3rd SIMPLAIX Workshop on "Machine Learning for Multiscale Molecular Modeling"

7 - 9 May 2025
Studio Villa Bosch Heidelberg, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg
https://simplaix-workshop2025.h-its.org/

Abstract Book

Organizing Committee: Daniel Sucerquia (HITS), Rostislav Fedorov (HITS), Jonathan Teuffel (HITS), David Hoffmann (KIT), Prof. Rebecca Wade (HITS), T.T.-Prof. Pascal Friederich (KIT), Prof. Marcus Elstner (KIT), Prof. Tristan Bereau (Heidelberg University)

Administrative Team: Britta Schläger, Angela Michel, Christina Blach, Marisa de Sá Almeida (HITS)

Contact
+49 6221 533 329
events@h-its.org

Bluesky: #SIMPLAIX25

Organizers:

HITS Heidelberg Institute for Theoretical Studies

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

KIT Karlsruher Institut für Technologie

Sponsors:

Klaus Tschira Stiftung

GRK 2450

Program

Wednesday, 07 May 2025

13:00 Registration

14:00 Opening and Welcome

Session 1
Chair: Rebecca Wade

14:15 [Talk 1] Modesto Orozco – "Nucleic acids in the frontier between AI and simulation" – (page 8)

14:55 [Short talk] Patrizia Mazzeo – "Excited-state dynamics of solvated molecules with multiscale machine learning" – (page 9)

15:15 [Short talk] Abhik Ghosh Moulick – "Towards Understanding of Chromatin Folding: The Role of Protein Interactions for Stability of Nucleosomes" – (page 10)

15:45 Coffee break

Session 2
Chair: Ullrich Köthe

16:30 [Talk 1] Carolin Müller – "Machine Learning in Photochemistry – Data is Key" – (page 11)

17:10 [Talk 2] Johannes Kästner – "Transferable and Uniformly Accurate Interatomic Potentials" – (page 12)

17:50 [Short talk] Leif Seute – "Learning conformational ensembles of proteins based on backbone geometry" – (page 13)

18:15 Picture speakers

Thursday, 08 May 2025

Session 3
Chair: Frauke Gräter

09:00 [Talk 1] Arne Elofsson – "Towards a Complete Structural Map of the Human Proteome Using AlphaFold" – (page 14)

09:40 [Talk 2] Matteo dal Peraro – "A Structure Transformer for Structural Biology and Molecular Design" – (page 15)

10:20 [Short talk] Sergio Suarez Dou – "Machine Learning Force Field modelling for quantum accuracy in biomolecule dynamics" – (page 16)

10:40 Coffee break

Session 4
Chair: Alice Allen

11:20 [Talk 1] Stefan Grimme – "g-xTB: DFT accuracy at tight-binding speed" – (page 17)

12:00 [Short talk] Oleksandra Kukharenko – "Utilizing generative machine learning models to improve determinination of glass transition in polymer melts" – (page 18)

12:20 Discussion session. Chair: Alice Allen

13:00 Lunch

14:20 Group picture

Session 5
Chair: Marcus Elstner

14:30 [Talk 1] Shirin Faraji – "On-the-fly hybrid quantum/classical dynamics in complex environment" – (page 19)

15:10 [Talk 2] Sandra Luber – "Excited states dynamics and beyond" – (page 20)

15:50 [Short talk] Henrik Schopmans – "Temperature-Annealed Boltzmann Generators" – (page 21)

16:10 Coffee break

Session 6
Chair: Andreas Dreuw

16:40 [Talk 1] Marc van der Kamp – "EMLE: Electrostatic Machine-Learned Embedding for accurate and efficient ML/MM simulations of enzymes and other biomolecules" – (page 22)

17:20 [Short talk] Sarah Bernart – "Machine Learning-Driven Insights into Active Species and Reaction Dynamics in Pd and Pt Catalysts Supported on Ceria" – (page 23)

17:45 Poster session

19:00 Workshop Dinner + Poster session (Studio)

Friday, 09 May 2025

Session 7
Chair: Tristan Bereau

09:00 [Talk 1] Lukas Stelzl – "Dynamic self organization of proteins in the cell nucleus" – (page 24)

09:40 [Talk 2] Elsa Sánchez-García – "Combining Machine-Learning and Physics-Based Approaches for Computer-Aided Drug Design and Protein Engineering" – (page 25)

10:20 [Short talk] Fabian Grünewald – "From CGsmiles to multiresolution GNNs for chemical space exploration" – (page 26)

10:40 Coffee break

Session 8
Chair: Pascal Friederich

11:10 [Talk 1] Antonia Mey – "From generative modelling for fragment-based drug design to property prediction based on large-language models" – (page 27)

11:50 [Short talk] Luis Walter – "Navigating Chemical Space: An Active Learning Strategy Using Multi-Level Coarse-Graining" – (page 28)

12:10 Discussion round table and round up – Chairs: Anya Gryn'ova; Tristan Bereau

13:00 Lunch and end of Workshop.

Talks

Nucleic Acids in the Frontier Between AI and Simulation.

Prof. Modesto Orozco (modesto.orozco@irbbarcelona.org)
Institute for Research in Biomedicine, Barcelona (BCN), España

Nucleic acids are large polymers with a complex physicochemical nature which makes challenging their simulation. In this context, AI techniques have emerged as a potential game-changing player, but how much real impact AI will have in the field is unclear. I will present a few studies done in Barcelona showing how AI and simulation methods can integrate to provide new information on nucleic acids systems.

Excited-state Dynamics of Solvated Molecules with Multiscale Machine Learning

Patrizia Mazzeo (patrizia.mazzeo@phd.unipi.it)
Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Pisa, Italy

Photoinduced processes typically occur in condensed phase, where the environment significantly influences the reaction outcome. This interplay is usually addressed using a quantum mechanical (QM) description of the chromophore conveniently coupled with a classical molecular mechanics (MM) representation of the environment. However, the computational cost of QM/MM methodologies remains a major challenge for molecular dynamics (MD), limiting accurate statistical sampling especially for excited states. Machine learning (ML) for potential energy surfaces has proven highly effective in accelerating ab initio MD simulations, but mainly focusing on isolated systems. Here, we present a computationally efficient ML/MM strategy for both ground- and excited-state simulations, designed to emulate ab initio electrostatic embedding QM/MM MD. Energies and forces are estimated as the sum of a vacuum contribution and an environment-induced shift, respectively accounting for the QM part's internal geometry and the polarization effect from external MM charges. This physically constrained model ensures the correct treatment of electrostatics in ML/MM and allows extrapolation to different environments. We apply our ML/MM strategy to the prototypical excited-state intramolecular proton transfer (ESIPT) of 3-hydroxyflavone in two solvents, effectively capturing the environmental effects on proton transfer dynamics in agreement with experiments.
Patrizia Mazzeo et al. Digit. Discov. 2024, 3, 12.

Towards Understanding of Chromatin Folding: The Role of Protein Interactions for Stability of Nucleosomes

Dr. Abhik Ghosh Moulick (abhik.moulick@kit.edu)
Karlsruhe Institute of Technology, Karlsruhe, Germany

Nucleosome dynamics governs DNA accessibility and chromatin function; however, studying their large-scale motions requires scalable computational approaches. This work integrates atomistic and coarse-grained (CG) molecular dynamics (MD) simulations using the SIRAH force field to first validate CG modeling for canonical nucleosomes and then unravel stability mechanisms in centromeric nucleosomes. For canonical systems (alpha satellite and Widom-601 DNA), CG simulations reproduce structural metrics (DNA base pair parameters) from atomistic benchmarks while capturing enhanced DNA end fluctuations and broader conformational sampling. Principal component analysis (PCA) reveals multiple free energy minima in CG trajectories, highlighting its utility for probing large-scale dynamics. For non-canonical centromeric nucleosomes, which contain CENP-A protein instead of H3, the employment of CG MD simulations improves the understanding of structural flexibility of nucleosomes and impact of regulating proteins, e. g., CENP-N on the nucleosome stability. This study establishes the SIRAH force field as a powerful tool for dissecting centromeric chromatin dynamics, laying the groundwork for multiscale models for chromatin folding.

Machine Learning in Photochemistry – Data is Key

Prof. Dr. Carolin Müller (carolin.cpc.mueller@fau.de)
FAU Erlangen-Nürnberg, Erlangen, Germany

Photochemistry has significant potential for sustainable chemistry but is limited by the lack of comprehensive design rules and detailed information about excited-state molecular structures. To explore molecular structure-property relationships in photoexcited processes, quantum chemical simulations are essential. However, their accuracy is constrained by the high computational costs of these methods, which limit their application to small systems (e.g., $\leq$100 atoms) and short timescales (fs to ps). Recently, machine learning techniques, such as PyRAI2MD, SchNarc, and SPaiNN, have shown promise in accelerating photodynamics simulations without sacrificing accuracy. Despite these advancements, several challenges remain: models are confined to the chemical space defined by the training data, and simulations over extended timescales require active learning loops, which result in larger training sets and increased computational demands. In this talk, I will focus on SPaiNN, a machine learning potential designed to accelerate non-adiabatic molecular dynamics simulations. I will discuss the importance of structuring training sets for optimal performance and highlight recent efforts to create maximal informative databases of photochemical reactions, specifically designed for excited-state learning, to further enhance model accuracy and efficiency.

Transferable and Uniformly Accurate Interatomic Potentials

Prof. Johannes Kästner (kaestner@theochem.uni-stuttgart.de)
University of Stuttgart, Stuttgart, Germany

The development of machine-learned interatomic potentials requires generating sufficiently expressive atomistic data sets. Active learning algorithms select data points on which labels, i.e., energies and forces, are calculated for inclusion in the training set. However, for batch mode active learning, i.e., when multiple data points are selected at once, conventional active-learning algorithms can perform poorly. Therefore, we investigate algorithms specifically designed for this setting and show that they can outperform traditional algorithms. We investigate selection based on the resulting training set's informativeness, diversity, and representativeness. We propose using gradient features specific to atomistic neural networks to evaluate the informativeness of queried samples, including several approximations allowing for their efficient evaluation. To avoid selecting similar structures, we present several methods that enforce the diversity and representativeness of the selected batch. Furthermore, we use transfer learning to improve the quality of the resulting potential, use training data from cluster calculations to predict bulk properties, and present a scheme to learn tensorial quantities, like magnetic anisotropy.

Learning Conformational Ensembles of Proteins Based on Backbone Geometry

Leif Seute (leif.seute@h-its.org)
Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

Deep generative models have recently been proposed for sampling protein conformations from the Boltzmann distribution, as an alternative to often prohibitively expensive Molecular Dynamics simulations. However, current state-of-the-art approaches rely on fine-tuning pre-trained folding models and evolutionary sequence information, limiting their applicability and efficiency, and introducing potential biases. We propose a flow matching model for sampling protein conformations based solely on backbone geometry. We introduce a geometric encoding of the backbone equilibrium structure as input and propose to condition not only the flow but also the prior distribution on the respective equilibrium structure, eliminating the need for evolutionary information. The resulting model is orders of magnitudes faster than current state-of-the-art approaches at comparable accuracy and can be trained from scratch in a few GPU days. In our experiments, we demonstrate that the proposed model achieves competitive performance with reduced inference time, across not only an established benchmark of naturally occurring proteins but also de novo proteins, for which evolutionary information is scarce.

Towards a Complete Structural Map of the Human Proteome Using AlphaFold

Prof. Arne Elofsson (arne@bioinfo.se)
Stockholm University, Solna, Sweden

Cellular functions are governed by molecular machines that assemble through protein-protein interactions. Their atomic details are critical to studying their molecular mechanisms. Today the structure of virtually all individual proteins is available from predictions using AlphaFold. However, these predictions are limited to individual chains and do not include interactions. In this talk I will describe our attempts to increase the structural coverage of protein-protein interactions. Today fewer than 5

A Structure Transformer for Structural Biology and Molecular Design

Matteo Dal Peraro (matteo.dalperaro@epfl.ch)
EPFL, Lausanne, Switzerland

Proteins are the fundamental building blocks of life, dictating the spatial and temporal occurrence of most biological functions. They have evolved to specifically interacting with other proteins, nucleic acids, metabolites, membranes, in order to form molecular complexes whose structural architecture is at the basis of any biological function. Predicting these specific protein binding interfaces remains however a major challenge. To tackle this task, we introduced a geometric transformer named Protein Structure Transformer (PeSTo at https://pesto.epfl.ch) that acts directly on protein atoms labelled with nothing more than element names. This deep learning approach demonstrated to surpass existing methods in accurately predicting protein-protein interfaces and to distinguish interfaces with nucleic acids, lipids, carbohydrates, ions and small molecules with high confidence. The low computational cost of this method enables processing high volumes of structural data, such as molecular dynamics ensembles and growing protein model databases, providing new means to discover unexplored biology. Moreover, building on the same architecture we introduced a new framework called CARBonAra (namely, Context-aware Amino acid Recovery from Backbone Atoms heteroatoms), a new protein sequence generator model trained on structural data available in the PDB. CARBonAra performs on par with state-of-the-art methods like ProteinMPNN and ESM-IF1 for sequence prediction of isolated proteins or protein complexes. However, and most importantly, the model has the ability to perform protein sequence prediction conditioned by any specific non-protein molecular context, which contributes to significantly improve sequence recovery. On in vitro testing, this approach was able to predict protein designs showing structural stability and enzymatic activity. This new concept is anticipated to improve the design versatility for engineering proteins with desired functions.

Machine Learning Force Field Modelling for Quantum Accuracy in Biomolecule Dynamics

Sergio Suárez Dou (sergio.suarezdou@uni.lu)
University of Luxembourg, Esch-Sur-Alzette, Luxembourg

Molecular dynamics (MD) is a computational method for modeling molecular behavior at the atomic level, providing insights into physical and chemical properties. Classical Force Fields (FF) have been essential in computational chemistry, simplifying molecular interactions into bonded and non-bonded terms. Recent advances in machine learning have led to the development of machine learning force fields (MLFFs), which offer quantum-level accuracy and efficiency. SO3LR, a foundational MLFF, integrates physics-based functions for comprehensive modeling, dynamically adjusting interactions for more accurate representations. MLFF accurately predicted experimental IR spectra and peptide folding. In water, it successfully modeled properties like radial distribution, density, diffusion, and proton transfer. For proteins in water, including crambin and ubiquitin, the MLFF was validated against NMR data. Finally, to test how it captures the collective modes of motion, explaining the stability of multi-chain proteins, it was tested with the p53 tetramer. This comprehensive testing highlights the MLFF's versatility and accuracy across different environments and molecular systems, demonstrating its potential for broad applications in computational chemistry and molecular simulations.

g-xTB: DFT Accuracy at Tight-Binding Speed

Prof. Dr. Stefan Grimme (grimme@thch.uni-bonn.de)
Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, Germany

g-xTB: DFT accuracy at tight-binding speed Stefan Grimme Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstrasse 4, 53115 Bonn, Germany I will present our third-generation tight-binding model g-xTB (g=general). This includes nonlocal Fock-exchange as well as other new, many-center Hamiltonian terms (e.g., atomic correction potentials, ACP). Another new ingredient is the adaptive q-vSZP minimal AO basis set[1] which provides in typical DFT applications results of about or better than DZ quality. The "breathing" of the AOs in the molecular environment is parameterized efficiently by on-the-fly computed effective atomic charges obtained by a new EEQ-bond-capacitor charge model[2] and coordination numbers. g-xTB aims at general purpose applicability in chemistry and more closely approaches DFT accuracy (actually ωB97M-V/aTZ[3]) than previous semi-empirical methods at only slightly increased computational cost (factor of 1.5 compared to GFN2-xTB). It will be consistently available for all elements Z=1-103 with f-electrons included for lanathanides/actinides. The talk describes a fresh look at the underlying TB theory as well as extensive benchmarking on a wide range of standard thermochemistry sets. [1] M. Müller, A. Hansen, S. Grimme, J. Chem. Phys. 159 (2023), 164108. Revision: JPC A, doi:10.1021/acs.jpca.4c06989 [2] T. Froitzheim, M. Müller, A. Hansen, S. Grimme, in preparation. [3] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. 144 (2016), 214110

Utilizing Generative Machine Learning Models to Improve Determination of Glass Transition in Polymer Melts

Dr. Oleksandra Kukharenko (kukharenko@mpip-mainz.mpg.de)
Max Planck Institute for Polymer Research, Mainz, Germany

Advances in computing power, computational methods, and simulation techniques open unique opportunities to computationally study (bio-)polymers across a wide range of length and time scales. Consequently, this has led to the necessity of developing and applying new methods to analyze, group, and ultimately explain processes derived from such data. In my talk, I will focus on the application of machine learning techniques to study the properties of polymer melts. Recently we proposed data-driven approach for the precise definition of the glass transition temperature of coarse-grained, weakly semi-flexible polymers from simulation trajectories. Instead of concentrating on the average properties of the melt, we utilize the high-resolution details accessible through molecular dynamics simulations and consider the structural information of single chains. Our data-driven method can determine the glass transition temperature with greater accuracy than classical approaches. I will demonstrate how the use of generative machine learning models can enable even higher accuracy with finer temperature resolution as well as helps to reduce simulation costs.

On-the-fly Hybrid Quantum/classical Dynamics in Complex Environment

Prof. Dr. Shirin Faraji (shirin.faraji@hhu.de)
Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Light-triggered processes, ubiquitous in nature and technology, are inherently quantum. Phenomena such as the photovoltaic effect, charge migration, and proton-coupled electron transfer require a quantum mechanical description. Understanding these processes is crucial for advancing technologies like optogenetics, photopharmacology, and photoresponsive materials. However, accurately modeling these processes in (supra)molecular systems remains challenging due to (i) the need for high-level electronic structure calculations, (ii) coupled electron-nuclear dynamics, and (iii) the importance of the environment. Recent years have seen significant growth in direct dynamics approaches, such as semiclassical trajectory surface-hopping and on-the-fly quantum dynamics within variational multi-configuration Gaussian. Particularly troublesome for trajectory-based methods is the huge amount of electronic structure calculations that need to be performed. Here, we present an innovative solution that integrates semiclassical and quantum direct dynamics with a database-driven approach, drastically reducing the number of electronic structure calculations by employing machine-learning algorithms and methods borrowed from the realm of artificial intelligence. Additionally, on-the-fly direct dynamics can be embedded into a quantum mechanics/molecular mechanics framework to explicitly account for environment effects. The degree of sophistication achievable with this implementation is demonstrated using test systems.

Excited States Dynamics and Beyond

Sandra Luber (sandra.luber@uzh.ch)
University of Zurich, Zürich, Switzerland

I will give an overview about selected parts of our recent research. Focus will be on our work for excited states such as the deltaSCF method and excited state dynamics for the condensed phase. Other research has concerned the use of machine learning, e.g. for learning correlation energies, spectroscopic signatures or collective variables, the latter being essential ingredients for enhanced sampling calculations auch as metadynamics.

Temperature-annealed Boltzmann Generators

Henrik Schopmans (henrik.schopmans@kit.edu)
Institute of Theoretical Informatics, Karlsruhe, Germany

Efficient sampling of unnormalized probability densities such as the Boltzmann distribution of molecular systems is a longstanding challenge. Next to conventional approaches like molecular dynamics or Markov chain Monte Carlo, variational approaches, such as training normalizing flows with the reverse Kullback-Leibler divergence, have been introduced. However, such methods are prone to mode collapse and often do not learn to sample the full configurational space. Here, we present temperature-annealed Boltzmann generators (TA-BG) to address this challenge. First, we demonstrate that training a normalizing flow with the reverse Kullback-Leibler divergence at high temperatures is possible without mode collapse. Furthermore, we introduce a reweighting-based training objective to anneal the distribution to lower target temperatures. We apply this methodology to three molecular systems of increasing complexity and, compared to the baseline, achieve better results in almost all metrics while requiring up to three times fewer target energy evaluations. For the largest system, our approach is the only method that accurately resolves the metastable states of the system.

EMLE: Electrostatic Machine-Learned Embedding for Accurate and Efficient ML/MM Simulations of Enzymes and Other Biomolecules

Dr. Marc Van Der Kamp (marc.vanderkamp@bristol.ac.uk)
University of Bristol, Bristol, United Kingdom

Combined quantum mechanics and molecular mechanics (QM/MM) can accurately describe enzyme reactions, for example, but simulations are limited to short timescales due to computational cost. This could be solved by using machine-learning (ML) atomic potentials that offer QM accuracy at a fraction of the cost. However, due to the absence of electrons, ML potentials are unable to describe the electrostatic interaction between ML and MM regions, which is crucial for capturing enzyme catalysis. We have developed the "electrostatic ML embedding" (EMLE) scheme that solves this issue, predicting the effect of the MM environment from the atomic properties of the QM system in vacuo. The EMLE scheme [1] thus enables ML/MM simulations with arbitrary ML potentials using existing QM/MM codes (https://chemle.github.io/emle-engine/). Here, we demonstrate that this allows to simulate various condensed-phase processes at DFT/MM level accuracy with only a fraction of the associated computational cost. The examples include two on enzyme reactions, showing how ML(EMLE)/MM can distinguish reactive binding poses in a Diels-Alderase and accurately captures the catalytic effect of chorismate mutase. In future, the use of ML(EMLE)/MM should allow efficient and accurate computational screening of enzyme variants for desired reactions. Reference [1] Zinovjev et al., JCTC 2024, 20, 4514.

Machine Learning-Driven Insights into Active Species and Reaction Dynamics in Pd and Pt Catalysts Supported on Ceria

Sarah Bernart (s.w.bernart@tue.nl)
TU/e, Karlsruhe, Germany

The catalytic activity of heterogeneous catalysts in combustion engines is highly sensitive to variations in their chemical state. The effectiveness of CO oxidation and O_2 dissociation relies not only on temperature and pressure but also on precisely tailoring the catalyst's chemical environment at the sub-nanometer scale. To systematically explore these complex interactions, we integrate machine learning (ML) models, specifically Artificial Neural Networks (ANNs), into the analysis of first-principles datasets. By leveraging a comprehensive dataset derived from Density Functional Theory (DFT) calculations, microkinetic modeling (MKM), and ab initio molecular dynamics, ANNs can efficiently process and predict catalytic behaviors. By systematically structuring and training ML models on theoretical and experimental datasets, we provide a predictive approach that bridges the gap between computationally intensive simulations and real-world catalyst optimization. This data-driven methodology enables an accelerated identification of optimal catalytic properties, reducing the reliance on exhaustive first-principles calculations while enhancing predictive accuracy in heterogeneous catalysis.

Dynamic Self Organization of Proteins in the Cell Nucleus

Prof. Dr. Lukas Stelzl (lstelzl@uni-mainz.de)
Johannes Gutenberg University Mainz, Mainz, Deutschland

In the nucleus of eukaryotic cells, multiple layers of control shape how genes are regulated and switched on and off. An important layer of control is provided by multivalent interactions of proteins, frequently involved disordered regions. The disordered C-terminal domain (CTD)of RNA polymerase II is a paradigmatic example of how disordered regions of proteins interact in the nucleus. I will discuss the impact of residual structure in the CTD. Coarse-grained molecular dynamics simulations show the CTD phase separates to form hubs for the transcription of genes. Intriguingly, the interactions preferences of CTD can be captured with simple neural networks and this holds the promise to elucidate regulatory networks in gene regulation in the nucleus.

Combining Machine-Learning and Physics-Based Approaches for Computer-Aided Drug Design and Protein Engineering

Prof. Dr. Elsa Sánchez-García
(professors.cbe.bci@tu-dortmund.de)
Lehrstuhl für Computational Bioengineering, Fakultät Bio- und Chemieingenieurwesen, Technische Universität Dortmund, Dortmund, Germany

Protein–protein interactions (PPIs) are involved in most biological processes. However, the efficient and reliable prediction of PPIs from the proteome remains a challenge given the vast search space involved. To address this challenge, we combine machine leaning (ML) tools with physics-based approaches. We developed PPI-Detect[1] as a sequence-based ML model for the prediction of the likelihood of protein– protein interactions. PPI-Detect allows exploring the peptidome for identifying protein partners of a given target. We also implemented PPI-Affinity[2] for the prediction of the binding affinity (BA) of protein–protein complexes and protein engineering applications. The Central Limit Free Energy Perturbation approach (CL-FEP)[3] allows evaluating the FEP identity directly from the energy samples of the end states of a system transformation, without fitted parameters or stratification. The CL-FEP approach delivered excellent accuracy estimating the binding Gibbs energy of a large set of benchmark systems and found real-life applications.[4,5,6,7] In this talk, I will briefly introduce these tools and discuss selected applications to the discovery and optimization of bioactive compounds as well as for protein engineering.

[1] Journal of Computational Chemistry 2019, 40 (11), 1233-1242, DOI: 10.1002/jcc.25780
[2] Journal of Proteome Research 2022, 21, 1829-1841. DOI: 10.1021/acs.jproteome.2c00020
[3] Journal of Chemical Theory and Computation 2020 6 (3), 1396-1410. DOI: 10.1021/acs.jctc.9b00725
[4] Nature Communications 2021 12 (1726). DOI: 10.1038/s41467-021-21972-0
[5] Cell Chemical Biology 2021 28, 1-11. DOI: 10.1016/j.chembiol.2021.03.013
[6] ChemBioChem 2022, e202100618. DOI: 10.1002/cbic.202100618
[7] ChemBioChem 2022, 23, e20220039, DOI: 10.1002/cbic.202200396

From CGsmiles to Multiresolution GNNs for Chemical Space Exploration

Dr. Fabian Grünewald (fabian.gruenewald@h-its.org)
HITS, Heidelberg, Germany

Coarse-grained (CG) models simplify molecular representations by grouping multiple atoms into effective particles. This approach is typically used to speed up classical particle-based simulations. However, it has been proposed that the CG resolution itself can be used in machine learning models for more effective exploration of chemical compound space (CCS). Yet, such methods have hardly advanced beyond the proof of principle stage. One major bottleneck is the lack of methods and data formats to generate and handle the crucial aspect of describing how the atoms are grouped (i.e., the mapping). To address this problem, we developed CGsmiles, a line notation that allows the representation of multiple resolutions of the same molecule in one string. Utilizing the capabilities of CGsmiles, we compiled the largest database of mappings within the popular Martini 3 coarse-grained force field and designed a program to generate mappings with high fidelity Interestingly, the produced mappings challenge the notion that CG models represent multiple molecules using the same mapping, thereby limiting the CCS - one of the central paradigms in CG CCS exploration. Instead of relying on the reduction of the CCS through the CG mapping, we propose to use the inductive bias of the CG model interactions directly in a multiresolution Graph Neural Network.

From Generative Modelling for Fragment-Based Drug Design to Property Prediction Based on Large-Language Models

Dr. Antonia Mey (antonia.mey@ed.ac.uk)
University of Edinburgh, Edinburgh, United Kingdom

Proteins are central to most biological processes, and understanding their function and interactions is crucial for advancing our knowledge of life and disease regulation. Small molecules can modulate protein activity, often by binding and inhibiting their function. However, identifying new molecules and developing fast and reliable computational methods to predict their binding efficiency and other desirable properties required to make them suitable drug-like molecules remains a significant challenge. Computational models can assist in tasks such as determining ligand poses—traditionally through docking algorithms or, more recently, co-folding methods—as well as estimating binding free energy and other relevant properties. I will introduce our approach to making use of refinement of generative models for small molecules to elaborate X-ray fragment hits to drug-like candidate molecules. To then be able to predict the newly generated molecules binding affinity to their target protein I will discuss different strategies from simulation-based approaches to leveraging large language-based methods highlighting successes and caveats.

Navigating Chemical Space: An Active Learning Strategy Using Multi-Level Coarse-Graining

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

Exploring the vast chemical compound space remains a significant challenge due to the immense number of possible molecules and limited scalability of conventional screening methods. To approach chemical space exploration more effectively, we have developed an active learning-based method that uses transferable coarse-grained models to compress chemical space into varying levels of resolution. By using multiple representations of chemical space with different coarse-graining resolutions, we balance combinatorial complexity and chemical detail. To identify target compounds, we first transform the discrete molecular spaces into smooth latent spaces. We then perform Bayesian optimization within these latent spaces, using molecular dynamics simulations to calculate target free energies of the coarse-grained compounds. This multi-level approach allows for an effective balance between exploration at lower and exploitation at higher resolutions. We demonstrate the effectiveness of our method by optimizing molecules to enhance phase separation in phospholipid bilayers. Our funnel-like strategy not only suggests optimal compounds, but also provides insight into relevant neighborhoods in chemical space. We show how this neighborhood information from lower resolutions can be used to guide the optimization at higher resolutions, thereby providing an efficient way to navigate large chemical spaces for free energy-based molecular optimization.

Posters

AI-Driven Peptide Design for Modulating the Phase Behavior of Protein Condensates

Kumar Gaurav (kugaurav@uni-mainz.de)
Institute of Molecular Biology, Mainz, Germany

Biomolecular condensates, formed through phase separation, help organize cellular components without membranes. C. elegans, the proteins MUT-8 and MUT-16 are key components of Mutator foci, a specialized condensates involved in RNA silencing. While these structures are essential for regulating gene expression, it remains unclear how the sequences of binding partners like MUT-8 and its mutants modulate the material properties of the MUT-16 condensate. To address this, we use coarse-grained molecular dynamics simulations to model and engineer the prion-like domain (PLD) of MUT-8, which interacts with the scaffold protein MUT-16. We apply an active learning framework, Wazy (Yang et al., 2022), which integrates Bayesian optimization and neural networks to iteratively guide the selection of the most informative protein variants for simulation. This reduces the number of simulated protein variant required while enabling the model to learn sequence–property relationships that govern condensate behavior. Our workflow unfolds in two stages: first, we optimize MUT-8's binding affinity to MUT-16; next, we fine-tune the condensate's internal dynamics, including chain mobility and molecular diffusivity. We use large language models (ProtGPT2) to generate a diverse and unbiased validation set for assessing generalizability. This physics-based, learning-guided framework offers a computational approach to understanding how disordered protein sequences influence condensate properties, supporting rational design strategies for synthetic condensates with potential applications in biological systems.

Stand № 1

ART-SM: A Fragment- and ML-based Approach to Backmapping

Christian Pfaendner (christian.pfaendner@simtech.uni-stuttgart.de)
University of Stuttgart - SimTech, Stuttgart, Germany

Multiscale MD simulations depend on accurate backmapping from coarse-grained to atomistic resolution. While generative machine-learning approaches have shown promise for proteins, their application to other molecular systems is hindered by the lack of sufficient training data. To address this, we developed a fragment-based method that leverages classical regression to identify optimal fragment conformations for a given coarse-grained structure. These fragments are then assembled using a novel optimization algorithm. The resulting structures require only minimal energy minimization and position restraint simulations for relaxation. In our tests, ART-SM demonstrated significant improvements in runtime and dihedral angle distributions compared to the widely used method Backward.

Stand Nº 2

Brownian Dynamics for Studying the Protein-Ligand Association Process in a Crowded Environment

Riccardo Beccaria (riccardo.beccaria@h-its.org)
HITS, Heidelberg, Germany

Intracellular crowding plays a crucial role in proteins behaviour, influencing their motion, the cell's kinetic and equilibrium properties. In this work, we present an adaptive model to simulate the association process between two biomolecules in a crowded environment by Brownian dynamics, and thereby compute the bimolecular association rate constant. In this model, interactions are represented with atomistic resolution when crowders are nearby and with a coarser force field when they are farther away. We applied this model to two contrasting protein-ligand systems with varying crowder types and densities. Markov State modelling of the encounter trajectories was used to analyse the binding pathways. The results reveal competing effects resulting in non-monotonic dependence of the association rate on crowder concentration. While the presence of the crowders can hinder binding by reducing the translational diffusion of the molecules in the medium and obstruction of binding sites, the crowders can also enhance association via caging, channeling, and other mechanisms. We are currently working on incorporating physics-informed machine learning to improve the calculation of intermolecular forces, further refining the accuracy of our simulations.

Stand Nº 3

Computational Tools for Biomolecular Systems: From Protein Engineering to Biocatalysis

Dr. Joel Mieres Perez (joel.mieresperez@tu-dortmund.de)
Technische Universität Dortmund, Essen, Germany

In this poster, we present an overview of machine learning (ML)-based tools developed by our group (Computational Bioengineering Chair at TU Dortmund University) for protein engineering, along with current projects in the field of biocatalysis. PPI-Detect is an ML tool designed to study protein-protein interactions (PPIs), focusing on the identification and design of protein sequences capable of interacting with specific protein targets.[1] It predicts the likelihood of a protein sequence binding to a target based on the characteristics of the PPI interface. PPI-Affinity is another ML tool that accurately predicts the binding affinity of protein-protein and protein-peptide complexes. In addition, its protein engineering module enables the design of new proteins or peptides—or the optimization of existing ones—to achieve high-affinity binding with desired protein partners.[2] In the field of biocatalysis, our group is actively investigating protein dynamics and solvent effects. We present two recent studies exploring the impact of different cosolvents on the dynamics and enzymatic activity of chymotrypsin and formate dehydrogenase from Candida boidinii, two highly relevant catalytic systems. By combining physics-based and ML approaches, we analyzed active site dynamics, interaction networks between cosolvent molecules and the enzymes, and water transport across the catalytic site, all of which were found to play critical roles in ligand binding and catalytic activity. References: [1] J. Proteome Res. 2022, 21, 1829-1841. [2] J. Comput. Chem. 2019, 40, 1233-1242

Stand № 4

Confinement-modulated Diffusion of Alkenes During Ethylene Oligomerization in Framework Materials

Aleksandr Avdoshin (aleksandr.avdoshin@kit.edu)
KIT, Karlsruhe, Germany

We present a combined molecular dynamics and grand canonical Monte Carlo simulation study of the hierarchical metal–organic framework (MOF) NU-1000 for ethylene oligomerization catalysis. By examining the diffusion behavior of linear α-olefins under confined versus non-confined conditions, we demonstrate that NU-1000's mesopores and micropores distinctly modulate molecular mobility. Under confinement, diffusion changes significantly, underscoring the crucial role of pore topology and thermal dynamics in controlling mass transport. Our results highlight how the framework's tunable structure influences key catalytic parameters such as site accessibility, selectivity, and kinetic profiles. The adsorption parameters calculated can be coupled with mass transfer for the scale-bridging simulations of the ethylene oligomerization reaction. This opens new avenues for tailoring MOF architectures for the optimization of catalytic performance, particularly in processes where diffusion limitations and selective adsorption dictate product distribution.

Stand № 5

Deciphering Promiscuity of HUWE1 for Protein Degradation Using Molecular Dynamic Simulations

Ritika Aggarwal (aggarwar@uni-mainz.de)
Johannes Gutenberg University Mainz, Frankfurt-am-main, Germany

HUWE-1 is a highly promiscuous enzyme which target a variety of protein substrates of high importance in regulatory mechanisms of cell for ubiquitination. However, the strength by which the substrate binds to HUWE-1 is unknown. Intrinsically disordered regions (IDRs) in proteins are sequences within the protein that lack a stable three-dimensional structure under physiological conditions. IDR-1 is mesh of negatively charged residues in the center of the HUWE1, and they recognize the substrate protein sequences based on overall surface basicity. Atomistic molecular dynamics recapitulates experimental trends in how HUWE-1 IDR-1 binds substrates. We did molecular dynamic simulations to explore the mechanisms and the specificity of interactions between HUWE-1 IDR1 and the substrate to understand these interactions better and determined the molecular basis of this substrate selection mechanism using simulations. We mixed in silico one IDR1 sequence with seven copies of the minimal substrate peptide and performed MD simulation. We observed rapid clustering of all the peptide copies with the IDR, and then stable contact throughout the rest of the simulation. For the control peptide, we instead observed very transient interactions, with peptides rapidly associating and dissociating and no long-term clustering. While the control simulation showed no specifically enriched contact points, for the wild-type peptide we observed many favored interaction modes. These interactions were diverse in nature, and while electrostatic interactions were most prominent, there were also some polar and hydrophobic interactions. This could potentially explain how HUWE1 could have some preference for disordered hydrophobic proteins.

Stand № 6

Elucidating the Exciton Transfer Mechanism in LHCII through Machine Learning

Ebru Akkus (ebru.akkus@kit.edu)
Karlsruhe Institute of Technology, Karlsruhe, Germany

The major goal of this project is to elucidate the excitation energy transfer (EET) mechanisms in light-harvesting complexes (LHCs), specifically light-harvesting complex II (LHC II). EET is driven by factors such as site energies, excitonic couplings, and structural rearrangements, requiring accurate modeling to capture exciton dynamics. To achieve this, we employ non-adiabatic molecular dynamics (NAMD) to simulate the combined electronic-nuclear dynamics using the long-range corrected time-dependent density functional tight binding (TD-LC-DFTB) method within quantum mechanical/molecular mechanical (QM/MM) framework, which provides an efficient modeling quantum and nuclear interactions during EET processes. However, these simulations remain computationally demanding for large systems. To address this challenge, we will utilize neural networks (NNs) that trained on TD-LC-DFTB data to accurately predict site energies, excitonic couplings, and forces at a significantly reduced computational cost. By integrating NNs, we aim to accelerate the investigation of EET processes and deepen our understanding of the structural and functional interactions that drive energy transfer in light-harvesting complexes.

Stand № 7

Enabling OF-DFT with Machine Learning

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

Orbital-free density functional theory (OF-DFT) holds the promise of computing ground state molecular properties at minimal cost. Despite this, decades of research have not yielded a kinetic energy functional with sufficient accuracy for general use in computational chemistry. We present a machine learning functional for the kinetic and exchange-correlation energy for OF-DFT calculations that a) achieves similar or better accuracy than previous state-of-the-art models, b) extrapolates from small molecules with up to 15 second-row atoms to significantly larger ones, c) is more robust to different initializations and d) converges in a mathematical sense to the ground state unlike the current state-of-the-art. We achieved this by leveraging and building upon our previously pioneered approach to generate more diverse training data for the non-interacting kinetic energy as well as improvements to the model architecture. Furthermore, we report on further developments to bring Orbital-free DFT a step closer to applicability for molecular systems.

Stand Nº 8

Proton Coupled Electron Transfer in Biomimetic Peptides and Ribonucleotide Reductase

Katharina Spies (katharina.spies@kit.edu)
Karlsruhe Institute of Technology, Karlsruhe, Germany

Proton-coupled electron transfer (PCET) is essential in biological redox processes, facilitating long-range charge transport in photosynthesis, respiration, and DNA synthesis. Aromatic residues like tyrosine and tryptophan act as redox intermediates, enabling electron and proton transfer through enzymatic pathways. Understanding how the environment influences PCET mechanisms is crucial for elucidating enzymatic reactions. Multiscale simulations provide critical insights: molecular mechanics force fields capture protein environmental effects, while quantum chemistry accurately describes electron and proton transfer. To characterize PCET we define two collective variables (CVs): 1) proton transfer, measured by the difference of the distance between the donor and acceptor atoms to the transferred proton and 2) electron transfer, quantified by Mulliken charge differences. Well-tempered metadynamics simulations, with biasing potentials applied to the proton transfer reaction coordinate, allow us to reconstruct unbiased free energy surfaces and determine reaction barrier heights. Our methodology provides a reliable description of PCET in biomimetic systems, where we found that the mechanism depends not only on the orientation of the donor and acceptor residues but also on environmental factors. In particular solvent exposure and nearby protein components influence the thermodynamics and kinetics of PCET. In addition, we apply the methodology to ribonucleotide reductase (RNR), a well-studied protein for PCET reactions, as it exhibits long-range PCET between its subunits.

Stand № 9

Estimation of Hydrogen Atom Transfer Reaction Barriers in Peptides by Learning Full Radical Potential Energy Surface

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

Hydrogen atom transfer is an important step in a wide range of chemical and biological processes, such as protein mechanics, but its mechanistic pathway is not yet fully understood. Simulating these reaction dynamics poses a challenge for fully classical approaches, which is why we use machine-learned potentials to learn a reactive force field in a data-driven way. Since this requires global information on the potential energy surface, we combine efficient data sampling strategies with state-of-the-art graph neural networks. The resulting full radical potential energy surface enables accurate estimations of hydrogen atom transfer reaction barriers in peptides.

Stand Nº 10

Evaluating Enhanced Sampling Methods for Conformational Transitions in Proteins

Ali Sharifian (ali.sharifian@kit.edu)
KIT, Karlsruhe, Germany

Understanding the conformational dynamics of proteins is crucial for elucidating their structure-function relationships. Enhanced sampling methods offer valuable tools to explore the complex energy landscapes governing these biomolecules' conformational transitions. We aim to assess the efficacy of various enhanced sampling methods in predicting conformational transitions of proteins based on their specific characteristics. Through this comparative analysis, we seek to provide insights into selecting the optimal enhanced sampling method tailored to the specific characteristics of proteins under investigation. Overall, our research contributes to advancing the field of molecular dynamics simulations by enhancing our understanding of conformational transitions in biomolecules.

Stand Nº 11

Explainable AI Enables Molecular Design

Jonas Teufel (jonas.teufel@kit.edu)
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

High-performing predictive models seem to have a good "understanding" of chemical properties. By learning how these models come to their predictions, we can learn about the underlying structure-property relationships of the properties themselves. These insights into the basic working principles can then inform the design of new molecules. We explore two explainable AI (xAI) approaches to extract structure-property relationships for molecular property predictions: Attributional explanations from a self-explaining graph neural network highlight substructures within the greater molecular graph that either positively or negatively impact the target property. Additionally, counterfactual explanations illustrate which small perturbations in the input graph structure are needed to cause a significant deviation in the model's prediction. We find that the generated explanations reproduce commonly known rules of thumb for well-explored tasks like water solubility regression and mutagenicity classification. These positive initial findings motivate future applications to more complex and lesser-explored properties.

Stand Nº 12

Exploration of Chemical Space of Target Reduction Potential

Rostislav Fedorov (rostislav.fedorov@h-its.org)
HITS gGmbH, Uni Heidelberg, Heidelberg, Germany

The prediction of Reduction Potential (RP) is a critical challenge in computational chemistry, offering insights into molecular electron transfer properties. In this study, we introduce a Graph Neural Network (GNN) model to predict RP across diverse molecular structures and solvent environments. We begin by benchmarking multiple Density Functional Theory (DFT) levels against experimental data to establish a reliable computational baseline. Recognizing the limitations of existing datasets, we construct an extensive synthetic database, ReSolved, comprising 19,307 molecules with redox potentials in five distinct solvents. Our GNN model outperforms state-of-the-art approaches across multiple solvents. Furthermore, we explore the model's capacity to generalize RP predictions to previously unseen solvents by integrating solvent descriptors into the learning framework, demonstrating robust predictive performance. We also validate the model's ability to handle both closed-shell molecules and open-shell radicals, expanding its applicability beyond conventional methods. Lastly, leveraging an evolutionary algorithm, we employ our trained GNN as an proxy-function to discover novel molecules with targeted RP.

Stand № 13

Expressive Equivariant Message Passing

Peter Lippmann (peter.lippmann@iwr.uni-heidelberg.de)
IWR, Heidelberg University, Heidelberg, Germany

In numerous applications of geometric deep learning, the studied systems exhibit spatial symmetries (rotations and reflections) and it is desirable to enforce these. While many approaches for equivariant message passing require specialized architectures, including non-standard normalization layers or nonlinearities, we here present a framework based on local reference frames ("local canonicalization") which can be integrated with any architecture without restrictions. We enhance equivariant message passing based on local canonicalization by introducing tensorial messages to communicate geometric information consistently between different local coordinate frames.

Stand № 14

Generating Highly Designable Proteins with Geometric Algebra Flow Matching

Simon Wagner (simon.wagner@iwr.uni-heidelberg.de)
IWR Heidelberg, Heidelberg, Germany

We introduce a generative model for protein backbone design utilizing geometric products and higher order message passing. In particular, we propose Clifford Frame Attention (CFA), an extension of the invariant point attention (IPA) architecture from AlphaFold2, in which the backbone residue frames and geometric features are represented in the projective geometric algebra. This enables to construct geometrically expressive messages between residues, including higher order terms, using the bilinear operations of the algebra. We evaluate our architecture by incorporating it into the framework of FrameFlow, a state-of-the-art flow matching model for protein backbone generation. The proposed model achieves high designability, diversity and novelty, while also sampling protein backbones that follow the statistical distribution of secondary structure elements found in naturally occurring proteins, a property so far only insufficiently achieved by many state-of-the-art generative models.

Stand № 15

Generative Diffusion Models for Backmapping in Chemical Compound Space

Dr. Luis Itza Vazquez-Salazar (vazquez@thphys.uni-heidelberg.de)
Institute for Theoretical Physics, Heidelberg, Germany

Chemical compound space (CS) is the set of all possible molecules or materials. Consequently, CS is extraordinarily large, and its exploration is a challenging but necessary task. In this regard, using coarse-grained (CG) models is a powerful tool to lower the complexity of the problem by reducing atomistic degrees of freedom by aggregating atoms in interaction centres called beads. Consequently, the same bead could represent multiple compounds in transferable CG models. Another advantage of CG models is the possibility to obtain outstanding results at low computational cost. By leveraging these advantages, CG models have emerged as a compelling option for exploring CS. Nevertheless, the inverse operation, back mapping, which transforms a CG representation into an all-atom (AA) resolution, is still very cumbersome. Recently, the use of machine learning (ML) techniques has been proposed as a way to build bridges between CG and AA representations. In particular, the use of generative ML models is a compelling strategy. In generative ML, the model learns a data distribution and generates new samples similar to the ones in the training distribution. Here, we explore the use of generative ML for backmapping by employing a classifier-free graph diffusion model in which molecular graphs are first transformed to a uniform noise. Later, the 'denoising' process is learned by using a graph transformer. The generation of molecules corresponding to specific beads is performed by explicitly conditioning it on the bead type fed to the model. Our model allows us to generate molecules with different numbers of heavy atoms (C, N, O, F) while conditioning on an individual bead type. Our results indicate a great validity, diversity and uniqueness (¿ 80

Stand № 16

Generative Modeling to Connect Molecular Simulations of Different Resolutions

Sander Hummerich (sander.hummerich@gmx.de)
Institute for Theoretical Physics Heidelberg, Heidelberg, Germany

Multiscale molecular simulations are essential for the study of complex systems, yet bridging different resolutions-from atomistic to coarse-grained representations-remains a fundamental challenge. In this work, we explore generative modeling approaches to learn mappings between resolutions, focusing on normalized flows as a bidirectional framework that explicitly models the marginalized degrees of freedom. We aim to develop methods that enable the computation of thermodynamic properties, such as free energies, across scales by exploiting the computationally inexpensive coarse-grained resolution in combination with our cross-resolution flow. Preliminary results show that our model can correctly capture the atomistic degrees of freedom and allows us to compute atomistic observables conditioned on a given coarse-grained representation. Furthermore, we are working on using our model as an interpolant between the all-atom and coarse-grained phase spaces, allowing us to perform cross-scale free energy calculations via thermodynamic integration. This work contributes to the broader goal of integrating machine learning into multiscale modeling, potentially improving the efficiency and accuracy of molecular simulations.

Stand Nº 17

Grain Boundaries and Charge Mobility in Organic Semiconductors: A Non-Adiabatic Molecular Dynamics Simulation Approach

Sonali Garg (sonali.garg@kit.edu)
Karlsruhe Institute of Technology, Karlsruhe, Germany

Organic Semiconductors (OSCs) have emerged as crucial materials in the development of electronic and optoelectronic devices due to their exceptional mechanical flexibility, lightweight nature, and cost-effectiveness. Charge carrier mobility is one key quantity to measure the performance of OSC devices. However, the presence of grain boundaries (GBs) can significantly impede charge carrier mobility. GBs can act as traps or barriers, hindering the charge transport and reducing device efficiency [1]. This study investigates the influence of the GB characteristics, including misorientation angles and GB width, on charge carrier mobility and compares the results with intrinsic mobility. Non- Adiabatic Molecular Dynamics (NAMD) simulations, employing Fewest Switches Surface Hopping (FSSH) approach [2,3], were used to model charge transport dynamics. The charge transfer Hamiltonian was constructed using a fragment orbital approach, with its elements computed via the Density Functional Tight Binding (DFTB) method [4,5]. These insights provide a deeper understanding of the effects of GB on charge carrier mobility in OSCs. References: [1] T. Meier et al., Adv. Optical Mater. 2021, 9, 2100115. [2] J. Spencer et al., J. Chem. Phys. 2016, 145, 064102. [3] S. Roosta et al., J. Chem. Theory Comput. 2022, 18 (3), 1264–1274. [4] M. Elstner et al., Phys. Rev. B 1998, 58, 7260. [5] T. Kubař et al., J. Phys. Chem. B 2010, 114 (34), 11221–11240.

Stand № 18

GRAPPA-based CHARMM Extender (GRACE): a Tool for Predicting CHARMM-compatible Force Field Parameters with Automated Refinement

Jonathan Hungerland (jonathan.hungerland@uol.de)
Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Molecules can be represented as graphs. By using a Local Graph Attentional Transformer architecture, the network learns to assign high-dimensional feature vectors called embeddings. The embeddings are then used to predict the complete set of force-field parameters. This allows to study novel molecules at the spatial and temporal scales of force fields.

Stand Nº 19

How a Stretching Force Differently Destabilizes Chemical Bonds on a Protein Backbone

Daniel Sucerquia (daniel.sucerquia@h-its.org)
HITS, Heidelberg, Germany

When subjecting a protein to pulling forces, bonds in the stretched backbone ultimately break. Nowadays, there are no standard methods to find which are the probabilities of having such breakages. In a first approximation, a protein backbone can be considered as a series of springs, each of which carries the same force. However, proteins are more complex than that and force will distribute across the various degrees of freedoms in the peptide, depending on the chemical environment. We here propose a method to study the changes of energy stored in the degrees of freedom of molecules at quantum level of accuracy. With this method, we first obtain different distributions of energies for the bonds in the backbone in a set of peptides with 3 different amino acids. We show how the energy is distributed in the bonds types for different peptides when pulling until the first bond breaks. We can use this data set to train a Machine Learning engine to learn such differences to find the probabilities of bond rupture in any pulled protein.

Stand № 20

Investigation of Novel Epigenetic Signals through Combinatorial Crosstalk Between Histone Isoforms

Hatice Döşeme (hatice.doseme@ibg.edu.tr)
Dokuz Eylül University/Izmir International Biomedicine and Health Institute, Izmir, Türkiye

Abstract Our genomic material is packed into a fiber known as chromatin. The smallest repeating architectural unit of chromatin is a disc-shaped nucleoprotein known as the nucleosome. Nucleosomes are dynamic entities that can modify their conformation, stability, and binding properties through multiple epigenetic mechanisms, such as the incorporation of different histone variants and post-translational modifications. These functions are influenced by epigenetic mechanisms, i.e., they go beyond the DNA sequence itself. While crosstalk between histone post-translations have been extensively studied as key epigenetic regulators, the crosstalk between histone variants remains largely unexplored. These variants, some significantly different from canonical histone proteins, could contribute to cellular signaling through complex interactions. We employ atomistic molecular simulations to examine how different combinations of these variants assemble and behave within chromatin. This in silico approach offers a cost-effective and rapid alternative to traditional lab experiments, particularly crucial in chromatin and epigenetics research because nucleosome reconstitution can be extremely cumbersome. The findings will not only advance our understanding of histone variant interactions but also serve as a foundation for international collaborations investigating their impact on health and disease.
Keywords: nucleosome, molecular dynamics, chromatin

Stand № 21

MACE for ML/MM Simulations

Lukas Petersen (lukas.petersen@kit.edu)
Karlsruhe Institute of Technology, Karlsruhe, Germany

The equivariant graph neural network architecture MACE has been successful in a range of applications. It needs fewer message-passing steps and provides better parallelization in comparison to alternative equivariant architectures such as NequiP, Allegro, and PaiNN. Through additional prediction of partial charges inspired by the third- and fourth-generation High-Dimensional Neural Network Potentials (HDNNPs) a pathway has been opened to a QM/MM variant of these machine learning potentials. Pure machine learning potentials would include solvent/environment molecules in the training data in an indiscriminate manner, but when it comes to complex environments like proteins, this approach becomes unfeasible. The QM/MM approach would facilitate this change of environment immensely. That's why we employ the QM/MM approach to incorporate information about the environment in the model inputs. The electrostatic potential caused by the MM-zone, typically solvent or protein backbone, is included to enable the network to take the environment into account, similar to the electrostatic embedding.

Stand Nº 22

Machine Learning Based Exploration of Potential Energy Surfaces

Kevin Steiner (uqvav@student.kit.edu)
Karlsruhe Institute of Technology (KIT), Gaggenau, Germany

Machine Learning Based Exploration of Potential Energy Surfaces

Stand Nº 23

Machine Learning Potentials for Accurate and Efficient Multiscale Enzymatic Diels – Alder Reaction Modelling.

Magdalena Kazmierczak (al23945@bristol.ac.uk)
University of Bristol, Bristol, United Kingdom

To perform efficient computational screening of reactions catalysed by enzymes, an efficient multiscale potential needs to be selected. Usually, this involves a trade-off between accuracy and efficiency: either substantial sampling with semi-empirical methods or limited sampling with more accurate DFT are applied in QM/MM. Here, we propose the use of machine learning potentials applied together with electrostatic machine learning embedding (EMLE) as an alternative to the traditional QM/MM calculations. AbyU is a powerful Diels-Alderase, which has been proven to be highly stable and mutable, and can act on multiple substrates. For the natural reaction of AbyU benchmarking using ML(EMLE)/MM indicated nanoseconds of simulations per day could be performed with the accuracy in energy and structures at the DFT-level. The EMLE approach is required to accurately capture the influence of the enzyme electric field on the reaction. Present research is focused on applying ML(EMLE)/MM to non-natural Diels-Alder reactions catalysed by AbyU with substrates previously synthesised and tested by our experimental partners. The ultimate goal is to test this workflow with wide range of enzyme variants and demonstrate the potential of ML(EMLE)/MM approach for efficient screening of enzyme variants.

Stand № 24

Machine Learning Potentials for Simulating Sigmatropic Reaction Dynamics

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

Machine learning (ML) potentials are becoming increasingly important for modeling complex chemical reactions. In this study, we benchmark state-of-the-art ML potentials for modeling sigmatropic rearrangements reactions that are simple yet structurally diverse, making them ideal for systematic exploration. Our comparative evaluation highlights key trade-offs among speed, accuracy, and transferability across different ML models. This research sets the foundation for efficient molecular dynamics simulations and detailed mechanistic studies of complex reactions enabled by ML-driven force fields.

Stand Nº 25

Molecular Simulations to Investigate the Membrane Interactions of a Peptide with Activity Against Heart Failure

Tommaso Bartoloni (tommaso.bartoloni@h-its.org)
HITS, Heidelberg, Germany

S100A1ct is a peptide derived from the C-terminal region of the S100A1 Ca2+-binding protein. Experiments have shown the ability of S100A1ct to interact with and promote the activity of SERCA2a, thereby providing a promising approach to treating heart failure. Recent molecular docking studies indicate that S100A1ct binds to SERCA2a as a transmembrane helix. The mechanism by which S100A1ct enters and interacts with the cellular membrane remains, however, elusive. Therefore, we investigated these mechanisms. First, we investigated the peptide's process of insertion in the membrane. We found that this process relies on the interaction of the peptide's aromatic residues with the hydrophobic patches exposed on the bilayer surface. Subsequently, we performed all-atom "assembly" MD simulations followed by Gaussian-accelerated MD to explore the conformational landscape of S100A1ct embedded in the membrane. Our results suggest that S100A1ct can adopt both transmembrane and single leaflet helical conformations, with the latter being more energetically favored. Finally, we performed Monte Carlo simulations using an implicit membrane model to explore how mutations could affect such conformations. The mechanistic insights from these simulations elucidate how S100A1ct behavior differs in a simple membrane model compared to in an aqueous environment and support the predicted binding mode to SERCA2a.

Stand № 26

On the Determinants of Electron Transfer Reorganization Energy in a Cytochrome P450:cytochrome B5 Complex: A Combined Quantum Mechanics and Molecular Dynamics Simulation Study

Jonathan Teuffel (jonathan.teuffel@h-its.org)
Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

Jonathan Teuffel1,2,3, Goutam Mukherjee1,4, Sungho Bosco Han1,5, Marcus Elstner6, Rebecca C. Wade1,2,4,5,7 The electron transfer (ET) steps in the catalytic cycle of cytochrome P450 (CYP) enzymes, ubiquitous proteins with key roles in processes such as drug metabolism and steroidogenesis, are often rate-limiting. To predict ET rates from atomistic molecular dynamics simulations using Marcus theory, values of the reaction free energy and the reorganization free energy are required from either experiment or computation. For the reduction of cytochrome P450 17A1 (CYP17A1) by the secondary redox protein cytochrome b5 (CYb5), a critical step in the regulation of steroidogenesis, experimental measurements of the reorganization ener are not available. We here describe the computation of *lambda* for this system from a combination of molecular mechanics/molecular dynamics simulations and quantum mechanics computations. Our results show that a quantum mechanical treatment of the redox-active cofactors is necessary, even though the surrounding protein and solvent, which are modeled classically, contribute most to the reorganization energy. We find that the reorganization energies computed for the individual soluble globular domains of the two proteins sum to approximately the values computed for the membrane-bound CYP17A1-CYb5 complex, indicating that additivity can be invoked in a computationally efficient approach to estimating reorganization energies for such protein-protein complexes.

Stand Nº 27

Parallel Active Learning and Application in Chemistry

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

PAL is an active learning library built for automation, modularity, and parallel execution. It brings together active learning tasks and efficiently manages their execution and communication on both shared- and distributed-memory systems using the Message Passing Interface (MPI). In this study, we demonstrate the application of PAL to Bismuth clusters, for predicting energies and forces across various shapes and positions, and further explore its capability to simulate cluster synthesis.

Stand Nº 28

Enhancing the Mechanism of Action Prediction Using Alchemical Relative Binding Free Energy Calculations

Jonas Paulus (jpaulu01@uni-mainz.de)
institute of Molecular Physiology Mainz, Mainz, DE

The overall aim is to resolve protein – ligand interaction in order to improve drug development. The tool used, Molecular Dynamics (MD) simulations, reveal a detailed and high-resolution picture of how the ligand interacts with and influences the protein. Using Protein Kinase A (PKA) as an example, I will show the influence of protonated benzoic acid, as opposed to benzamide, on the structure of PKA, even though the crystal structure indicates stereoisomerism. Additionally, I will give an outlook on a simulation that uses alchemical multi-state transitions in a coarse grained simulation. The aim is to to sample a large part of the chemical space of small molecules for binding candidates.

Stand Nº 29

Relative Binding Free Energy Calculations for Ligands in Complex with Protein Kinase A Derived From Atomistic Molecular Dynamics Simulations

Rebecca Ziora (rziora@students.uni-mainz.de)
Johannes Gutenberg-Universität Mainz, Mainz, Deutschland

In order to characterise protein-ligand binding behaviour as well as quantifying the Relative Binding Free Energy (RBFE) as a measure of the binding affinity, Protein Kinase A (PKA) is studied in complex with three distinct ligands. The calculations are based on atomistic molecular dynamics simulations. PKA is a well-studied protein of medical importance as it is a cell signalling protein. It is studied in complex with the ligands benzoic acid, deprotonated benzoic acid and benzamide. By studying the protein-ligand interactions, which are essential for biological functions, potential drug candidates can be identified. Analysing the stability and binding behaviour of the three complexes reveals that the position of the deprotonated benzoic acid relative to the protein varies much more than compared to the other ligands. Furthermore, alchemical non-equilibrium simulations are conducted to estimate the RBFE between the ligands benzoic acid and deprotonated benzoic acid as well as between benzoic acid and benzamide. For this, an alchemical thermodynamic cycle is constructed. Since there is a change in charge when morphing between benzoic acid and its deprotonated form, the single box simulation approach is used. This overcomes the problem of having different charges in the two states of the system.

Stand № 30

Revisiting Acene Dimers: A Comprehensive Theoretical Study of A More Stable Conformer

Nikita Matsokin (nikita.matsokin@kit.edu)
Karlsruhe Institute of Technology, Stutensee, Germany

Acenes are linear polycyclic aromatic hydrocarbons known for their unique electronic and optical properties, making them promising materials in organic electronics and photonics. Their aromatic properties facilitate the formation of various non-covalent aggregates-such as dimers, trimers, and tetramers-primarily stabilized by π-π interactions, which significantly influence the physical properties and performance of the resulting materials. In this work, we present a detailed theoretical investigation of an alternative acene dimer configuration that has received limited attention despite its potentially higher stability compared to the conventional glike conformers. Geometric optimizations were performed using Density Functional Theory and subsequently validated using the post-Hartree Fock method such as CCSD(T). Furthermore, Symmetry-Adapted Perturbation Theory was applied to quantitatively assess the contributions of various non-covalent contributions to the overall stabilization of the dimer. Additionally, UV spectral analysis was performed to obtain the influence of different interactions. The results not only deepen our understanding of non-covalent interactions in acene systems but also pave the way for designing materials with optimized properties for advanced technological applications.

Stand № 31

Understanding the Catalytic Mechanism of Creatine Kinase through QM/MM and Metadynamics Simulations

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

Mitochondrial creatine kinases (uMtCK) are crucial for cellular energy metabolism, reversibly transferring phosphate between creatine and phosphocreatine. Phosphocreatine, a high-energy phosphate reservoir, supports rapid ATP regeneration in tissues like muscle and brain. By converting ATP to phosphocreatine, uMtCK facilitates ATP regeneration in energy demanding tissues, ensuring a steady supply of ATP. While structural studies offer a general understanding of the reaction mechanism, details regarding the energetic landscape and the influence of surrounding amino acids remain unclear. Previous computational studies, limited by simplified models and short timescales, have hindered a complete understanding of protein dynamics and environmental effects on catalysis. To overcome these limitations, we employ quantum mechanics/molecular mechanics (QM/MM) simulations with DFTB3 as QM method, combined with metadynamics as an enhanced sampling technique, to gain deeper insight into the catalytic mechanism of uMtCK. Our results highlight the two-step deprotonation and phosphorylation mechanism, capturing the protein environment's influence on the reaction pathway, enabling the exploration of otherwise inaccessible rare catalytic events, and providing a detailed free energy landscape. Furthermore, we investigated the contributions of key amino acids and protein dynamics. This integrated QM/MM and enhanced sampling approach refines our understanding of creatine kinase catalysis and the protein environment's role.

Stand № 32

Using Denoising Diffusion Models to Learn Free Energies From Non-Equilibrium Steady States

Daniel Nagel (nagel@thphys.uni-heidelberg.de)
ITP, Heidelberg, Heidelberg, Germany

Accurate estimation of free energy profiles is essential for predicting the behavior of complex molecular systems. While biased molecular dynamics simulations enhance the sampling of rare events, the extraction of reliable free energy landscapes from these simulations remains challenging. On the other hand, the exploitation of the Fokker-Planck equation provides valuable insights into the dynamics of complex systems in nonequilibrium states. However, its application is constrained by its computational complexity. This work presents a novel approach that combines the principles of statistical physics with the established machine learning technique of denoising diffusion models to efficiently estimate free energy profiles from biased non-equilibrium steady states. By establishing a connection between the diffusion and simulation times, it is demonstrated that the training objective, the so-called score, can be decomposed into a non-trivial conservative contribution from the equilibrium potential and a trivial non-conservative part determined by external driving forces. To demonstrate the robustness of our approach and its ability to learn equilibrium free energy profiles, we apply it to a driven toy model and a Martini force field molecular dynamics simulation of a small molecule biased through a lipid bilayer.

Stand Nº 33

Author Index