edu)
Karlsruhe Institute of Technology, Germany

A transition-conditioned framework is introduced for the extraction of dynamical fingerprints of designated rare transitions. By restricting analysis to reactive-corridor ensembles, a low-dimensional representation of the transition motion is identified together with a ranking of microscopic players that are most informative for barrier crossing. Across benchmark proteins, the resulting hotspot players are shown to yield more faithful mechanism-preserving acceleration than conventional or random baselines. A practical route is thereby provided for the identification of transition-specific motions, players, and control points in rare-event dynamics.

Stand № 1

Graph Diffusion Models for Backmapping in Chemical Compound Space

Luis Itza Vazquez-Salazar (l.i.vazquez-salazar@thphys.uni-heidelberg.de)
Heidelberg University, Germany

Chemical compound space (CS), which is the collection of all possible molecules and materials, is extremely vast. This makes exploring it both challenging and necessary. Coarse-grained (CG) models provide a strong way to simplify this complexity by grouping atoms into interaction centers called beads. In transferable CG models, one bead can symbolize several compounds. This positions CG methods as an effective way to navigate CS. However, backmapping, which means reconstructing an all-atom (AA) representation from a CG description, is still quite complicated. Machine learning (ML) techniques have become a promising approach for connecting CG and AA resolutions. In this work, we use generative ML for backmapping across CS with a classifier-free graph diffusion model. Molecular generation is based on specific bead types by directly supplying the bead identity as input. Our method creates molecules with different numbers of heavy atoms (C, N, O, F) based on individual bead types. It achieves high validity, diversity, and uniqueness, with over 80% of the generated structures meeting these criteria.

Stand № 2

Efficient and Accurate Prediction of Infrared Spectra of Small Molecules Using Machine-Learning Force Fields

Ruslan Levochkin (ruslan.t.levochkin@gmail.com)
Constructor University, Bremen, Germany

The study aims at developing efficient computational workflows for accurately predicting in- frared (IR) spectra of small molecules. While the methods of computing IR spectra using linear response theory are widely used, recent advances in the field of Machine Learning and, namely, so-called Machine-Learning Force Fields (MLFFs) in Molecular Dynamic Simulations hold a promising ground for being orders of magnitude faster than pure quantum mechanical calculations while maintaining the same or close level of accuracy and showcasing extensive scalability. The main focus of the study is on benchmarking multiple simulation techniques that vary by the level of theory behind the calculations, integration technique, and starting conditions against the highly accurate experimental spectra using High-Performance Computing (HPC) to challenge the limits of MLFFs and, hopefully, yield the most optimized method that can be scaled up to process larger arrays of small molecules.

Stand № 3

Tracking Conformational Transitions in Malarial PfCRT Using Enhanced Sampling Simulations

Kushal Singh (kushal.singh@h-its.org)
Heidelberg Institute for Theoretical Studies, Germany

Malaria is a life-threatening, endemic vector-borne parasitic disease caused by parasites of the genus Plasmodium, has been responsible for an estimated 263 million cases and 597,000 deaths worldwide in 2023. Resistance to commonly prescribed antimalarial drugs such as chloroquine (CQ) and piperaquine (PPQ) is primarily driven by specific point mutations in the Plasmodium falciparum chloroquine resistance transporter (PfCRT). These mutations enable PfCRT to recognize and efflux drugs from the parasite's digestive vacuole, thereby reducing drug efficacy. Similar to other trans-membrane transport proteins, PfCRT hypothetically operates via an alternating access mechanism, whereby it undergoes large-scale conformational changes between inward-facing, occluded and outward-facing states. However, the detailed dynamics underlying these transitions remain poorly understood. Here, we utilize a computational workflow which combines machine learning, conventional, and enhanced-sampling molecular dynamics simulations to characterize the conformational landscape of PfCRT in both ligand-bound and ligand-free states. This approach aims to provide a comprehensive mechanistic understanding of PfCRT dynamics and to elucidate how resistance-associated mutations can alter drug transport.

Stand № 4

Predicting Surfactant Properties from an MD Database

Richard Beckmann (ricbec@chalmers.se)
Chalmers University of Technology, Gothenburg, Sweden

Surfactants, a portmanteau of "surface-active agents", are chemical compounds affecting the interfacial properties between two phases, whether this interface is liquid-gas, liquid-solid, or liquid-liquid. Here we present an extension of an existing database of experimental properties of surfactant molecules by carrying out large-scale MD simulations for all compounds in the database. With the extended database, we created a suite of accurate machine learning models to predict surfactant properties based on SMILES strings alone.

Stand № 5

Generation of Open States for Ion Channels using AlphaFold Subsampling and Intrinsic Protein Properties

Cesar Mendoza Martinez (cmendozamartinez001@dundee.ac.uk)
Contributing: Ulrich Zachariae

University of Dundee, Scotland

Ion channels are major drug targets, but structure-based discovery is limited by the scarcity of experimentally resolved open, ion-conductive states. Because these conformations are transient and underrepresented in structural databases, most available ion channel structures correspond to closed states. To address this limitation, we developed a computational workflow that focuses on the pore domain to generate and prioritise candidate open-like conformations. Our approach uses AlphaFold2 subsampling with reduced multiple sequence alignment depth to increase conformational diversity within the pore-forming region. This produces ensembles of alternative pore conformations that may capture transitions toward more conductive states. Candidate models are then filtered using intrinsic properties derived from AlphaFold2 and the protein structure itself, including confidence-related metrics and structural features linked to pore architecture and steric plausibility. Selected conformations are subsequently subjected to molecular dynamics simulations to track their structural behaviour over time and assess whether predicted open-like states remain stable or relax to alternative conformations. We applied this strategy to human TRPC, TRPM, and TRPV families, using experimentally determined closed-state structures as references. The workflow generated structurally diverse pore conformations with varying degrees of expansion relative to closed templates. This provides a scalable framework for exploring ion channel gating and supports future structure-based design of modulators targeting conductive states.

Stand № 6

Protonation Modulates Architecture and Hydration in Peptide Condensates

Vasileios Xenidis (xenidisv@uni-mainz.de)
Max Planck Graduate Center, Mainz, Germany

To gain deeper mechanistic insight into the self-mediated assembly of peptide-based condensates, we conducted all-atom molecular dynamics (MD) simulations of short, sequence-defined peptides under varying protonation states. Our results demonstrate differences in the aggregation mechanism under various protonation rates. Specifically, protonation modulates spatial organization within the condensate: protonated peptides preferentially localize to the outer periphery, whereas neutral peptides are enriched in the interior core. The overall protonation fraction strongly influences condensate morphology, with systems containing a higher proportion of protonated peptides forming more diffuse and less compact assemblies, while predominantly deprotonated systems yield denser, more compact structures. Hydration analysis including solvent-accessible surface area (SASA) and local water density profiles, reveals distinct solvent partitioning across protonation setups. Further analysis of non-covalent interactions revealed that π-π stacking and hydrogen bonding act as driving forces for the self-organization process, with T-shaped interactions emerging as dominant. These findings provide mechanistic insight into how protonation state regulates condensate architecture and stability.

Stand № 7

Towards a Mechanistic Understanding of Kinase-Catalysed N-Phosphorylations Using Multiscale Simulations

Lena Eichinger (lena.eichinger@kit.edu)
Karlsruhe Institute of Technology, Germany

Phosphorylation of nitrogen atoms are fundamental in biological systems, including histidine autophosphorylation in signal transduction, yet their detailed molecular mechanisms remain incompletely understood. Hybrid QM/MM methods enable the description of enzymatic reactions at atomic resolution by treating the active site quantum mechanically and the surrounding protein classically, while enhanced-sampling techniques extend accessible timescales and allow exploration of the free-energy landscape. Semi-empirical methods such as density-functional tight-binding (DFTB) provide high computational efficiency but often lack sufficient accuracy. To improve the accuracy of the QM description, a Δ-machine learning (ΔML) correction based on high-dimensional neural network potentials (HDNNPs) is applied to DFTB, improving its accuracy towards density functional theory (DFT) while retaining its computational efficiency. The ΔML model is trained and evaluated in vacuum for a diverse set of N-phosphorylation model reactions and subsequently tested in aqueous simulations to assess transferability between environments. Application to a histidine kinase system provides both methodological and mechanistic insights into the autophosphorylation reaction, particularly regarding the role of Mg2+ coordination and the influence of artificial restraints. Overall, this work demonstrates how combining QM/MM methods, enhanced sampling, and ΔML corrections enables mechanistic insight into kinase-catalysed N-phosphorylation reactions and establishes a foundation for future applications to complex enzymatic systems.

Stand № 8

Creation of a Charge Propagation Workflow to calculate the Charge Mobility of Organic Semiconductors

Sebastian Will (sebastian.will@kit.edu)
Heidelberg University, Germany

Charge mobility is an important quantity to measure the effectiveness of organic semiconductors, calculating it is difficult and involves a lot of manual steps, therefore a workflow is being created in order to simplify those steps so that more people have access to charge mobility for research.

Stand № 9

Computing Residence Time for Flexible Peptide-Antibody Complexes using Enhanced Sampling and Clustering for Conformational Classification

Mislav Brajković (mislav.brajkovic@h-its.org)
Darius Szablowski (darius.szablowski@h-its.org)
Contributing: Anton Hanke, Hedda Wardemann, Rebecca C Wade

Heidelberg Institute for Theoretical Studies, Germany

The dissociation (k_{off}) rate constant, or its inverse residence time ($\tau=1/k_{off}$), of a protein-ligand complex is often a good predictor drug efficacy in the non-equilibrium conditions of living organisms.[1] τRAMD (τ Random Acceleration Molecular Dynamics) has been developed for efficiently computing relative residence times and ranking different ligands according to their ligand-target τ.[2] This method uses an additional randomly-oriented force applied to the ligand to accelerate its dissociation during Molecular Dynamics (MD) simulations. τRAMD was initially developed for small molecule-protein complexes. In this study, we report on the application of τRAMD to complexes where the ligands are highly flexible repeating 20-amino acid peptide sequences from Circumsporozoite Surface Protein (CSP) from Plasmodium falciparum, a parasite that causes malaria, and the target proteins are mutants of a high affinity antibody. We evaluate the use of τRAMD with two different force fields: CHARMM36m and AMBER99SBdisp.[3] Initial conformational analysis of the MD trajectories revealed that the bound peptides can adopt two distinct conformational ensembles. Leveraging this knowledge, an automatic clustering pipeline involving two-step clustering with remapping was developed in order to separate all peptide conformations sampled during MD simulations into two major clusters. Classifying τRAMD-replicas based on one of the identified clusters, rather than the full set of conformations, leads to an improved agreement between computed and experimental residence times. This finding enables the identification of amino acid residues that determine antibody-peptide residence times.

Stand № 10

---

[1] Copeland, R. A. et al., Nature reviews Drug discovery, 5(9), 730-739 (2006).
[2] Kokh, D.B. et al., Journal of chemical theory and computation, 14(7), pp.3859-3869 (2018).
[3] Robustelli, P. et al., Proceedings of the National Academy of Sciences, 115(21), pp.E4758-E4766 (2018).

One Year of Bio-Structure Hub: Insights into AI Structure Prediction

Christine Schulz (christine.schulz@iwr.uni-heidelberg.de)
Contributing: Inga Ulusoy, Rebecca Wade

Heidelberg University, Germany
Heidelberg Institute for Theoretical Studies, Germany

The explosion of AI tools for predicting protein, RNA, and other macromolecular structures is reshaping everything from basic biochemistry to drug design and synthetic biology. In practice, however, getting the most out of these methods still requires specialist knowledge and robust computing resources. One year ago, the Bio-Structure Hub was created to lower these hurdles and give researchers a reliable, easy-to-use entry point. The Hub provides domain expertise and research-software development support, with a focus on validation tools. In addition, the Hub offers a central infrastructure that provides access to tested and optimised AI-software installations on suitable compute resources, emphasizing usability for non-expert users. Here we present the lessons learned from the first twelve months: we have assisted researchers across diverse projects, identified common technical bottlenecks, and curated a script collection that abstracts away the technical steps, enabling new users to run AI-prediction jobs by following step-by-step tutorials.

Stand № 11

Mobility-Driven Carbohydrate Recognition by the Carbohydrate-Binding Module 21 Directs Glycoside Hydrolase 15 Toward Productive Catalysis

Vinicius Ávila Cabral (vinicius.cabral@h-its.org)
Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
Heidelberg Institute for Theoretical Studies, Germany

Carbohydrate-binding modules (CBMs) enhance the catalytic efficiency of glycoside hydrolases (GH) by mediating substrate recognition and positioning. CBM21 from Rhizopus oryzae is linked to GH15 via a disordered, highly O-glycosylated linker. Despite its biotechnological potential, its low binding affinity and highly dynamic behavior make its mechanism elusive. Here, we combine extensive molecular dynamics and metadynamics simulations using classical Molecular Mechanics (MM) and multiscale Quantum Mechanics/Molecular Mechanics (QM/MM) to characterize CBM21 interactions with amylose. Our results indicate that the unusual 1S3 glucose conformation found in crystal structures of CBM21 is likely a crystallographic artifact rather than a functional state. Simulations with long amylose chains reveal that CBM21 binds transiently, alternating between two canonical binding sites through repeated unbinding and rebinding events. Remarkably, CBM21 exhibits directional motion along amylose fibers toward the non-reducing ends, the target sites for the associated GH15. This movement is driven primarily by polysaccharide entropic contributions. Markov State Model analysis suggests an additional transient binding region facilitates this guided transport. These findings support a mobility-dependent mechanism where CBM21 acts as a dynamic guide, steering substrates toward the catalytic domain. This atomistic framework provides essential insights for engineering CBMs with tailored properties for industrial biocatalysis.

Stand № 12

Molecular Optimization of the Coarse-Grained Martini Model from Differentiable Simulations

Immanuel Sulzer (immanuel.sulzer@stud.uni-heidelberg.de)
Heidelberg University, Germany

Coarse-grained models such as Martini 3 enable molecular simulations at scales inaccessible to all-atom approaches, but their Lennard-Jones parameters have traditionally been fit by hand. We are developing a framework for the automated, gradient-based optimization of these parameters using differentiable molecular dynamics. To this end, we have implemented soft-core potentials, thermodynamic integration, and Martini 3 support within the differentiable simulation engine Molly.jl, enabling the computation of exact gradients of thermodynamic observables, such as solvation free energies, radial distribution functions, and densities, with respect to interaction parameters. Because backpropagation through entire trajectories is impractical, we employ a Monte Carlo gradient estimator that aggregates derivatives from independent short trajectories. The framework is validated against GROMACS across systems of increasing complexity, from small-molecule solvation to bilayer membranes with restrained particles. Current work focuses on closing the optimization loop to perform end-to-end parameter updates driven by solvation free energy targets, while future efforts aim to differentiate bilayer demixing free energies, opening the door to optimizing coarse-grained parameters for lipid membrane thermodynamics. This work establishes a path toward systematic, data-driven reparametrization of the Martini force field.

Stand № 13

Title to be announced

Morteza Chehel Amirani (mchehelamirani@svanteinc.com)
Svante, Calgary, Canada

To be announced

Stand № 14

Reactive Learning Construction of a Reactive Machine-Learned Potential for Bismuth Cluster Assembly

Yumeng Zhang (yumeng.zhang@kit.edu)
Karlsruhe Institute of Technology, Germany

Understanding size-selective formation in heavy-element clusters remains a major challenge due to complex reactive landscapes and the high cost of quantum chemical simulations. In particular, the emergence of Bi_{11} as a dominant product in bismuth aggregation raises fundamental questions about its stability and formation mechanisms. Here, we develop a reactive machine-learned interatomic potential based on the MACE architecture, trained using an uncertainty-driven active learning framework. Starting from a small initial dataset, the model is iteratively refined through molecular dynamics exploration and selective first-principles labeling. The resulting potential enables large-scale reactive simulations across a wide temperature range. We observe a pronounced, temperature-dependent increase in Bi_{11} formation.

Stand № 15

End-to-End Nonbonded Parameterization for Machine Learning Force Fields via Physics-Informed Regularization

Berkay Günes (guenesb@mpip-mainz.mpg.de)
Max Planck Institute for Polymer Research, Mainz, Germany
Heidelberg University, Germany

Classical molecular mechanics force fields rely on rigid, human-defined atom-typing rules that cannot capture the full nuance of continuous chemical environments. While Graph Neural Networks (GNNs) offer an automated path to generate bespoke parameters, their application to nonbonded interactions is complicated by the inherent degeneracy between long-range electrostatics and van der Waals (vdW) dispersion when trained on vacuum monomer energies. We present a unified, end-to-end parameterization framework that extends the Grappa force field to include fully self-consistent nonbonded interactions. To break the electrostatic-dispersion degeneracy, our architecture avoids direct total-energy fitting for nonbonded terms. Instead, it predicts the nonbonded interaction parameters directly. By employing a gradient-stopping mechanism, the bonded parameters learn the residual energy landscape while the nonbonded terms remain anchored to their physical regularizers. Our prototype successfully reproduces established nonbonded interaction energies with an RMSE of 0.50 kcal/mol, while maintaining rigorous topological masking (1-4 scaling) and exact molecular charge conservation. This work paves the way for force fields that possess the accuracy of quantum physical proxies with the computational efficiency of classical mechanics.

Stand № 16

SITH: A Quantum-Chemical Framework for Predicting Bond Destabilization in Stretched Molecules

Daniel Sucerquia (daniel.sucerquia@mpip-mainz.mpg.de)
Max Planck Institute for Polymer Research, Mainz, Germany

Mechanical forces can selectively destabilize chemical bonds of molecular systems, particularly in biological and synthetic polymers. While experimental and theoretical methods have advanced our understanding of mechanochemical processes, predicting where energy concentrates within a molecule remains a significant challenge. To address this, we introduce SITH (Splitting Intramolecular Tension due to stretcHing), a novel method that decomposes the total electronic energy change of a stretched molecule into the contributions from its internal degrees of freedom--such as bond lengths, angles, and dihedrals--using numerical integration of the work-energy theorem. Unlike previous approaches that rely on harmonic approximations, SITH provides high accuracy and robustness to study the distribution of energies of stretched molecules up to a first bond cleavage. Although SITH uses 3N-6 degrees of freedom for the energy decomposition, we show that it can work even for ring structures like prolines. We apply SITH to a dataset of tripeptides and demonstrate that glycine and proline exhibit significantly different energy distributions in their Cα–C backbone bonds under tension: proline requieres less energy to be elongated, making it more prone to rupture, while glycine has the opposite behaviour. These findings reveal intrinsic differences in mechanochemical susceptibility across amino acids, offering more accurate predictions of bond rupture in proteins and other (bio)polymers. SITH thus provides a powerful, interpretable tool for understanding energy distribution at the quantum level, with possible applications in mechanochemistry and force field validation.

Stand № 17

Multiscale Simulations to study Electron Transfer Complex Formation of Membrane-bound Cytochrome P450 17A1 and its Redox Proteins

Jonathan Teuffel (jonathan.teuffel@h-its.org)
Contributing: Goutam Mukherjee, Sungho Bosco Han, Marcus Elstner, Rebecca C. Wade

Heidelberg Institute for Theoretical Studies, Germany

Cytochrome P450 enzymes form a family of heme-containing monooxygenases that are present in all domains of life and play essential roles in many catabolic and anabolic processes. Their canonical reaction constitutes the chemical activation of molecular oxygen through a reductive process that requires an external supply of electrons from secondary redox proteins. The cytochrome P450 17A1 (CYP17A1) is a central enzyme in steroid hormone synthesis and is one of a subgroup of cytochrome P450 enzymes that interacts with two redox proteins: cytochrome b5 (CYb5) and NADPH cytochrome P450 oxidoreductase (CPR). We applied a multiresolution simulation protocol[1] to investigate how human CYP17A1 forms binary and ternary complexes with these two redox proteins in a phospholipid bilayer. Extensive molecular dynamics simulations of the complexes were analysed to yield information on the binding modes and structural rearrangements of the redox partners upon complex formation and on electron-transfer kinetics. Additionally, Marcus theory parameters required to quantitatively describe electron transfer kinetics were computed from combined quantum mechanics- and molecular mechanics simulations[2]. Our findings show how association to different redox proteins differentially impacts the activity of CYP17A1 through conformational rearrangements.

Stand № 18

---

[1] Mukherjee G. et al. Communications Biology 4, 55 (2021).
[2] Teuffel J. et al. The Journal of Chemical Physics 162(19), 195101 (2025).

From DFT to CCSD(T): Method-conditioned Multi-fidelity Energy Learning

Marlen Neubert (marlen.neubert@kit.edu)
Karlsruhe Institute of Technology, Germany

High-level quantum-chemical methods such as CCSD(T) provide highly accurate reference energies but are too expensive for large-scale data generation. In contrast, many lower-cost density-functional approximations are available, but they are less accurate. We investigate whether a single machine-learning potential can learn molecular energies across many quantum-chemical levels of theory and transfer toward CCSD(T)-quality predictions through method conditioning. Our approach uses a method-conditioned MACE model in which the quantum-chemistry method is represented not as a discrete label, but as a continuous descriptor. This descriptor is derived from a small calibration set and maps methods into a continuous space. The resulting representation enables interpolation across methods and querying at a coupled-cluster descriptor, while training the model only on DFT data. On the MSR-Acc dataset, the calibrated method descriptor improves prediction accuracy substantially over unconditioned and simpler conditioned baselines, both for known and unseen DFT methods. Querying at the coupled-cluster descriptor also improves over the available DFT baseline on unseen coupled-cluster geometries, indicating that the model learns a geometry-dependent correction toward higher-level quantum chemistry.

Stand № 19

Towards Accurate Density Functionals for Molecular Adsorption on Metal Surfaces

Benjamin Xu Shi (bshi@flatironinstitute.org)
Contributing: Timothy C. Berkelbach

Flatiron Institute, New York, USA

Machine-learning interatomic potentials (MLIPs) are fundamentally limited by the accuracy of their underlying electronic structure methods. In particular, for catalytic transition-metal surfaces, commonly used density functional approximations within density functional theory (DFT) often fail to reach the ~3 kcal/mol (13 kJ/mol) threshold associated with transition-metal chemical accuracy, frequently exhibiting errors nearly twice as large. These limitations are evident in well-known failures, such as the incorrect description of CO adsorption on Pt(111). In this poster, I present a new class of density functionals that systematically achieve this level of accuracy while maintaining high computational efficiency. These advances enable the construction of more reliable MLIPs for complex chemical environments. I will demonstrate the impact of this approach through applications to molecular adsorption on surfaces, highlighting both improved accuracy and practical scalability.

Stand № 20

Computation of Electron Transfer Reorganization Energies in Cytochrome P450 Systems

Levi Miederer (levi.miederer@h-its.org)
Contributing: Jonathan Teuffel, Rebecca C. Wade

Heidelberg Institute for Theoretical Studies, Germany

Electron transfer (ET) reactions are central to biological redox processes and commonly described by Marcus theory, where the reorganization energy λ is a critical parameter for ET rates. Because experimental determination of λ in proteins remains challenging, computational approaches based on molecular dynamics (MD) simulations have been established. Such approaches have been applied to cytochrome P450 (CYP) enzymes, which catalyze essential reactions in drug metabolism and steroid biosynthesis. Their catalytic cycle includes two ET steps involving their heme cofactor, making them challenging systems for studying protein ET. We have shown how the QM+MM approach can be used to compute λ for CYP17A1-redox protein complexes but observed unexpected ET energy distributions for some complexes[1]. Here, we examine the reason for such deviations from Marcus theory and find that they result from the parameterization of the molecular mechanics force field. We evaluate different parameterizations for CYP17A1 and the structurally highly flexible CYP2B4 using μs-scale MD simulations. We show that λ is modulated by conformational transitions of the protein and gain insights into the impact of heme force-field parameterization on computed λ values. These results yield generalized guidelines for computing λ values in complex biomolecular systems such as membrane-bound CYP-redox protein complexes.

Stand № 21

---

[1] Teuffel J. et al., J. Chem. Phys., 2025.

From Molecular Dynamics Simulations to Scaffold Discovery: A Multiscale Computational Framework

Guilherme Pinheiro da Silva (guilhermepinheirodasilva@gmail.com)
Federal University of Rio Grande do Sul, Porto Alegre, Brazil

A central challenge in machine-learning-driven molecular modeling is the construction of representations that encode the physicochemical complexity and dynamics of molecular environments. However, binding-site characterization in structure-based drug design is often based on static snapshots, neglecting the rich dynamical information present in molecular simulations. This work proposes a conceptual framework for learning dynamic representations of protein binding sites directly from molecular dynamics trajectories to support ligand scaffold discovery. The approach integrates time-resolved physicochemical descriptors of protein–ligand binding sites with representation-learning strategies, modeling the binding site as a continuous scalar field defined over a fixed spatial grid. Within this reference grid, spatiotemporal multichannel descriptors (e.g., aromaticity, hydrogen bonding, hydrophobicity) can be extracted throughout the simulation and processed using machine-learning models. The resulting patterns can be translated into interpretable ligand design principles to support scaffold hopping. This multiscale perspective bridges atomistic simulations and data-driven molecular design and is formulated as a generalizable strategy applicable to diverse protein targets. Conceptual considerations regarding descriptor selection, sampling requirements, and integration with data-driven models are highlighted as methodological challenges for operationalizing the proposed framework in computer-aided drug design workflows.

Stand № 22

Coarse-Grained High-Throughput Screening of Chemical Space for Protein–Ligand Binding

Luis Walter (walter@thphys.uni-heidelberg.de)
Heidelberg University, Germany

Estimating absolute protein-ligand binding free energies remains a key challenge in computational drug discovery. All-atom free-energy calculations are computationally expensive, limiting their applicability to large-scale screening of chemical space. Coarse-grained (CG) models offer a promising route to accelerate such calculations, yet their accuracy for protein-ligand binding has received limited attention so far. Here, we evaluate the ability of the Martini 3 CG force field to predict absolute binding free energies and present a workflow that integrates these calculations with Bayesian optimization for efficient exploration of chemical space. We apply our approach to the T4 lysozyme L99A mutant, a well-characterized benchmark system for binding studies. Our results show that Martini 3 reproduces relative binding trends across a diverse set of small molecules. By coupling CG free-energy calculations with a data-driven search strategy, we can systematically explore large regions of chemical space and derive insights into relevant chemical features for binding.

Stand № 23

Benchmarking and Deriving Triplet–Triplet Exciton Couplings for Density Functional Tight-Binding

Kevin Mächtel (kevin.maechtel@kit.edu)
Karlsruhe Institute of Technology, Germany

Triplet–triplet exciton interactions are central to multi-exciton processes in organic semiconductors, yet their quantitative description remains challenging for large molecular systems. In this work, we focus on establishing and validating a density functional tight-binding (DFTB)-based framework for computing triplet–triplet exciton couplings. First, we benchmark DFTB triplet excitation energies and relevant electronic properties against higher-level electronic-structure methods for representative organic semiconductor molecules and dimers. This provides a systematic assessment of the accuracy and transferability of the underlying DFTB parametrization for triplet states. On this basis, we derive a working expression for the effective triplet–triplet coupling that separates Coulomb and exchange-like contributions, formulated in terms of DFTB quantities such as transition densities or state-specific density matrices. The resulting scheme is designed to be computationally efficient while retaining a clear physical interpretation of the individual coupling components. These benchmarking and derivation steps lay the groundwork for future applications of DFTB to multi-exciton phenomena in extended organic semiconductor systems, such as triplet–triplet annihilation and singlet fission.

Stand № 24

Orbital-Free Density Functional Theory Powered by Machine Learning

Marc Ickler (marc.ickler@iwr.uni-heidelberg.de)
Heidelberg University, Germany

Density functional theory is central to computational chemistry and materials science, but its most accurate practical formulations remain computationally expensive. Orbital-free density functional theory offers a promising alternative by predicting electronic energies directly from the electron density, though achieving sufficient accuracy has remained challenging. In this work, we build on our earlier paper, *Stable and Accurate Orbital-Free Density Functional Theory Powered by Machine Learning*, and show that a newer machine-learning architecture combined with improved data generation and training strategies yields better results than our previous model. We also take first steps toward periodic systems, bringing machine-learned orbital-free density functional theory closer to large-scale materials applications.

Stand № 25

Computational Design and AI-guided Optimization of the Lanthanide Binding de novo TFD-EE Scaffold

Ghulam Mustafa (g.mustafa@tum.de)
Technical University of Munich, Germany

De novo protein design has gained momentum in recent years with the advancement of new computational algorithms and artificial intelligence. These de novo designed proteins are further engineered to incorporate metal cofactors and can be developed into biocatalysts for new-to-nature reactions[1,2]. Here, we present the original Rosetta based de novo designed TIM-Ferredoxin (TFD) dimer with 4 GLU residues located in the center of TIM-barrel that shows tight binding to different lanthanides[3] and compare it with the deep learning based MPNN sequence redesign[4]. We have used classical molecular dynamics (MD) simulations in combination with biophysical/biochemical characterization, nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography experiments to study different variants of a de novo designed lanthanide binding TFD scaffolds. In this work, we aim to understand the structure, function and dynamics of different TFD variants with and without metal. The MPNN redesign structures showed improved dimer assembly, high-affinity towards metal and better enzyme catalysis. Combined NMR and MD simulations have been useful to distinguish the dynamic differences in different scaffolds.

Stand № 26

---

[1] Klein, A. S. & Zeymer, C. Protein Engineering, Design and Selection 34, (2021).

[2] Klein, A.S. et al. Journal of the American Chemical Society. 146, 25976-25985 (2024).

[3] Caldwell, S. J. et al. Proceedings of the National Academy of Sciences 117, 30362–30369 (2020).

[4] Dauparas, J., Anishchenko, et.al. Science 378, 49-56 (2022).

Simulating Large Scale Conformational Changes in Histidine Kinase

Leonie Pfeffer (leonie.pfeffer@kit.edu)
Karlsruhe Institute of Technology, Germany

Histidine kinases (HKs) are essential two-component signaling proteins in microorganisms. As they are absent in eukaryotic cells, they represent promising antibacterial drug targets. A cascade of conformational changes culminates in autophosphorylation and signal transfer to a response regulator, triggering adaptive cellular responses. However, these transitions occur on millisecond timescales, making them inaccessible to conventional molecular dynamics (MD) simulations. Enhanced sampling methods like metadynamics could probe these mechanisms but require robust collective variables (CVs)—a major challenge, since suitable CVs are often non-trivial to identify. We address this through coarse-grained (CG) Gō-Martini simulations, which accelerate conformational sampling by orders of magnitude. A conceivable approach would be to introduce custom dual-basin contact potentials that stabilize both states, thereby enabling their mutual sampling and generating transition trajectories. Principal component analysis (PCA) of these CG input trajectories could then identify reaction coordinates that adequately cover both states. These CVs could subsequently be transferred to all-atom representations for refined simulations.

Stand № 27

Machine-Learned Collective Variables for the Thiol-Disulfide Exchange in Glutaredoxin

Maike Vogelbacher (maike.vogelbacher@kit.edu)
Karlsruhe Institute of Technology, Germany

Thiol-disulfide exchange in proteins occurs on a timescale much shorter than the slow conformational motions that control active-site hydration. Conventional Metadynamics using only a reaction-based collective variable fails to converge because it cannot capture the configurational transition that opens or closes a water channel. However, finding a CV that describes this configurational change is highly non-trivial. We therefore trained a non-linear autoencoder on QM/MM snapshots to extract a machine-learned collective variable describing both these slow protein motions and the reaction. This collective variable is used in a Metadynamics simulation, which provides new input for the autoencoder, until the free energy surface converges. This workflow was implemented and tested so far on a small model system for the thiol-disulfide exchange, with the aim to transfer this workflow to proteins.

Stand № 28

Machine Learning Driven Advancement of Density Functional Tight Binding Method (DFTB)

Frederic Teich (frederic.teich@kit.edu)
Pascal Schmidt (pascal.schmidt@kit.edu)

Karlsruhe Institute of Technology, Germany

This project aims to advance the Density Functional Tight Binding (DFTB) method by improving both its theoretical foundations and practical accuracy. Machine learning potentials will be integrated to enhance the predictive accuracy of DFTB toward the level of full Density Functional Theory (DFT) while maintaining its inherent computational efficiency. Together, these developments seek to bridge the gap between speed and accuracy in atomistic simulations, enabling more reliable and scalable modeling of complex molecular systems.

Stand № 29

Energy Landscape of a Photosystem II Subcomplex Revealed by Multiscale Machine-Learning Simulations

Joel Santos (joel.santos@dcci.unipi.it)
Contributing: Chris John, Lorenzo Cupellini, Benedetta Mennucci

University of Pisa, Italy

In higher plants, Photosystem II captures solar energy through an extended network of light-harvesting complexes that funnel excitation energy to the reaction centre where charge separation occurs. The LHCII–CP29–CP24 pentamer (M-CP2) is a subcomplex of Photosystem II whose 67 chlorophylls (Chls) mediate excitation energy transfer from the peripheral antenna to the reaction centre. Characterising the excitonic structure of such a large multichromophoric assembly requires an accurate calculation of excitation energies in the protein environment, accounting for its structural disorder. We address this challenge with a combination of microsecond molecular dynamics (MD) and a recently developed machine-learning framework, which reproduces the excitation energies of pigments computed with hybrid quantum mechanics/polarizable molecular mechanics (QM/MMPol) methods. We demonstrate that the polarizable description is essential to capture assembly-induced changes in the excitation energies of Chls. Our simulations reveal that a single chlorophyll (Chl a611) in one LHCII monomer undergoes a ~245 cm^{-1} red-shift upon pentamer formation, driven entirely by the collective electrostatic potential of CP29 and CP24 rather than structural distortion or individual residues. This generates a deeper energy trap in the a610–a611–a612 terminal emitter site at the interface between LHCII and the minor antenna complexes, favoring a net energy flow towards the reaction center.

Stand № 30

Author Index