

CHIMIA

CHIMIA 2025, Volume 79

ISSN 0009-4293

www.chimia.ch

Supplementa to Issue 7-8/2025



SCS

Swiss Chemical
Society

SCS Fall Meeting 2025

Lecture, Short Talk and Poster Abstracts

Computational Chemistry

September 4, 2025

University of Zurich, Irchel Campus

<https://fm25.scg.ch>

Swiss Chemical Society
Marktgasse 32
3011 Bern
Switzerland
info@scg.ch
www.scg.ch

Free energies from score-based generative modelsT. Bureau¹¹Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany - bureau@uni-heidelberg.de

Thermodynamic integration (TI) offers a rigorous method for estimating free-energy differences by integrating over a sequence of interpolating conformational ensembles. However, TI calculations are computationally expensive and typically limited to coupling a small number of degrees of freedom due to the need to sample numerous intermediate ensembles with sufficient conformational-space overlap. We propose to perform TI along an alchemical pathway represented by a trainable neural network, which we term Neural TI. Critically, we parametrize a time-dependent Hamiltonian interpolating between the interacting and non-interacting systems, and optimize its gradient using a denoising-diffusion objective. The ability of the resulting energy-based diffusion model to sample all intermediate ensembles allows us to perform TI from a single reference calculation. We apply our method to Lennard-Jones fluids, where we report accurate calculations of the excess chemical potential, demonstrating that Neural TI is capable of coupling hundreds of degrees of freedom at once [1]. Extensions to solvation free energies for atomistic systems will be discussed [2].

[1] B. Máté, F. Fleuret, T. Bureau, *The Journal of Physical Chemistry Letters*, **2024**, 15(45), 11395-11404.

[2] B. Máté, F. Fleuret, T. Bureau, *The Journal of Chemical Physics*, **2025**, 162(12).

Bridging the Gap Between Variational and Perturbational DFT-based Methods for Calculating Excited States

A. Sinyavskiy¹, M. Mališ^{1*}, S. Luber^{1*}

¹Department of Chemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Obtaining the excited electronic states and their properties is a vital part of simulating photochemical reactions. The most precise wavefunction-based methods scale poorly with the system size, which sparked the development of density functional theory (DFT)-based perturbational methods like Linear-Response Time-Dependent DFT also known as TDDFT. It is affordable, but comes with certain drawbacks, as it fails to describe the charge transfer and Rydberg states. On top of that, the multi-electron excitations cannot be simulated by TDDFT due to the adiabatic approximation. Variational DFT-based methods are a promising alternative which should alleviate the aforementioned issues.¹ However, they require a good initial guess as well as some sort of verification procedure to ensure that the resulting excited state is the one we need.

A restricted-open Kohn--Sham delta self-consistent field (ROKS- Δ SCF) approach allows for direct calculation of unmixed singlet and triplet excited electronic states without the need for spin purification.² With a proper initial guess, i.e. the fractional occupation numbers which can be obtained e.g. from a TDDFT calculation, and a molecular orbital tracking procedure to rectify the root flipping during the diagonalization, even distinctly multi-reference states with no clearly dominant contribution can be targeted and converged which, to our best knowledge, has not been achieved before.³ The previously developed method for calculating the transition dipole moments⁴ has been extended to the multi-reference states as well. Moreover, we conducted a large-scale benchmark and demonstrated that the excited electronic states converged with our ROKS- Δ SCF method are indeed the targeted ones in every case by comparing the electron density differences and transition dipole moments with both TDDFT and the second-order algebraic diagrammatic construction method.

[1] Nicola Bogo, Christopher J. Stein, *Physical Chemistry Chemical Physics*, **2024**, 26, 21575-21588

[2] Eva Vandaele, Momir Mališ, Sandra Luber, *Journal of Chemical Physics* **2022**, 156, 130901

[3] Andrey Sinyavskiy, Momir Mališ, Sandra Luber, in revision

[4] Momir Mališ, E. Vandaele, Sandra Luber, *Journal of Chemical Theory and Computation* **2022**, 18, 4082

Non-equilibrium Alchemical Transformation with Enveloping Distribution Sampling

S. Chen¹, E. Ruijsenaars^{1*}, S. Riniker^{1*}

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

In recent years, non-equilibrium (NEQ) approaches have attracted increasing attention [1-3]. Rather than spending computational resources on sampling intermediate states, non-equilibrium methods focus on the end states given a fixed NEQ protocol. In theory, a locally optimal NEQ protocol with the minimal entropy production can be achieved by setting the switching speed proportional to the inverse of the square root of the alchemical friction [4]. In practice, however, when the alchemical friction differs in the forward and backward transition due to a locally non-positive potential curvature, only a sub-optimal protocol can be found to minimise the average entropy production.

In this study, we demonstrate that enveloping distribution sampling (EDS) [5] is a potential alternative to the commonly used linear interpolation (LI) potential. The behaviour of the EDS potential is characterised by the smoothing parameter s that modulates the local curvature of the potential energy surface and recovers the LI potential in the limit of $s \rightarrow 0$. Using a range of systems – from simple harmonic traps and anti-traps to particle insertions and dipole inversion – we demonstrate that the parameter s serves as a promising tuning variable for optimizing NEQ protocols in real-world alchemical transformations.

[1] Wan, S.; Bhati, A. P.; Coveney, P. V. Comparison of Equilibrium and Nonequilibrium Approaches for Relative Binding Free Energy Predictions. *J. Chem. Theory Comput.* **2023**, *19*, 7846– 7860,

[2] Bhati, A. P.; Wan, S.; Coveney, P. V. Equilibrium and nonequilibrium ensemble methods for accurate, precise and reproducible absolute binding free energy calculations. *J. Chem. Theory Comput.* **2025**, *21*, 440– 462,

[3] Albergo, M. S. and Vanden-Eijnden, E. Nets: A non-equilibrium transport sampler. arXiv preprint arXiv:2410.02711, **2024**.

[4] D. A. Sivak and G. E. Crooks, Thermodynamic metrics and optimal paths, *Phys. Rev. Lett.* **108**, 190602 **2012**.

[5] Christ, C. D.; van Gunsteren, W. F. Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. *J. Chem. Phys.* **2007**, *126*, 184110,

Compressing Quantum States of Identical Particles with Neural Network StatesL. Ding¹, V. Barandun^{1*}, M. Reiher^{1*}¹ETH Zürich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 1-5/10, 8093, Zürich, Switzerland

The growing entanglement in scaled-up or time-evolved strongly correlated many-body systems poses a fundamental challenge in an accurate, yet affordable description of their quantum states. In the light of neural-network quantum states (NQS) [1], such an entanglement barrier may be finally overcome. Evidence exists that such ansätze can indeed simulate highly entangled (volume-law) states with only polynomial cost [2,3], thereby achieving a quantum state compression. That being said, there is an apparent mismatch between the entangled states, that each NQS ansatz targets, and the physically or chemically relevant ones [4]. Recent developments in fermionic NQS ansätze [5,6] showed that, despite the validity of the universal approximation theorem, a network agnostic of the indistinguishability of the particles may have a harder time achieving the same accuracy than a network that is physically aware. In this work, we set out to understand the reason behind the compact expression of highly entangled quantum states by a physically aware NQS design, specifically the neural backflow ansatz for fermions [5]. We then incorporate these insights into the construction of a bosonic analogue, where these advantages could be even more pronounced. Applications will then consider vibrational ground and time-evolved states calculations.

[1] Carleo G., Troyer M., Science, **2017**, 355, 6325, 602

[2] Deng D.-L., Li X., S. D. Sarma, Phys. Rev. X, **2017**, 7, 2, 021021

[3] Levine Y., Sharir. O, et al., Phys. Rev. Lett., **2019**, 122, 6, 065301

[4] Passetti G., Hofmann D., et al. Phys. Rev. Lett., **2023**, 131, 3, 036502

[5] Liu A.-J., Clark, B. K., Phys. Rev. B., **2024**, 11, 115137

[6] Zakari, D., Allessandro, S., et al., arXiv preprint, **2023**, arXiv:2309.11534

Broadening the Applicability of FDET-Based Methods through Bottom-Up Approximations of the Non-Additive Kinetic Potential

P. Roy¹, T. Wesolowski^{1*}

¹University of Geneva, department of physical chemistry

Frozen Density Embedding Theory (FDET) [1-3] provides a rigorous framework for embedding quantum subsystems within an environment. In this formalism, the total electron density is constructed from two contributions: the embedded subsystem's density, described by a wavefunction, and the environment's density, which remains frozen. This partitioning allows for a flexible choice of computational methods, where the embedded subsystem can be treated using wave function-based approaches that are accurate for the observables of interest, while the environment's density can be obtained from coarser, computationally efficient approximations. The embedding potential is determined by minimizing the total energy—expressed as a functional of both the embedded subsystem's wavefunction and the environment's density—with respect to variations in the subsystem's density.

The accuracy of FDET-based methods hinges on reliable approximations to the non-additive kinetic energy (NAKE) bifunctional, which accounts for Pauli repulsion between the embedded system and its environment. However, for most properties of interest, it is not the NAKE itself but rather its functional derivative—the non-additive kinetic potential (NAKP)—that must be accurately described. A promising strategy for improving NAKP approximations is the bottom-up approach, which models the NAKP directly rather than deriving it from an approximate kinetic energy functional. This method offers key advantages, notably the ability to enforce exact constraints and recover correct physical limits, leading to improved accuracy and transferability. We demonstrate that incorporating these strategies enhances the applicability of FDET-based methods, enabling us to tackle challenging embedding problems in both ground and excited states. Specifically, we show that these approaches improve the description of charge transfer to the environment and covalent bonding between the embedded subsystem and its environment, where standard embedding methods often fail.

(1) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calcu-

lations of solvated molecules. *The Journal of Physical Chemistry* 1993, 97, 8050–8053.

(2) Wesolowski, T. A. Embedding a multideterminantal wave function in an orbital-free

environment. *Physical Review A* 2008, 77, 012504.

(3) Wesolowski, T. A. Embedding potentials for excited states of embedded species. *The*

Journal of Chemical Physics 2014, 140, 18A530.

Towards the new types of memory devices using molecular electronicsV. Slama¹, U. Röthlisberger^{1*}¹ École Polytechnique Fédérale de Lausanne, Rte Cantonale, 1015 Lausanne, Switzerland

Molecular electronics is seen as a route to further miniaturize microelectronics and to extend Moor's law beyond the limits of semiconductor-based platforms. The molecular electronics devices make use of the quantum nature of electrons and their control to facilitate specific functions. Our main focus here is on molecular electronics memories, where the specific resonance interaction between an electrode surface plasmon and the charge transfer state of an active molecule is used to influence the charge distribution by an external electric or magnetic field. In this project, we focus on investigating the nature of the coupling between active molecules and metallic electrodes. Understanding the interplay of the different interactions such as exciton-exciton, charge transfer and exciton-plasmon interactions and their dependence on the coupling between the active molecule and the surface, is a crucial step in developing novel molecular electronic devices. We aim to develop a computationally affordable approach correctly describing intrinsic quantum effects like mixing of the locally excited or charge transfer states with the surface plasmon resonance. This is essential for large-scale investigations of the new type of molecular devices. Promising system that can serve as prototypes for future molecular memories are paracyclophane molecules deposited on a gold surface. These molecules contain two aromatic units stacked parallel and fixed by linkers, offering a well-defined structure with controlled inter-unit conformation. When paracyclophane is formed by two different units e.g. naphthalenediimide-Pyrene or naphthalenediimide-Azulene, we observe large charge transfer (CT). By varying the linker's length and flexibility, we can systematically adjust the inter-unit distance and orientation, thereby modifying CT state properties and obtaining excited state energies close to the gold plasmon resonance with the CT transition dipole moment perpendicular to the gold surface. Resonant coupling between the CT state and the gold plasmon resonance could then be utilized for future molecular memories. I'll present computational strategy for accurate treatment of the CT states as well as the locally excited states of the paracyclophanes and investigate electronic interaction between the paracyclophane and the gold surface. Comparison of computed and experimental optical spectra and STM show excellent agreement and validating the proposed computational methodology.

Process Development Fueled by Enabling TechnologiesS. Wagschal¹¹Advanced Chemistry Technologies, Global R&D Small Molecule Lonza, Visp

Despite the increasing focus on biologics, small molecules continue to dominate pharmaceutical sales, accounting for about 90% of the market. With advancing research, the complexity of novel small molecule active pharmaceutical ingredients (APIs) is increasing. This escalation results in more extended synthetic pathways, decreased yields, and elevated costs. Process chemists face the challenge of developing reliable manufacturing processes while managing supply chains for raw materials and intermediates. Lonza's approach integrates AI-enabled informatics and automation technologies, including computer-aided synthesis planning (CASP) tools, to design efficient retrosynthetic paths. Combined with high-throughput experimentation (HTE), this strategy accelerates clinical readiness, reduce costs, and enhance strategic decision-making for pharmaceutical innovators. A recent application to an early-phase project will be presented to showcase the benefits of these technologies.

Generative molecular design with steerable and granular synthesizability control

V. Sabanza Gil^{1,2}, J. Guo^{1,2*}, Z. Jončev^{1*}, J. Luterbacher^{3*}, P. Schwaller^{1,2*}

¹Laboratory of Artificial Chemical Intelligence (LIAC), EPFL (Switzerland), ²NCCR Catalysis, EPFL (Switzerland), ³Laboratory of Sustainable and Catalytic Processes (LPDC), EPFL (Switzerland)

Molecular generative design is crucial for identifying and developing novel molecules with targeted physicochemical and biological properties [1]. Despite recent progress facilitated by advancements in generative artificial intelligence, translating computationally designed molecules into experimentally viable candidates remains challenging [2]. A critical bottleneck in the validation process is the requirement to synthesize these computationally proposed molecular structures in laboratory settings. Current generative models often overlook synthetic feasibility, resulting in molecules that, while theoretically promising, present substantial practical difficulties in experimental synthesis [3]. In this work, we introduce a generative molecular design framework that allows selective and granular control over multiple elements in the predicted synthetic routes of the generated molecules. By combining a pretrained language model with a policy gradient optimization algorithm coupled to a retrosynthesis model, the model learns to enforce the presence of pre-defined reactions and/or building blocks, or to avoid the inclusion of selected transformations in the generated molecule's final synthetic route. The steerable control over the synthetic parameters allows high flexibility and is more amenable to experimental validation. Overall, this method opens a new capability for accelerating the discovery of promising molecules with tailored properties.[4]

[1] Yuanqui Du, Arian Jamasb, Jeff Guo, et al. *Nature Machine Intelligence*, **2024**, 6, 589–604

[2] Camille Bilodeau, Wang Jin, Tommi Jaakkola, Regina Barzilay, Klavs Jensen, *WIREs Computational Molecular Science*, **2022**, 12:e1608.

[3] Wenhao Gao, Connor Coley, *Journal of Chemical Information and Modeling*, **2020**, 60, 5714–5723

[4] Jeff Guo, Víctor Sabanza Gil, Zlatko Jončev, Jeremy Luterbacher, Philippe Schwaller, 2025, <https://arxiv.org/abs/2505.08774>

Machine learning-guided design of rhenium tricarbonyl complexes for next-generation antibiotics

M. Nedyalkova¹, G. Demirci^{1*}, Y. Cortat^{1*}, M. Lattuada^{1*}, F. Zobi^{1*}

¹University of Fribourg

The escalating prevalence of antibiotic-resistant bacteria and the increasing complexity of managing severe infections emphasize the critical need for novel and effective antibiotics. Herein, we present a novel computational strategy focused on metal-based antibiotics, specifically rhenium (Re) complexes, for the rational design of next-generation antibacterial agents. Our approach integrates machine learning (ML) classification models to predict antibacterial potency, particularly against multidrug-resistant pathogens. A recognized limitation of conventional ML-driven antibiotic discovery is its dependence on structural similarity to known antibiotics, which hinders the exploration of structurally diverse and innovative antibiotic classes. To address this, we developed predictive ML models based on Multi-Layer Perceptron (MLP) and Random Forest (RF) algorithms to estimate the minimum inhibitory concentration (MIC) of Re-complexes against methicillin-resistant (MRSA) and methicillin-sensitive (MSSA) *Staphylococcus aureus* strains. Utilizing structural descriptors, these models demonstrated strong predictive performance and were successfully applied to evaluate 26 novel Re-complexes. Additionally, SHapley Additive exPlanations (SHAP) analysis provided insights into the structural features influencing antibacterial activity predictions. The outcomes of the study affirm the effectiveness of our ML-guided approach as a promising pathway for the rational, de novo design of potent Re-based antibiotics capable of combating antibiotic-resistant bacterial infections.

Transformer Model for Structure Elucidation from Tandem Mass Spectroscopy data

L. Mismetti^{1,2}, M. Alberts^{1,3*}, M. Graziani^{1,2*}, T. Laino^{1,2*}

¹IBM Research, Säumerstrasse 4, 8803 Rüschlikon, Switzerland, ²NCCR Catalysis, Switzerland, ³University of Zürich, Department of Chemistry, 11, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Tandem mass spectrometry (MS/MS) is a crucial technique in multiple fields, including pharmaceutical analysis, drug discovery, diagnostics, metabolomics, proteomics, and environmental science. It enables the identification of molecules present in complex mixtures. However, the annotation and interpretation of MS/MS data remains challenging due to the complexity of the data and vastness of chemical space. To annotate MS/MS data, experimental spectra are typically compared against existing spectral databases in order to find the most similar matches. However, this approach is strongly limited by the size and diversity of the reference database used and does not allow for the elucidation of structures of novel compounds. In contrast, previous work has shown that Artificial Intelligence (AI) models can annotate MS/MS data without the need for database retrieval and direct comparison to existing references. Due to the difficulty of the task and scarcity of high-quality data, current state-of-the-art AI models adopt a sequential approach: first predicting molecular fingerprints and subsequently generating molecules as SMILES (Simplified Molecular-Input Line-Entry System [1]). While this two-step strategy partially mitigates the challenge of MS/MS data annotation, it also introduces complexity with the current state of the art, exhibiting both low performance and failing to generalise beyond the training set.

We propose a novel Transformer model, which predicts SMILES directly from input MS/MS spectra in an end-to-end manner. This approach eliminates the need for intermediate steps, such as identifying molecular fingerprints. Leveraging transfer learning from simulated data, we achieve state-of-the-art performance on the challenging MassSpecGym dataset. This dataset is known for its extreme difficulty, with the best existing Top-1 zero-shot accuracy at 2.3% [2]. Our model surpasses this benchmark by 50%, establishing a new upper bound of test performance at 3.4% Top-1 zero-shot accuracy. These results highlight the potential of transformer-based models in simplifying and improving structure elucidation process from MS/MS data.

[1] D. Weininger, Journal of Chemical Information and Computer Sciences, 1988, 28, 31-36.

[2] M. Bohde, M. Manjrekar, R. Wang, S. Ji, C. W. Coley, arXiv:2502.09571v1, 2025.

Feature Selection in Multivariate Linear Regression Models for Data-Efficient Reaction Optimization

A. A. Schoepfer^{1,2}, R. Laplaza^{1*}, J. Waser^{2*}, C. Corminboeuf^{1*}

¹Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, ²Laboratory of Catalysis and Organic Synthesis, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Reliable machine learning models are difficult to develop in the early stages of reaction optimization due to data scarcity. To address this limitation, chemists have turned to multivariate linear regression (MLR) models that rely on a small set of carefully selected molecular features, such as charges, angles, and volumes. Despite MLR's simplicity and high interpretability, its practical application depends on selecting only a few relevant features from a typically large pool. However, existing feature selection approaches for reaction optimization are often inconsistent, leading to suboptimal models and providing limited insight into model quality.

To overcome these limitations, we introduce ReaFS (Reaction Feature Selection), a comprehensive framework of data, feature, and model diagnostics designed to identify statistically robust MLR models for any given reaction. ReaFS facilitates model selection by computing relevant metrics and enables an *a priori* assessment of model quality. Combined with Bayesian optimization and the interpretability of MLR, this approach enables more efficient and streamlined reaction optimization.

Computational design of single-atom catalysts for sustainable vinyl chloride synthesis

A. R. Ferrando¹, S. M. -^{1*}, N. L. Alonso^{2*}, J. P. Ramírez^{1*}

¹ETH Zurich, ²ICIQ-CERCA

Acetylene hydrochlorination, a key technology for vinyl chloride monomer synthesis, urgently needs to transition away from toxic HgCl₂ catalysts in pursuit of more sustainable alternatives. Single-atom catalysts (SACs) based on transition metals, particularly platinum, have emerged as promising candidates, combining high durability and productivity. [1,2] Enhancing their activity and fully exploiting their potential requires atomistic insight into how active sites form during synthesis, evolve under reaction conditions, and ultimately govern catalytic performance. As the smallest possible metal ensembles, SACs provide a unique opportunity for such understanding due to their comparatively well-defined structures. However, current theoretical approaches often fall short by neglecting synthetic history, oversimplifying support effects, and using stability descriptors disconnected from realistic conditions. [3] To overcome these gaps, we introduce a thermodynamically grounded, coordination-resolved computational framework. By explicitly simulating the synthesis-structure-property-performance relations of carbon-supported SACs under relevant synthesis and reaction conditions (e.g. 473 K and p HCl ~ 0.5 atm), the model captures dynamic changes in site speciation. [2] Applied to relevant transition metals (Pt, Au, Cu, Ru), our approach reproduces over 80% of coordination environments observed via operando X-ray absorption spectroscopy, and reveals how structural evolution under working conditions affects both catalytic activity and deactivation. These results demonstrate the framework's predictive value for guiding catalyst selection and optimization in acetylene hydrochlorination and beyond.

[1] S.K. Kaiser, E. Fako, G. Manzocchi, F. Krumeich, R. Hauert, A.H. Clark, O. Safonova, N. López, J. Pérez-Ramírez, *Nat. Catal.* 2020, 3, 376.

[2] V. Giulimondi, A. Ruiz-Ferrando, G. Giannakakis, I. Surin, M. Agrachev, G. Jeschke, F. Krumeich, N. López, A.H. Clark, J. Pérez-Ramírez, *Nat. Commun.*, 2023, 14, 5557.

[3] F. Kraushofer, G.S. Parkinson, *Chem. Rev.* 2022, 122, 14911.

Fast and scalable retrosynthetic planning with a transformer neural network and speculative beam search.

M. Andronov¹, N. Andronova^{2*}, M. Wand^{1,3*}, J. Schmidhuber^{4*}, D. Clevert^{5*}

¹IDSIA, USI, SUPSI, 6900 Lugano, Switzerland, ²Independent researcher, ³Institute for Digital Technologies for Personalized Healthcare, SUPSI, 6900 Lugano, Switzerland, ⁴AI Initiative, KAUST, 23955 Thuwal, Saudi Arabia,

⁵Machine Learning Research, Pfizer Research and Development, Friedrichstr. 110, 10117 Berlin, Germany

Fast and accurate computer-aided synthesis planning (CASP) is an old technology that becomes increasingly in demand today as a crucial part of computational drug discovery workflow. As with many things today, it is being transformed by artificial intelligence (AI), which replaces the expert systems relying on manually encoded rules. An AI-powered CASP system typically combines a machine learning-based single-step retrosynthesis model with a planning algorithm, such as Monte Carlo Tree Search. The single-step model generates a set of candidate retrosynthetic chemical transformations for a given molecule, and the planning algorithm constructs multiple synthesis trees given those candidates.

The most accurate neural models for single-step retrosynthesis prediction are the models based on the Transformer, the architecture that is also the technology powering Large Language Models such as ChatGPT. Transformers pose the single-step retrosynthesis problem as SMILES-to-SMILES translation. The high accuracy of such models makes them increasingly interesting for integration in multi-step CASP systems, such as AiZynthFinder and Syntheseus. However, the transformer generates SMILES "one letter per run", and is therefore very slow on inference compared to its less accurate alternatives [1], which is a major obstacle in the way of its successful application in multi-step synthesis planning. Therefore, a need arises to accelerate the inference of the SMILES-to-SMILES transformer.

To accelerate SMILES-to-SMILES transformers, researchers have turned to knowledge distillation and speculative decoding [2][3] - a technique that allows the generation of more than one token per one model call by "drafting" multiple tokens ahead and then verifying them. Our recent work [2] introduced the speculative beam search algorithm - the first algorithm that makes speculative decoding suitable for multiple parallel SMILES generation and, therefore, for single-step retrosynthesis.

We build on our recent work and propose using the speculative beam search method in combination with the drafting method called "Medusa," [4] which demonstrates much better scalability to larger input batch sizes than the previously described method [2]. Our method accelerates single-step retrosynthesis with the 18M-parameter transformer model by 4 times with batch size 16. We also demonstrate our method in multi-step synthesis planning and quantify its advantage in inference time to state-of-the-art transformer models without compromising accuracy. Our method will interest chemistry specialists who often face the problem of synthesis planning.

[1] P. Torren-Peraire, A.K. Hassen, S. Genheden, J. Verhoeven, D.A. Clevert, M. Preuss, I.V. Tetko, Digital Discovery, 2024, 3, 558-572.

[2] M. Andronov, N. Andronova, M. Wand, J. Schmidhuber, D.A. Clevert, Journal of Cheminformatics, 2025, 17, 31.

[3] Y. Leviathan, M. Kalman, Y. Matias, International Conference on Machine Learning, 2023, 19274-19286.

[4] T. Cai, Y. Li, Z. Geng, H. Peng, J.D. Lee, D. Chen, T. Dao, arXiv preprint, 2024, arXiv:2401.10774.

MiMiC: A Flexible Multiscale Simulation Framework for the Next-Generation HPC

A. Antalík¹, J. M. Olsen^{2*}, U. Röthlisberger^{1*}

¹Laboratory of Computational Chemistry and Biochemistry, EPFL, Lausanne, Switzerland, ²DTU Chemistry, Technical University of Denmark, Lyngby, Denmark

These days, computational chemists can choose from a wide variety of software packages, each with its own unique features and strengths. Many of these programs also offer some support for modeling chemical processes requiring treatment at multiple resolutions, be it due to the need of a higher level of theory to capture relevant phenomena or because they unfold on considerable time scales. However, writing interfaces for such simulations involving several independent programs is always a balancing act between performance and flexibility.

Addressing this challenge, the MiMiC framework [1,2] provides a general interface for different programs handling individual subsystems at varying levels of theory. It facilitates data exchange between these concurrently running programs and potentially calculates interactions that might arise between the subsystems. To minimize the implementation and computational overheads, we adopted a loose-coupling paradigm combined with a multiple-program multiple-data model (MPMD). The key advantage of this approach is that MiMiC does not interfere with the underlying parallelization of different external programs, thus maintaining their high efficiency that can be fully exploited on modern modular high-performance computing (HPC) architectures.

In recent years, MiMiC has undergone a major refactoring that further enhanced its flexibility, made it more accessible, and introduced several new programs into its ecosystem [3,4]. In this contribution, we highlight some of these latest developments with a particular focus on its modularity and the excellent performance of its generalized electrostatic embedding QM/MM scheme on modern supercomputers. We believe that all these features will make MiMiC an excellent choice for multiscale molecular modeling in the forthcoming era of exa-scale HPC.

[1] A. Antalík et al., *J. Chem. Phys.*, **2024**, 161, 022501

[2] <https://mimic-project.org>

[3] A. Antalík, A. Levy, S. K. Johnson, J. M. H. Olsen, U. Röthlisberger, *J. Chem. Inf. Model.*, **2025**, [DOI: 10.1021/acs.jcim.5c00409]

[4] A. Levy, A. Antalík, J. M. H. Olsen, U. Röthlisberger, **2025**, [arXiv:2502.06539]

Tango*: Constrained synthesis planning using chemically informed value functions

D. Armstrong¹, Z. Jončev^{1*}, J. Guo^{1,2*}, P. Schwaller^{1,2*}

¹École Polytechnique Fédérale de Lausanne, ²NCCR Catalysis

Computer-aided synthesis planning (CASP) has made significant strides in generating retrosynthetic pathways for simple molecules in a non-constrained fashion. Recent work introduces a specialised bidirectional search algorithm with forward and retro expansion to address the starting material-constrained synthesis problem, allowing CASP systems to provide synthesis pathways from specified starting materials. In this work, we introduce a simple guided search—Tango*—which allows solving the starting material-constrained synthesis planning problem using an existing uni-directional search algorithm, Retro*. We show that by optimising a single hyperparameter, Tango* outperforms existing methods in terms of efficiency and solve rate. We also highlight the effectiveness of our computed node cost function in steering synthesis pathways.

On the single-Hessian thawed Gaussian wavepacket dynamics for electronic spectroscopyD. Barbiero¹, J. J. Vaniček^{1*}¹Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland

Single-Hessian Gaussian wavepacket dynamics (GWD) enables simulations of spectra of high-dimensional, slightly anharmonic systems at the cost of ab initio classical molecular dynamics [1,2]. This semiclassical method significantly reduces the computational burden of Heller's local harmonic GWD, while maintaining comparable accuracy in approximating spectra. The unexpected but very useful behavior of this method has been hypothesized to stem from the superior geometric properties of the single-Hessian approach over the local harmonic version [3]. Here, we prove explicitly that, similar to the optimal variational GWD, the single-Hessian GWD conserves the non-canonical symplectic structure on the manifold of Gaussian wavepackets, and we derive expressions for numerical verification. To accelerate numerical simulations, while preserving the geometric properties of the method, we implement high-order geometric integrators that are time-reversible and conserve the norm and symplectic structure exactly, regardless of the time step. Using on-the-fly ab initio Gaussian wavepacket dynamics on the first excited-state surface of ammonia, we numerically assess the conservation of geometric properties by these integrators, demonstrating that high-order integrators can enhance both accuracy and computational efficiency. Employing the geometric integrators, we compute the photoelectron spectrum of the difluorocarbene anion and the absorption spectrum of methylamine. Single-Hessian GWD reliably reproduces the experimental spectra, outperforming global harmonic models and matching the accuracy of local harmonic GWD. Comparing results across the two systems, we identify which spectral features are influenced by the choice of the reference Hessian.

[1] T. Begušić, M. Cordova, and J. Vaniček, *J. Chem. Phys.*, **2019**, 150, 154117.

[2] T. Begušić, E. Tapavicza, and J. Vaniček, *J. Chem. Theory Comput.*, **2022**, 18, 3065.

[3] J. J. L. Vaniček, *J. Chem. Phys.*, **2023**, 159, 014114.

Transfer Learning in Heterogeneous Catalysis: Overcoming Hurdles in Applying Machine Learning Algorithms

V. Böttcher¹, C. Copéret^{1*}

¹ETH Zurich, D-CHAB, Laboratory of Inorganic Chemistry

Machine learning in heterogeneous catalysis suffers from many obstacles but among the most crucial are the low-data regime and the absence of a unique character and therefore distinctive computational representation for catalysts themselves. Atomistic descriptors and elemental loadings are used to supply models with a notion of the catalyst identity yet face struggles with the low data available.[1–3]

In applying transfer learning for heterogeneous catalysis, the pool of available data can be increased through combining data from different reactions to optimize catalysts for another reaction system. Transferring data from only one reaction that shares similarities with the studied system can already severely increase dataset size. This poster will discuss how to utilize the data obtained from testing catalysts in one reaction as training data for another seemingly related reaction. While increasing the dataset without running additional tests, the catalyst performance can serve as highly effective characterization descriptors – mending both stated limitations in the application of ML.

[1] Wang, G., Mine, S., Chen, D *et al.* *Nat Commun* **2023**, 14 (1), 5861.

[2] Ramirez, A., Lam, E., Gutierrez, D. P. *et al.* *Chem Catalysis* **2024**, 4 (2), 100888.

[3] Suzuki, K., Toyao, T., Maeno, Z. *et al.* *ChemCatChem* **2019**, 11 (18), 4537–4547.

Solution-state structure refinement of cyclic peptides leveraging NOE restraints in REST2 sampling

P. Brandl¹, L. Glazer^{1*}, M. Ebert^{2*}, S. Riniker^{1*}

¹Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland, ²Laboratory of Organic Chemistry, ETH Zürich, D-CHAB, Vladimir-Prelog-Weg 3, 8093 Zürich, Switzerland

Sampling the conformational space of cyclic peptides poses significant challenges due to their high flexibility. Cyclosporin A is a structurally well-understood example with a dominant conformation (ca. 94%) in chloroform [1, 2]. Sampling this conformation in molecular dynamics (MD) simulations when starting from a random starting structure remains challenging as it requires overcoming a cis-trans peptide bond isomerization. Even when applying distance restraints from Nuclear Overhauser Effect (NOE) experiments during the simulation, the correct conformation could not be found, unless starting from a near-by conformation [3]. With enhanced-sampling techniques like REST2 [4] these barriers can be overcome but typically converge slowly.

Hence, to improve its practical application for NMR structure refinement, we combined REST2 sampling with NOE distance restraints experiments to steer the sampling to the relevant conformational space. Furthermore, time-averaging enables the generation of an appropriate multi-conformation ensemble. We applied the protocol to the even larger cyclic peptide SaD3 in methanol [5]. The resulting ensemble satisfies NOE restraints significantly better than that of long unrestrained REST2 simulations. This is further supported by our data on J-couplings, which indicate the absence of a single dominant conformation. We therefore anticipate that combining the enhanced sampling REST2 methodology with available NMR data is a promising route for structure refinement of this emerging class of conformationally challenging therapeutics.

[1] Jagna Witek et al., *Journal of Chemical Information and Modeling*, **2016**, 56, 8, 1547–62

[2] Horst Kessler et al., *Helvetica Chimica Acta*, **1990**, 73, 7, 1818–32

[3] Jagna Witek, Doctoral Thesis, *ETH Zurich*, **2018**

[4] Lingle Wang, et al., *The Journal of Physical Chemistry B*, **2011**, 115, 30, 9431–38

[5] Renier H.P Neer, et al., *ACS Chemical Biology*, **2022**, 17, 8, 2284–95

AI-Driven Molecular Generation for GDB-20Y. Buehler¹, J. Reymond^{1*}¹Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Switzerland

AI-driven molecular generation presents a transformative approach for exploring chemical space, particularly in the context of GDB-20. This study proposes a dual-component research framework comprising molecular decoration and graph enumeration, which operate in parallel.

The molecular decoration phase leverages a sequence-to-sequence transformer model trained on a mixed GDB dataset (GDB-11, GDB-13s^[1, 2], and a subset of GDB-17), ensuring data diversity and mitigating potential data leakage through clustering and graph-based data splitting.

In the graph enumeration phase, mathematical graphs with up to 20 nodes are generated and filtered to retain only planar graphs. These graphs are subsequently SMILESified into hydrocarbons (carbon and hydrogen atoms only), forming the GDB-20 graph set^[3]. Upon completion, the trained transformer model translates these graphs into molecular structures by decorating them with double/triple bonds, heteroatoms, and aromatic systems.

The proposed framework provides a rapid and automated alternative to traditional molecular generation methods, significantly enhancing efficiency while preserving chemical diversity.

[1] Ye Buehler, Jean-Louis Reymond, J. Chem. Inf. Model., 2023, 63 (2), 484–492.

[2] Ye Buehler, Jean-Louis Reymond, J. Chem. Inf. Model., 2024, 64 (4), 1234–1245.

[3] Ye Buehler, Jean-Louis Reymond, J. Chem. Inf. Model., 2025, 65 (1), 567–578.

De novo material design using diffusion generation and reinforcement learning

J. Chen^{1,2}, J. Guo^{1,2*}, P. Schwaller^{1,2*}, P. Schwaller^{1,2*}

¹Laboratory of Artificial Chemical Intelligence (LIAC), Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ²National Centre of Competence in Research (NCCR) Catalysis, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Recent advances in diffusion models have enabled increasing capabilities for inverse materials design. The key capability is achieving tailored design towards desired property profiles, with wide applications for climate change, semiconductor design, and catalysis. In this work, we present MatInvent, a reinforcement learning (RL) framework tailored to optimize diffusion models for goal-directed crystal generation. By formulating equivariant denoising as a multi-step decision-making problem, MatInvent leverages policy optimization with reward-weighted KL regularization, including experience replay and diversity filters to enhance sample efficiency and diversity. Experimental results demonstrate that MatInvent outperforms existing baselines, offering an effective strategy for crystal generation with single or multiple property optimization.

Polypharmacology Browser PPB3: A Web-based Deep Learning Tool for Target Prediction Using ChEMBL Data

M. Darsaraee¹, J. Reymond^{1*}

¹Department of Chemistry, Biochemistry and Pharmacy, University of Berne, Freiestrasse 3, 3012 Berne, Switzerland

Polypharmacology is the concept that a single small molecule drug can interact with multiple targets simultaneously and identifying drug-target interactions plays a key role during the early stages of drug discovery. To address this, several web-based target prediction tools have been developed to identify the potential targets of small molecules using bioactive compound data from databases such as ChEMBL, employing similarity searching, logistic regression and machine learning models to predict the polypharmacology of any query compound and potential drug candidates.^[1] In our group we have developed the polypharmacology browsers (PPB and PPB2),^[2,3] which assign possible targets to a query molecule based on molecular fingerprint similarities to ChEMBL molecules, and provided critical insights in several practical case studies. Unlike the previous versions of PPB, which its dataset is restricted to a single protein type and three organisms and employs similarity searching for target prediction, PPB3 overcomes this limitation by incorporating all target types and organisms, creating a broader and more comprehensive dataset featuring a total of 1.1 million molecules and 2.5 million molecule-target associations. Furthermore, we developed PPB3 using a new set of molecular fingerprints including six single fingerprints and a fused fingerprint to train on seven deep neural networks models. Notably, we applied PPB3 on (*R,R*)-**1a**, which is recently reported as a norepinephrine transporter by our group,^[4] and it successfully predicted all four off-targets of (*R,R*)-**1a** with predicting norepinephrine transporter as its first ranked prediction.

Figure 1. Overview of the PPB3 web-based target prediction tool workflow. (a) Query compound input panel in PPB3, featuring the JSME editor with an example molecule (*R,R*)-**1a** and its corresponding SMILES. (b) Selection menu for prediction models in PPB3, offering seven DNN-based methods using various molecular fingerprints. (c) Prediction output window for query compound (*R,R*)-**1a** showing the top-ranked targets along with ChEMBL IDs, names, confidence scores, and visual summaries of predicted protein classes, organisms, and target types. (d) Example of the nearest neighbor panel displaying the top ten similar compounds for a selected target ranked by the Tanimoto coefficient, accessible via the “Display NN” button in the results table.

Keywords: target prediction, polypharmacology, deep neural networks, molecular fingerprints, ChEMBL, drug–target interactions

[1] M. Darsaraee, S. Javor, J.-L. Reymond, in ‘Polypharmacology’, John Wiley & Sons, Ltd, **2025**, pp. 223.

[2] M. Awale, J. L. Reymond, *J. Cheminf.* **2017**, 9, 11.

[3] M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2019**, 59, 10.

[4] A. Carrel, A Yiannakas, J. Roukens, I. Reynoso-Moreno, M. Orsi, A. Thakkar, J. Arus-Pous, D. Pellegata, J. Gertsch, and J.-L. Reymond, *Journal of Medicinal Chemistry*, **2025**, doi.org/10.1021/acs.jmedchem.4c02549.

Ab initio Calculations for Fe₄ L-edge X-ray Adsorption Spectroscopy (XAS)

N. Dongfang¹, F. Totti^{2*}, M. Iannuzzi^{1*}

¹Department of Chemistry, University of Zurich, Zurich, 8057 Switzerland, ²Dipartimento di Chimica “Ugo Schiavoni” & INSTM RU, Università degli Studi di Firenze, Via della Lastruccia 3, Sesto Fiorentino (FI) 50019, Italy

X-ray absorption spectroscopy (XAS) at the 2p ($L_{3,2}$) edge is a powerful, element-specific technique for probing the local electronic structure of 3d transition metals in both molecular and condensed phases. However, the accurate theoretical modeling of 2p XAS remains challenging due to strong electron correlations and spin-orbit effects. In this work, we employ linear-response time-dependent density functional theory (LR-TDDFT), including spin-orbit coupling, to simulate the 2p XAS spectra of a complex multinuclear magnetic system: the tetrairon(III) Fe₄H cluster and its Li-doping derivatives. Our method successfully reproduces key spectral features observed in experiments, including the $L_{2,3}$ edge splitting and the contributions from both spin-alpha and spin-beta excitation channels. We demonstrate that accurate reproduction of the electronic structure—particularly oxidation and spin states—is more critical than an exact magnetic ground state for obtaining reliable spectra, alleviating common concerns regarding the limitations of DFT for strongly correlated systems. Furthermore, we show that reduced peripheral Fe by Li-doping has a 1.58 eV shift in $L_{3,2}$ -edge without affecting the $L_{3,2}$ -edge splitting. This study establishes a practical and accurate framework for simulating 2p XAS in complex magnetic molecules, providing valuable insight into their electronic behavior and guiding future experimental and theoretical investigations in quantum materials and molecular magnetism.

Bayesian Optimization of Surface Reaction Dynamics via Surrogate Molecular RepresentationsE. Fako^{1,2}, P. Schwaller^{1,2*}¹Ecole Polytechnique Federale de Lausanne (EPFL), Rte Cantonale, Lausanne, 1015, Switzerland, ²National Centre of Competence in Research (NCCR) Catalysis, Switzerland

Catalysis and surface science are essential to addressing societal challenges in energy, environment, and chemical manufacturing, where atomic-level understanding of catalytic mechanisms is critical. These processes are often controlled by rare events, which can only be effectively probed using atomistic simulations.[1-3] Yet, even the most advanced methods face limitations in sampling the complex and high-dimensional potential energy surfaces (PES) where such events occur. To overcome this challenge, Bayesian optimization (BO) has been proposed as a strategy to actively steer on-the-fly simulations toward informative regions of chemical space.[4-7] In this work, we demonstrate that encoding reactive fragments using surrogate SMILES (***SMILES**) enables a simple yet expressive parametric description of surface–fragment bonding. A Gaussian process model, is used to bias molecular dynamics toward uncertain or chemically interesting regions of the parameter space. This allows efficient exploration of the PES and targets catalytically relevant, otherwise hard-to-access states. The approach is broadly generalizable across materials and chemical species and enables the discovery of new reaction pathways and mechanistic insights.

[1] Nørskov, J.K., Abild-Pedersen, F., Studt, F., Bligaard, T., Proceedings of the National Academy of Sciences, 2011, 108(3), 937–943

[2] Dattila, F., Seemakurthi, R.R., Zhou, Y., Lopez, N., Chem. Rev., 2022, 122(12), 11085–11130

[3] Medford, A.J., Vojvodic, A., Hummelshøj, J.S., Voss, J., Abild-Pedersen, F., Studt, F., Bligaard, T., Nilsson, A., Nørskov, J.K., Journal of Catalysis, 2015, 328, 36–42

[4] Reuter, K., Scheffler, M., Physical Review B, 2006, 73(4)

[5] Todorovic, M., Gutmann, M.U., Corander, J., Rinke, P., npj Computational Materials, 2019, 5(1)

[6] Carr, S., Garnett, R., Lo, C., Proceedings of Machine Learning Research, 2016, 48, 898–907

[7] Packwood, D., Briefs in the Mathematics of Materials. Springer, Singapore, Singapore, 2017

Polaritonic chemistry: Modifying chemical reactivity with mirrorsM. R. Fiechter¹, J. O. Richardson^{1*}¹Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland

Placing a pair of mirrors around a reaction mixture allows one to strongly couple molecular vibrations to a trapped light mode. This results in the formation of new hybridized quantum states called “polaritons”, which consist in equal parts of matter and light excitations.

Experiments have demonstrated that this vibrational strong coupling can significantly change molecular properties, such as ground-state reactivity. However, despite the large body of theoretical work in this field, the mechanism behind these changes in reactivity is still not well-understood. In this presentation, I aim to introduce the field, give an idea of why the experimental results have been so hard to capture theoretically, and propose effects that have been missing from our descriptions so far that could be responsible for the experimental observations.

[1] M.R. Fiechter, J.E. Runeson, J.E. Lawrence, J.O. Richardson, *J. Phys. Chem. Lett.*, **2023**, *14*(36), 8261-8267.

[2] M.R. Fiechter, J.O. Richardson, *J. Chem. Phys.*, **2024**, *160*(18).

Evaluating vibronic spectra beyond the Herzberg-Teller approximation with Hagedorn wavepackets

M. Gandolfi¹, J. Vaníček^{1*}

¹École Polytechnique Fédérale de Lausanne

In most calculations of vibrationally resolved electronic spectra, the transition dipole moment is assumed to be either a constant or a linear function of nuclear coordinates. We describe an efficient method, which goes beyond these so-called Condon and Herzberg-Teller approximations and allows for an arbitrary polynomial dependence of the transition dipole moment on nuclear coordinates. To propagate a wavepacket in the form of a Gaussian multiplied by an arbitrary polynomial, we expand the initial state in Hagedorn's basis (1-3). In harmonic systems, the coefficients of Hagedorn functions stay constant, and one only needs to evolve the guiding Gaussian's parameters and evaluate the autocorrelation function of Hagedorn wavepackets using an exact recursive expression (4). To validate the method, we compare its results with exact quantum benchmarks for displaced, distorted, and Duschinsky-rotated harmonic models and for different polynomial dependences of the transition dipole moment. In *ab initio* applications, the bottleneck is the evaluation of the electronic structure required for the propagation of the guiding Gaussian. Therefore, the proposed method can evaluate corrections to the Condon and Herzberg-Teller spectra at a negligible additional cost, similarly to how Hagedorn wavepacket dynamics has been used to evaluate single-vibronic level fluorescence spectra from excited levels (5,6).

1. G. A. Hagedorn, "Raising and Lowering Operators for Semiclassical Wave Packets," Ann. Phys. (NY) 269, 77–104 (1998).
2. C. Lubich, From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis, 12th ed. (European Mathematical Society, Zurich, 2008).
3. C. Lasser and C. Lubich, "Computing quantum dynamics in the semiclassical regime," Acta Numerica 29, 229–401 (2020).
4. J. J. L. Vaníček and Z. T. Zhang, "On Hagedorn wavepackets associated with different Gaussians," J. Phys. A 58 085303 (2025)
5. Z. T. Zhang and J. J. L. Vaníček, "Single vibronic level fluorescence spectra from Hagedorn wavepacket dynamics," J. Chem. Phys. 161, 111101 (2024)
6. Z. T. Zhang, M. Visegrádi, and J. J. L. Vaníček, "On-the-Fly Ab Initio Hagedorn Wavepacket Dynamics: Single Vibronic Level Fluorescence Spectra of Difluorocarbene," Phys. Rev. A 111, L010801 (2025).

REST2-AMP/MM: Integrating Enhanced Sampling with Machine Learning Potentials for Molecular Conformational Sampling

I. Gordiy¹, R. Solazzo^{1*}, S. Riniker^{1*}

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

Numerical potentials from machine learning (ML) have been proposed to replace computationally expensive quantum-mechanical (QM) calculations. Anisotropic message passing (AMP) is an equivariant graph neural network potential that incorporates directional and long-range interactions, making it particularly suitable for QM/MM hybrid schemes [1,2]. While AMP is computationally much cheaper than DFT (at similar accuracy), ML/MM molecular dynamics (MD) simulations can still encounter sampling difficulties. In classical MD, replica exchange with solute scaling (REST2) is a popular sampling-enhancement method, which works by selectively scaling solute interactions while maintaining a fixed bath temperature [3]. This reduces the number of replicas required compared to standard temperature replica exchange and thus increases the computational efficiency. Here, we present the combination of REST2 with AMP, which is made possible because the AMP Hamiltonian – in contrast to QM Hamiltonians – can be decomposed into solute-solute, solvent-solvent, and solute-solvent interaction contributions. Due to this feature, solute interactions in AMP can be selectively scaled by REST2. We validate the REST2-AMP approach on the popular test system alanine dipeptide, achieving good sampling coverage and capturing dihedral transitions within a short simulation time. These results suggest that REST2-AMP enables efficient conformational sampling while providing a more accurate description of biomolecular systems, capturing effects that may be omitted by classical force fields.

[1] Moritz Thürlmann, Sereina Riniker, *Proceedings of the International Conference on Learning Representations*, **2023**.

[2] Felix Pultar, Moritz Thürlmann, Igor Gordiy, Eva Doloszeski, Sereina Riniker, *Journal of the American Chemical Society*, **2025**, *147*, 6835-6856.

[3] Lingle Wang, Richard A. Friesner, Bruce J. Berne, *The Journal of Physical Chemistry B*, **2011**, *115*, 9431-9438.

Computational Design of Enzymes to Degrade Short-chain Per- and Polyfluoroalkyl Substances (PFAS)

S. Guilbert¹, U. Röthlisberger^{1*}, P. Schwaller^{1*}

¹Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

Increasing concern over the persistence of per- and polyfluoroalkyl substances (PFAS), often referred to as “forever chemicals”, has pushed for the need to design technologies for their remediation [1]. Specifically, trifluoroacetate (TFA) is the main degradation product of longer-chain PFAS and its exposure is linked with various health defects. Bioremediation stands as an attractive, cost-effective and sustainable solution for degrading such pollutants, although no bacteria or enzymes have yet been discovered to achieve the full degradation of TFA.

The *in silico* rational design of novel enzymes for PFAS degradation first involved the selection of a suitable biological target. A minimal active site was created with the catalytic residues of fluoroacetate dehalogenase, an enzyme capable of breaking the strong C-F bond (~460 kJ/mol) in fluoroacetate [2]. Machine learning tools such as RFDiffusion, ProteinMPNN and AlphaFold were used to design a scaffold around the active site, since a smaller biomimetic design would facilitate further computational optimisation [3,4]. The design pipeline yielded 180,000 proteins. The best biomimetic designs were selected based on various criteria, and were optimised to accept the novel substrate: TFA.

Substrate binding was first enhanced using MM-PB(GB)SA in the first place, then alchemical free energy methods [5]. Genetic algorithm tools such as EVOLVE were then used to improve both the reactivity and thermal stability of the enzymes [6]. The EVOLVE physics based fitness function could assess the stability of multiple mutants while increasing the electrostatics of the catalytic pocket. Finally, to uncover the reaction mechanism of the novel enzymes with TFA, QM/MM simulations will be performed, to compute C-F bond dissociation energies.

- [1] E. Sunderland, X. Hu *et al.*, *Journal of Exposure Science & Environmental Epidemiology*, **2019**, 29, 131–147
- [2] S. Farajollahi, N. Lombardo *et al.*, *ACS Omega*, **2024**, 9, 28546–28555
- [3] R. Krishna ., *Science*, **2024**, 384
- [4] J. Jumper, R. Pritzel *et al.*, *Nature*, **2021**, 596, 583–589
- [5] B. Miller *et al.*, *Journal of Chemical Theory and Computation*, **2012**, 8, 3314–3321
- [6] N. Browning, M. Perez, E. Brunk, U. Rothlisberger, *ChemRxiv*, **2023**

Implementation of non-empirical DFT+U in the Gaussian and plane waves framework

K. Hanasaki¹, S. Lubert^{1*}

¹Department of Chemistry, University of Zürich

We present our implementation of non-empirical DFT+U^[1] and its application to real-time TDDFT (RT-TDDFT) in the CP2K software suite^[2]. Non-empirical DFT+U refers to DFT+U calculations using Hubbard parameters calculated by ab initio methods. It realizes cost-efficient calculations of strongly correlated systems without empirical parameters. In this work, we examined two approaches, one is based on the linear-response theory^[3,4], in which the Hubbard parameters are computed from the dynamical response of the system to small perturbations. The other one is the new framework of DFT proposed by Agapito, Curtarolo, and Buongiorno Nardelli^[5], known as ACBN0, in which the Hubbard parameters are computed using a renormalized occupation number reflecting the static correlation. The former approach is conceptually clear, and its validity has been established in the literature through a number of ground-state calculations, whereas the latter has the advantages of small computational cost and applicability to dynamical RT-TDDFT calculations^[6]. We implemented both approaches in the k-point sampling DFT+U using an efficient local projection scheme^[7] whose validity has been examined by static and dynamical calculations^[8]. We report details of implementation in the Gaussian and plane waves framework in CP2K and discuss the validity and limitations of these approaches.

[1] V. I. Anisimov, J. Zaanen, O. K. Andersen, *Phys. Rev. B*, 1991, 44, 943-954.

[2] <https://cp2k.org/>

[3] M. Cococcioni, and S. de Gironcoli, *Phys. Rev. B*, 2005, 71, 035105.

[4] G. Moynihan, G. Teobaldi, and D. D. O'Regan, *arXiv preprint*, 2017, arXiv:1704.08076.

[5] L. A. Agapito, S. Curtarolo, and M. Buongiorno Nardelli, *Phys. Rev. X*, 2015, 5, 011006.

[6] N. Tancogne-Dejean, M. J. T. Oliveira, and A. Rubio, *Phys. Rev. B*, 2017, 96, 245133.

[7] Z. Chai, R. Si, M. Chen, G. Teobaldi, D. D. O'Regan, and L.-M. Liu, *J. Chem. Theory Comput.*, 2024, 20, 8984–9002

[8] K. Hanasaki, S. Lubert, *J. Chem. Theory Comput.*, 2025, 21, 1879–1891

Generative Modeling of Cyclic ConformersA. Hartgers^{1,2}, L. Schaufelberger^{1,2*}, K. Jorner^{1,2*}¹Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland, ²NCCR Catalysis, Switzerland

Molecules containing cycles are omnipresent across chemical space, ranging from playing a central role in drug discovery and catalysis to appearing in the majority of biochemical structures, with their three-dimensional conformational landscape governing important properties such as binding affinities and catalytic activities. Generating diverse conformers of cyclic molecules remains highly challenging due to steric constraints. Here, we introduce PuckerFlow, a generative machine learning model that efficiently samples conformers of cyclic compounds. By performing the generation process on the Cremer-Pople coordinates, which characterize out-of-plane elevation of atoms in cycles, we capture the relevant internal degrees of freedom and guarantee the generation of only closed rings. Our model surpasses the performance of knowledge-based cheminformatics methods such as RDKit's ETKDG algorithm in terms of capturing conformational diversity and naturally scales across ring sizes. PuckerFlow enables targeted design of cyclic drugs and catalysts, and we aim to extend its applications to macrocycles in future work.

Efficient 3D Virtual Screening of Combinational Spaces

S. M. Hönig^{1,2}, R. Klein^{1*}, M. Rarey^{2*}, C. Lemmen^{1*}

¹BioSolveIT GmbH, An der Ziegelei 79, 53757 Sankt Augustin, Germany, ²University of Hamburg, ZBH - Center for Bioinformatics, Albert-Einstein-Ring 8-10, 22761 Hamburg, Germany

Over the last few years, make-on-demand chemical libraries have gained enormous popularity. Their extension beyond the limits for exhaustive enumeration brought forth combinatorial approaches to give access to these ever-growing sets of compounds as a so-called Chemical Fragment Space.[1, 2] Combinatorial approaches enable the screening of trillions of compounds or more with moderate hardware and in a reasonable time frame. Established and extremely performant ligand-based 2D methods exist in this context,[3] however, recent developments focus on the need for 3D methods.[4-6]

Responding to this new challenge in cheminformatics, we present efficient ways to explore make-on-demand Chemical Fragment Spaces with 3D shape-based virtual screenings. The advantages of combinatorial approaches in runtime and the generation of superior results become tangible in comparison to methods operating on enumerated libraries. [7]

A drug discovery workflow was investigated with the above-described combinatorial strategy to identify compounds with similar binding capabilities and molecular novelty in different design scenarios.

- [1] T. Hoffmann and M. Gastreich, *Drug Discov Today*, **2019**, 24 (5), 1148–1156.
- [2] W. A. Warr et al., *J Chem Inf Model*, vol. 62, **2022**, 9, 2021–2034.
- [3] M. Korn et al., *Curr Opin Struct Biol*, **2023**, 80, 102578.
- [4] P. Beroza et al., *Nat Commun*, **2022**, 13 (1), 6447.
- [5] J. Müller et al., *J Med Chem*, **2022**, 65 (23), 15663–15678.
- [6] J. Lyu et al., *Nature*, 566, **2019**, 7743, 224–229.
- [7] S. M. N. Hönig et al., *J Comput Aided Mol Des*, **2024**, 38 (1), 13.

Lifelong Machine Learning Potentials for Reaction Network Explorations

R. T. Husistein¹, M. Reiher^{1*}

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

Automated exploration [1-3] of chemical reaction spaces is a powerful new avenue to understand complex reaction mechanisms at unprecedented depth. However, ab initio calculations are too slow for exhaustive explorations, which is why semi-empirical methods such as GFNn-xTB are often used. An accurate and very efficient alternative to semi-empirical methods are machine learning models to represent the potential energy surface.

One such approach is based on high-dimensional neural network potentials [4]. They can handle systems with a large number of atoms and they respect rotational, translational, and permutational invariance. However, by using atom-centered symmetry functions (ACSF), the number of inputs increases with the number of different elements in the system, making them hard to apply to large systems. To improve on this unfavorable scaling, we have introduced element embracing (ee)ACSF [5]. They ensure that the input size remains constant regardless of the number of elements.

Since machine learning potentials provide poor predictions for structures that are very different from the training set, it is necessary to continuously train the machine learning potentials when such a situation is detected. For this purpose, we have developed a new lifelong-adaptive data-selection training approach and the new continual resilient optimizer [5]. We note that the new optimizer also performs very well for other training tasks [6].

While this allows us to continuously train the network, we also need to ensure that the capacity of the network is sufficient to learn the new data. For this purpose, we have developed the Network Expressivity by Activation Rank (NEAR) metric [7]. This metric can be computed on an untrained network and shows a high correlation with the final accuracy. It demonstrates state-of-the-art performance on several neural architecture search spaces. In addition, our tests show that we can use NEAR to define a good layer size for lifelong machine learning potentials.

- [1] G. N. Simm, A. C. Vaucher, M. Reiher, *J. Phys. Chem. A*, **2018**, *123*, 385-399.
- [2] J. P. Unsleber, M. Reiher, *Annu. Rev. Phys. Chem.*, **2020**, *71*, 121-142.
- [3] A. Baiardi, et al., *Acc. Chem. Res.*, **2022**, *55*, 35-43.
- [4] J. Behler, M. Parrinello, *Phys. Rev. Lett*, **2007**, *98*, 146401.
- [5] M. Eckhoff, M. Reiher, *J. Chem. Theory Comput.*, **2023**, *19*, 3509-3525.
- [6] M. Eckhoff, M. Reiher, *Mach. Learn.: Sci. Technol.*, **2024**, *5*, 015018.
- [7] R. T. Husistein, M. Reiher, M. Eckhoff, *arxiv: 2408.08776*, **2024**.

Substrate Decoration by Combining 3D Generative Modelling with High-throughput DFT Calculations for Reactivity Design

Z. Ivkovic^{1,2}, K. Jorner^{1,2*}

¹Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland, ²NCCR Catalysis, Switzerland

Scaffold elaboration, or substrate decoration, is a cornerstone of molecular design in drug discovery [1,2,3] but remains underutilized in the context of reactivity design. For instance, scaffold elaboration can be used for further tuning reactive properties of substrates obtained from larger search campaign. [4] This gap is largely attributed to data scarcity and the lack of generative models explicitly tailored for reaction-centric applications. While existing 3D generative frameworks possess the capacity for scaffold decoration [5,6], they are not purpose-built for this task, limiting their utility in conditioning generation on reactive properties. In this work, we present a specialized generative modelling approach aimed at scaffold elaboration for reactivity design. We construct a curated dataset of around 10'000 DFT-level keto-enol tautomerization reactions by subsampling a combinatorial library sharing a common *N*-Salicylideneaniline core scaffold. We train a 3D structure-generating model on this dataset to produce transition state geometries that exhibit desired reactive properties. The inherent advantage of our approach over existing 2D approaches is easy validation due to generation of transition states, which can directly be used as input to transition state optimization. Our approach demonstrates the potential of targeted scaffold elaboration to advance computational reactivity design and provides a foundation for future developments towards catalyst design.

[1] Thomas E. Hadfield, Fergus Imrie, Andy Merritt, Kristian Birchall, and Charlotte M, *Journal of Chemical Information and Modeling*, **2022**, 62, 2280–2292.

[2] Xuhan Liu, Kai Ye, Herman W. T. van Vlijmen, Adriaan P. IJzerman, Gerard J. P. van Westen, *Journal of Cheminformatics*, **2023**, 15, 24.

[3] Jaechang Lim, Sang-Yeon Hwang, Seokhyun Moon, Seungsu Kim, Woo Youn Kim, *Chemical Science*, **2020**, 11, 1153–1164.

[4] Thijs Stuyver, Connor W. Coley, *Chemistry Europe Journal*, **2023**, 29, 28

[5] Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom L. Blundell, Pietro Lio, Max Welling, Michael Bronstein, Bruno Correia, *Nature Computational Science*, **2024**, 4, 899–909.

[6] Julian Cremer, Ross Irwin, Alessandro Tibo, Jon Paul Janet, Simon Olsson, Djork-Arné Clevert, *ArXiv*, **2025**, 2504.10564

Rotamer-dependent amino acid descriptors library

L. Jacot-Descombes^{1,2}, K. Jorner^{1,2*}

¹Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland, ²NCCR Catalysis, Switzerland

Abstract: While machine learning methods offer a promising approach for enzyme design, the models are often hindered by the limited accuracy of molecular representations^[1]. To address this, we introduce a novel amino acid descriptor library which explicitly accounts for rotameric structures, moving beyond conventional single-conformation descriptors. Rotamer geometries are optimized using GFN2-xTB^[2], and stereo-electronic properties such as Sterimol parameters^[3], buried volume^[4], solvent-accessible surface area^[5], HOMO/LUMO energies, Fukui functions^[6] and bond orders^[7], are computed. We envision that these descriptors will enhance amino acid representations for bioactivity prediction, improving machine learning models in enzyme engineering. On-going work includes the development of a Python package enabling descriptors query of targeted rotamer or frequency-weighted averaging across rotamers.

References

- [1] R. Michael et al., *PLOS Computational Biology* **2024**, 20 (5), e1012061.
- [2] C. Bannwarth et al., *J. Chem. Theory Comput.*, **2019**, 15 (3), 1652–1671.
- [3] A. Verloop, The Sterimol approach: Further development of the method and new applications, Pergamon Press, **1983**, pp. 339–344.
- [4] S. Escayola et al., *Chem. Soc. Rev.*, **2024**, 53 (2), 853–882.
- [5] B. Lee et al., *J. Mol. Biol.*, **1971**, 55 (3), 379-IN4.
- [6] P. Geerlings et al., *Chem. Rev.*, **2003**, 103 (5), 1793–1874.
- [7] I. Mayer, *J. Comput. Chem.*, **2007**, 28 (1), 204–221.

Multiscale Modeling for DNA Oxidative Damage due to Endocrine Disrupting Chemicals: From Onset to Detection

S. K. Johnson¹, S. Guilbert^{1*}, M. Venkatanarayanan^{1*}, U. Rothlisberger^{1*}

¹École Polytechnique Fédérale de Lausanne

Our DNA faces over 10 000 damage incidents daily per cell, of which the majority are oxidative lesion sites. The species which promote oxidative damage and the systems which repair these lesions are of paramount important in our genomic integrity. Current theoretical modeling of redox properties of DNA regions only provides a perspective on small, localized regions. While useful for comparisons of local, oxidative damage types across similar systems [1,2], there is a broader question as to the overall accuracy of the measured electronic properties, especially as stacks of nucleobases can easily delocalize electron density over sizeable regions [3]. Thankfully, advances in high performance computing (HPC) and parallelization schemes within software ensure a continual improvement in computational modeling. We apply a well-known multiscale modeling approach, QM/MM MD (quantum mechanics/molecular mechanics molecular dynamics) using the new highly-scalable MiMiC framework code in order to study DNA oxidative damage from onset to its detection by repair enzymes [4]. First, we determine the modeling parameters necessary to achieve better convergence of DNA redox properties with respect to the size of its subsystem (from three to seven base pair stacks) that is treated at quantum mechanical level within the context of a solvated helix treated at the molecular mechanics level in order to provide further evidence (or lack thereof) toward the proposed charge transfer scheme for oxidative damage detection by DNA repair enzymes [5, 6]. With these better modeling parameters, we simulate the possible mechanisms of oxidative damage in DNA due to radical, semiquinone metabolites of estrogenic, endocrine disrupting chemicals (EDCs) – a chemical class garnering increasing concern for human and environmental health. Estrogenic EDCs share structural similarities to endogenous estrogens and are suspected to undergo the same metabolic processes as estrogens to form reactive quinone and semiquinone species [7]. Finally, we provide a rationalization of potential genotoxicity of different subtypes of EDC metabolites, both semiquinone and quinone, based on conceptual DFT studies of reactivity. Together, we offer a practical set of modeling parameters for the efficient and accurate simulation of solvated DNA systems for measuring DNA electronic properties as well as provide insights into the reactivity of small molecules with DNA bases.

- [1] P. Diamantis, I. Tavernelli, U. Rothlisberger, *J. Chem. Theory Comput.*, **2020**, 10, 6690–6701.
- [2] M. Kılıç, P. Diamantis, S. K. Johnson, O. Toth, U. Rothlisberger, *J. Chem. Theory Comput.*, **2023**, 19, 8434–8445.
- [3] J. C. Genereux, S. M. Wuerth, J. K. Barton, *J. Am. Chem. Soc.*, **2011**, 133, 11, 3863–3868.
- [4] J. M. H. Olsen, V. Bolnykh, S. Meloni, E. Ippoliti, M. P. Bircher, P. Carloni, U. Rothlisberger, *J. Chem. Theory Comput.*, **2019**, 15, 3810–3823.
- [5] K. J. McDonnell, J. A. Chemler, P. L. Bartels, E. O'Brien, M. L. Marvin, J. Ortega, R. H. Stern, L. Raskin, G.-M. Li, D. H. Sherman, J. K. Barton, S. B. Gruber, *Nature Chem*, **2018**, 10, 873–880.
- [6] T. P. Mui, J. O. Fuss, J. P. Ishida, J. A. Tainer, J. K. Barton, *J. Am. Chem. Soc.*, **2011**, 133, 16378–16381.
- [7] E. Cavalieri, K. Frenkel, J. Liehr, E. Rogan, D. J. Roy, *Natl. Cancer Inst., Monogr.*, **2000**, 27, 75–93.

Femtosecond Excitation and Charge Migration in hBN/C₃N₄–Electrophile Systems studied by Real-Time TDDFT

F. Keller¹, M. R. Coates^{1*}, M. Iannuzzi^{1*}

¹University of Zurich, Department of Chemistry, Winterthurerstrasse 190, CH-8057 Zurich

Real-time time-dependent density functional theory (RT-TDDFT) has emerged as a popular method to investigate ultrafast phenomena on the nanoscale. In our work we present the ultrafast electron dynamics in molecular electrophiles; 2D layered hBN and g-C₃N₄ materials, and their corresponding surface-electrophile interfaces, excited by a resonant laser pulse in the framework of the adiabatic local-density approximation using RT-TDDFT. The induced change in dipole moment and the evolution of excited-state populations provide valuable insight into the validity of simplified models for capturing essential features of ultrafast charge dynamics. By applying this approach, we can quantitatively characterize the transfer of charge carriers from light-absorbing 2D materials to adsorbed molecular states demonstrating the number of excited electrons, temporal evolution of the electronic populations, and the spatial distribution of transient charge density.

Our results offer a detailed interpretation of light-induced charge-transfer processes at the nanoscale and establish a robust computational framework for probing charge-carrier dynamics in complex material systems. This methodology provides a pathway toward understanding and controlling ultrafast optoelectronic phenomena relevant to photovoltaics and photocatalysis.

Autonomous Bifunctional Catalyst Design for the Morita-Baylis-Hillman reaction

S. Kopf^{1,2}, J. Guo^{1,3*}, J. Martin^{4*}, J. Schörgenhumer^{4*}, P. Schwaller^{1,3*}, C. Nevado^{4*}

¹Laboratory of Artificial Chemical Intelligence (LIAC), ISIC, EPFL, ²Department of Chemistry, University of Zurich, National Centre of Competence in Research (NCCR) Catalysis, EPFL, ³National Centre of Competence in Research (NCCR) Catalysis, EPFL, ⁴Department of Chemistry, University of Zurich

Efficient catalyst design remains a central challenge—and often dubbed the “holy grail”—of computational catalysis [1]. A common strategy toward this goal is virtual screening, where a scoring function (either physics-based or data-driven) is used to evaluate a library of candidate catalysts [2]. However, such approaches are limited by the scale and diversity of the input libraries.

De novo molecular design uses generative machine learning models to propose new molecules, enabling open-ended exploration of chemical space [2]. Goal-directed generative models can propose structures with tailored property profiles, as evaluated by a scoring function. However, in *de novo* settings, property prediction poses unique challenges: models may exploit spurious features of the scoring function to produce unrealistic or chemically irrelevant solutions—a phenomenon known as oracle exploitation or reward hacking [3].

Previous work by Seumer et al. introduced TS Score, an oracle for predicting the performance of small tertiary amine catalysts in the Morita-Baylis-Hillman (MBH) reaction [4]. In concert with a genetic algorithm, they identified the first experimentally-verified efficient *de novo* designed catalyst.

Here, we develop a pipeline for organocatalyst discovery that couples a language-based generative model, Saturn, with TS Strict—an adapted version of TS Score that better accommodates token-based generative models and large, flexible molecules [5]. Our pipeline autonomously discovers bifunctional amino-urea organocatalysts, with diverse urea-family groups mediating key noncovalent interactions. Notably, candidates include semicarbazides which are underrepresented in organocatalysts. Our results position generative models, paired with meaningful scoring functions, as both design engines and idea generators for the next generation of catalyst development.

[1] K. Houk, F. Liu, Accounts of Chemical Research, 2017, 50, 539-543

[2] B. Sanchez-Lengeling, A. Aspuru-Guzik, Science, 2018, 361, 360-365

[3] M. Langevin, R. Vuilleumier, M. Bianciotto, Journal of Cheminformatics, 2022, 14, 20

[4] J. Seumer, J. Kirschner Solberg Hansen, M. Brøndsted Nielsen, J. H. Jensen, Angewandte Chemie International Edition, 2023, 62, e202218565

[5] J. Guo, P. Schwaller, arXiv, 2024, preprint arXiv:2405.17066

Improving the Δ SCF method for excited electronic states

M. Malis¹, A. Sinyavskiy^{1*}, S. Lubert^{1*}

¹Department of Chemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

The simulation of photochemical and photophysical processes, vital for an in-depth understanding of these phenomena and their further development, relies on efficient and reliable electronic structure calculations of excited electronic states and their corresponding properties. These become ever more challenging with the system size, especially when applied to condensed phase systems. As each stationary excited electronic state has a corresponding electronic density, its energy can be obtained by the variational density functional theory (DFT) method, also known as delta self-consistent field (Δ SCF).[1,2] Obtaining all related excited electronic state properties accurately is not always straightforward, and while Δ SCF has a long track record for the calculation of excited electronic energies, it was rarely used for nonadiabatic molecular dynamics (NAMD) because of a cumbersome construction of singlet excited electronic states and their gradients.[1] Instead, constructing the corresponding singlet excited electronic state density using fractional occupation numbers, we were able to directly obtain its energy and nuclear gradients and used them for NAMD simulations in gas and condensed phase systems (with periodic boundary conditions), also using DFT embedding.[2,3,4] We also derived and implemented other excited electronic state properties, like transition dipole moments,[4,5] and spin-orbit couplings.[5] Here we present our recent work focused on improving the Δ SCF method with two new algorithms for maintaining the proper electronic density during the SCF optimizations.[6] This enabled a larger number of excited electronic states to converge. By examining the differences between converged electronic properties and reference properties generated with other electronic structure methods for excited states, it becomes evident that without proper tracking of excited electronic state density, not all converged excited state energies at the Δ SCF level correspond to the right targeted excited electronic states, and that energy alone cannot be used as a sole criterion for convergence. The new algorithms ensure that the right targeted electronic density is always reached. Furthermore, they are employed for converging electronic densities constructed of several singly excited electronic state determinants, enabling direct construction of the multi-configurational character excited electronic states and their corresponding properties with the Δ SCF method at the DFT level of theory. We also show how utilizing hybrid-based DFT functionals with fractional occupation numbers can lead to spurious shifts in singlet electronic state energies computed with Δ SCF.[7] The origin of this strange behavior is fully explained, and a mitigation procedure is developed. All these represent a small but important step forward in improving the Δ SCF method and its applications towards NAMD for simulating ever larger nonadiabatic processes.

- [1] E. Vandaele, M. Mališ, S. Lubert, *Journal of Chemical Physics*, **2022**, 156, 130901
- [2] M. Mališ, S. Lubert, *Journal of Chemical Theory and Computation*, **2020**, 16, 4071-4086
- [3] M. Mališ, S. Lubert, *Journal of Chemical Theory and Computation*, **2021**, 17, 1653-1661
- [4] E. Vandaele, M. Mališ, S. Lubert, *Physical Chemistry Chemical Physics*, **2022**, 24, 5669-5679
- [5] M. Mališ, E. Vandaele, S. Lubert, *Journal of Chemical Theory and Computation*, **2022**, 18, 4082-4094
- [6] A. Sinyavskiy, M. Mališ, S. Lubert, *in revision*
- [7] M. Mališ, S. Lubert, *submitted*

Spin-Exchange via Temperature-Dependent Transition States

D. Mukherjee¹, J. O. Richardson^{1*}

¹Institute of Molecular Physical Sciences

Isotopes of noble gases have a nonzero nuclear spin, which is well-shielded by the electron cloud and resist spin contamination by external agents[1]. However, spin-exchange phenomenon of the nuclear spin with another electronic spin causes coherent coupling of the spins [2], making them important candidates for building quantum materials [3]. Thus, it becomes increasingly important to understand the mechanism of the spin-exchange processes in these systems and to design various reactions that mimic the same phenomenon.

Here, we derive a new semiclassical transition-state theory (SCTST) which accurately describes the nonadiabatic transition between identical reactant and product potential energy surfaces (PESs), where classical TST and Landau–Zener theory fail. We apply our theory to describe the inelastic spin-exchange collision between the nuclear spin of ^3He and the electronic spin of ^{23}Na . SCTST shows how spin-exchange collision proceeds via a temperature-dependent transition state, determined by an intricate interplay of the potential energy and the hyperfine coupling and successfully explains the weak temperature dependence of the rates. Moreover, since SCTST requires calculation of a single point in the whole PES, the computation cost for the rate calculation is also reduced. In addition, we are also successful in formulating a new instanton theory[4] for reactions where the coupling has a nontrivial dependence on the nuclear geometry.

[1] T. R. Gentile, P. J. Nacher, B. Saam, and T. G. Walker, Optically polarized ^3He , Rev. Mod. Phys. 89, 045004 (2017).

[2] R. Shaham, O. Katz, and O. Firstenberg, Strong coupling of alkali-metal spins to noble-gas spins with an hour-long coherence time, Nature physics 18, 506 (2022).

[3] O. Katz, R. Shaham, and O. Firstenberg, Quantum interface for noble-gas spins based on spin-exchange collisions, PRX Quantum 3, 010305 (2022).

[4] J. O. Richardson, Ring-polymer instanton theory, Int. Rev. Phys. Chem. 37, 171 (2018).

Reinforcement learning environment for mechanism proposal in chemistry

T. A. Neukomm¹, Z. Jončev^{1*}, E. Fako^{1*}, P. Schwaller^{1,2*}

¹EPFL - Laboratory of Artificial Chemical Intelligence (LIAC), ²National Centre of Competence in Research (NCCR) Catalysis

Reinforcement learning (RL) has been used in the past to infuse suprahuman performances into machine learning (ML) agents playing games, surpassing the best human players at chess [1] or Go [2] among others [3]. In this work, we show that mechanism proposal for organic chemistry can be treated as a game in which the reward is based on the energy profile of the proposed pathway. This allows RL algorithms to be applied for improvement by self-play. In particular, we are releasing an environment with a formal set of rules framing the arrow-pushing formalism for ionic and radical chemistry over a large set of organic and transition-metal atoms. Based on this organic chemistry game, agents can compete in mechanism prediction for a given reaction, submitting their pathways to scoring algorithms that are either density functional theory (DFT) or machine learned interatomic potentials (MLIP)-based. By releasing such an environment, we hope to spark a fair and standardized competition that can be used as a benchmark. We also made this environment modular and open so that it is easily adaptable to new scoring algorithms better optimized for speed, precision or other specific requirements. In this work, we show that models pretrained in a supervised manner manage to improve their performance over a set of reactions through self-play, when getting feedback from a good scoring function.

[1] T. McGrath, A. Kapishnikov, N. Tomašev, A. Pearce, M. Wattenberg, D. Hassabis, ..., V. Kramnik, *Proceedings of the National Academy of Sciences*, 2022, 119(47), e2206625119.

[2] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, ..., D. Hassabis, *nature*, 2017, 550(7676), 354-359.

[3] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, ..., D. Hassabis, *Science*, 2018, 362 (6419), 1140-1144.

Modelling of Quantum-Controlled Molecular Collisions

S. patel¹, M. Meuwly^{1*}, S. Willitsch^{1*}

¹Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland

A high-accuracy potential energy surface (PES) for the ground electronic state of N_2^+ ($X^2\Sigma_g^+$) has been developed using the CCSD(T)/aug-cc-pVQZ level of theory. Key spectroscopic constants, including the equilibrium bond length (R_e), harmonic frequency (ω_e), and rotational constant (B_e), were computed and benchmarked against experimental data through a detailed analysis of ro-vibrational states [1].

Building on this work, a new PES for the N_2^+ ($X^2\Sigma_g^+$)–He(^1S) complex was constructed using a reproducing kernel Hilbert space (RKHS) approach [2]. This surface was based on an extensive set of ab initio energy points at the same level of theory. All bound ro-vibrational levels of the N_2^+ –He complex in its electronic ground state were computed. Furthermore, a comparative analysis of the para- and ortho- N_2^+ states was performed using wavefunction characterisation, providing accurate quantum state assignments.

The dynamics of collisions between N_2^+ and He atoms will be explored using both quantum mechanical and classical methods. The constructed PES will be employed to calculate scattering matrices and cross sections, which will subsequently be used to derive rate constants for the relevant collisional processes. This work lays the foundation for quantum-controlled studies of cold ion-neutral collisions, with implications for astrochemistry and controlled reactivity.

[1] D. Shi, W. Xing, J. Sun, Z. Zhu, Y. Liu, *Comput. Theor. Chem.* **966** (2011) 44–53. <https://doi.org/10.1016/j.comptc.2011.02.011>

[2] O. T. Unke, M. Meuwly, *J. Chem. Inf. Model.* **57** (2017) 1923–1931. <https://doi.org/10.1021/acs.jcim.7b00090>

QM/MM RE-EDS: an efficient multiscale and multistate free-energy methodD. Pregeljic¹, S. Riniker^{1*}¹Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

In computational chemistry, methods based on molecular dynamics (MD) simulations are arguably the most rigorous way of calculating free energies. However, the attainable accuracy is limited by the available computational resources. In practice, a trade-off has to be struck between model approximations and sufficient phase-space sampling. In the present work we address these two limitations of MD simulations by integrating multiscale QM/MM modeling with the multistate enhanced sampling technique replica-exchange enveloping distribution sampling (RE-EDS). We discuss the details of its implementation and validate it by calculating hydration free energies. We obtained converged results within chemical accuracy. Effects of QM method and classical water model selection are also highlighted. The developed methodology is QM-model agnostic and represents the first time that RE-EDS is utilized beyond classical force fields. This work also opens up new avenues for the introduction of any Hamiltonian, including machine-learned interatomic potentials (MLIPs), as well as applications to highly polarizable and systems poorly parameterized by fixed-charge force fields.

Advancing Yield Prediction in Transition Metal–Catalyzed Cross-Coupling Reactions via Graph Neural Networks

C. RAJALAKSHMI¹, S. SALIM^{1*}, V. I. THOMAS^{1*}

¹CMS COLLEGE KOTTAYAM, KERALA, INDIA

The integration of data science into catalysis research is revolutionizing the field by enabling unprecedented predictive capabilities. Catalytic reactions, particularly transition metal-catalyzed cross-coupling reactions, are indispensable tools in modern organic synthesis. Among the many machine learning applications in this domain, yield prediction has gained significant attention due to its potential to accelerate reaction optimization and discovery. Despite recent advancements, accurately predicting reaction yields across chemically diverse and structurally complex datasets remains a formidable challenge. In this study, we present a comparative evaluation of several Graph Neural Network (GNN) architectures—specifically MPNN, ResGCN, GraphSAGE, GAT, GATv2, GCN, and GIN—for yield prediction in a broad set of cross-coupling reactions. The dataset encompasses a range of well-known reactions, including Suzuki, Sonogashira, Cadiot–Chodkiewicz, Ullmann-type, and Buchwald–Hartwig couplings. Our analysis reveals that GNN-based models demonstrate robust and consistent performance across this heterogeneous dataset. Notably, the Message Passing Neural Network (MPNN) model achieved the highest predictive accuracy, with an R^2 value of 0.75. To enhance model interpretability, we applied the Integrated Gradients technique to quantify the contribution of individual RDKit-derived molecular descriptors to the predicted yields. This approach not only improves transparency but also offers mechanistic insight into the molecular features influencing reaction outcomes. Overall, this work underscores the promise of explainable, graph-based machine learning models in predicting chemical reaction yields. It marks a significant step toward the rational design and optimization of catalytic systems, contributing to advancements in sustainable chemistry and organic synthesis.

References

1. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science*. **360**, 186–190 (2018).
2. A self-driving lab for accelerated catalyst development. *Nat. Chem. Eng.* **1**, 206–207 (2024).

Dielectric-dependent hybrid functional based on meta-GGA

S. Riemelmoser¹, X. Xu^{1,2*}, A. Pasquarello^{1*}

¹Chaire de Simulation à l'Echelle Atomique (CSEA), Ecole Polytechnique Fédérale de Lausanne (EPFL), ²Beijing Computational Science Research Center

Semi-local density functionals such as PBE typically underestimate experimental band gaps by 50%. Hybrid functionals address this "band gap problem" by admixing a fraction of exact exchange to semi-local exchange. The optimal mixing parameter depends on the specific material and can be identified as the inverse dielectric constant [1]. Recently, we have shown that dielectric constants obtained using the r^2 SCAN meta-GGA functional [2] are significantly more accurate than dielectric constants obtained using PBE [3]. This can be understood through the improved treatment of electronic self-interaction within the meta-GGA framework [4,5].

In this talk, we will show that a dielectric-dependent hybrid functional based on r^2 SCAN can outperform the standard PBE based hybrid in terms of band gaps (see figure). Particularly impressive improvements are obtained for narrow-gap semiconductors such as Ge and InAs, where PBE wrongly predicts a metallic phase, but r^2 SCAN can open a gap. The hybrid functional based on r^2 SCAN also yields accurate effective masses and ionization potentials. By studying the yellow luminescence in GaN, we showcase that our new hybrid functional is an excellent choice for semiconductor applications such as defect calculations. Finally, we discuss how our new hybrid functional can be applied to molecular systems.

[1] Audrius Alkauskas, Peter Broqvist, Alfredo Pasquarello, *Physica Status Solidi (b)*, **2011**, 248(4), 775-789.

[2] James W. Furness, *et al.*, *The Journal of Physical Chemistry Letters*, **2016**, 11(19), 8208-8215.

[3] Stefan Riemelmoser, Xun Xu, Alfredo Pasquarello, (to be submitted).

[4] John P. Perdew, *et al.*, *Proceedings of the National Academy of Sciences.*, **2017**, 114(11), 2801-2806.

[5] Yubo Zhang, *et al.*, *Physical Review B*, **2020**, 102(4), 045112.

Overcoming Injectivity Challenges in Chemical Reaction Properties via Tagged Physics-Inspired Representations

J. V. Schleicher¹, Y. Calvino Alonso^{1*}, K. R. Briling^{1*}, C. Corminboeuf^{1*}

¹Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Physics-inspired representations were initially developed to predict a range of quantum chemical properties. More recently, these methods have been extended to model properties associated with chemical transformations. In this latter context, key design principles critical for creating high-performing representations of chemical reactions were identified [1]. Specifically, preserving a meaningful notion of the difference between products and reactants (*i.e.*, additivity), as well as effectively separating the description of bonding interactions (*i.e.*, preserving bagging), were found to be essential. These principles enabled accurate prediction of energy barriers across diverse datasets encompassing a wide variety of reaction types, catalysts (including metals and ligands), and catalyst conformations [2].

In this work, we highlight additional challenges that arise when addressing substrate scope. Notably, the presence of multiple chemically similar reactive sites, such as sp^2 C–H bonds, on a single substrate introduces an injectivity problem. Furthermore, the local nature of individual reactive sites, particularly those involving light atoms, can be obscured within the global molecular representation, making it difficult to accurately capture site-specific reactivity.

Here, we present strategies that leverage the inherent flexibility of physics-based representations to restore site-specific uniqueness. In particular, we demonstrate the effectiveness of the *meaningful difference framework* [1] and introduce a tagging scheme to explicitly encode the reactive site. We apply these approaches to the predictions for C–H activation reactions across systems of increasing complexity.

Figure 1: Illustration of the tagging strategy used to encode the reactive site.

[1] Puck van Gerwen, Alberto Fabrizio, Matthew D Wodrich and Clemence Corminboeuf, *Machine Learning Science and Technology*, **2022**, 3, 04005.

[2] Puck van Gerwen, Ksenia R. Briling, Yannick Calvino Alonso, Malte Franke and Clemence Corminboeuf, *Digital Discovery*, **2024**, 3, 932-943.

CurryBO: How to Obtain General Chemical Conditions via Bayesian Optimization over Curried Functions

S. P. Schmid^{1,2}, E. M. Rajaonson^{3,4*}, C. Ser^{3,4*}, M. Haddadnia^{5*}, S. Leong^{3,6*}, A. Aspuru-Guzik^{3,4*}, A. Kristiadi^{4*}, K. Jorner^{1,2*}, F. Strieth-Kalthoff^{7*}

¹Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich CH-8093, Switzerland, ²National Centre of Competence in Research (NCCR) Catalysis, Switzerland, ³Department of Chemistry, University of Toronto, ⁴Vector Institute, ⁵Department of Biological Chemistry & Molecular Pharmacology, Harvard University, ⁶School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, ⁷School of Mathematics and Natural Sciences, University of Wuppertal

General reaction conditions, i.e. reaction conditions that deliver satisfactory outcomes for multiple starting materials, are highly desirable, as they can be used and reused for multiple transformations, e.g. on one plate in high-throughput experimentation [1]. Within the development of new reactions, conditions are commonly optimized for one model substrate and the generality of these conditions is subsequently evaluated on a substrate scope [1]. However, this approach omits any consideration for conditions generality, as the optimization of reaction conditions for one substrate does not guarantee a broad substrate scope.

To optimize reaction conditions, traditional human-centric optimization workflows have been complemented by data-driven methods such as Bayesian Optimization in order to increase optimization efficiency [2]. Coupled with robotic reaction execution, autonomous reaction optimization has emerged as a key application of self-driving labs to date [3]. Even though initial studies have investigated the usage of data-driven methods for the identification of general reaction conditions, their optimization behaviour as well as limitations remain yet to be understood [4,5].

In order to address this, we present *CurryBO*, an expansion of the Bayesian Optimization framework for optimization of general reaction conditions by considering multiple substrates throughout the reaction optimization. We investigate multiple transition-metal catalyzed cross-coupling reactions and the nucleophilic addition of thiols to imines catalyzed by chiral phosphoric acids. For all reactions, we show how optimizing over multiple substrates leads to more general reaction conditions. Further, we demonstrate that *CurryBO* efficiently identifies such general reaction conditions, enabling the development of broadly applicable reactions.

[1]: C.C. Wagen, S.E. McMinn, E.E. Kwan, E.N. Jacobsen, *Nature*, 2022, 610 (7933), 680–686.

[2]: A.D. Clayton, J.A. Manson, C.J. Taylor, T.W. Chamberlain, A.B. Taylor, G. Clemens, A.R. Bourne, *Reaction Chemistry & Engineering*, 2019, 4 (9), 1545-1554.

[3]: G. Tom, S.P. Schmid, S.G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E.M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G.D. Akkoc, F. Strieth-Kalthoff, M. Seifrid, A. Aspuru-Guzik, *Chemical Reviews*, 2024, 124 (16), 9633-9732.

[4]: N.H. Angello, V. Rathore, W. Beker, A. Wołos, E.R. Jira, R. Roszak, T.C. Wu, C.M. Schroeder, A. Aspuru-Guzik, B.A. Grzybowski, M.D. Burke, *Science*, 2022, 378 (6618), 399-405.

racer^{TS}: Rapid generation of transition-state conformer ensembles via constrained distance geometry

S. P. Schmid^{1,2}, H. Seng^{1*}, T. Kläy^{1*}, K. Jorner^{1,2*}

¹Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich CH-8093, Switzerland, ²National Centre of Competence in Research (NCCR) Catalysis, Switzerland

The consideration of transition state (TS) conformer ensembles is required to accurately model a reaction, and thus plays a key role in computational catalyst design [1]. While CREST [2] and GOAT [3] are established methods for TS conformer ensemble generation, the associated computational cost remains a major bottleneck for generating large machine learning datasets for catalyst design.

To this end, we present racer^{TS} (Rapid and Accurate Conformer Ensembles with RDKit for Transition States), a method for swift TS conformer ensemble generation, based on the established ETKDG method within RDKit [4]. Focusing on usability and with RDKit as its sole dependency, racer^{TS} is easily integrable into existing computational chemistry workflows and is further able to generate conformer ensembles with 3 lines of python code. To demonstrate its suitability for TS conformer ensemble generation in high-throughput computational chemistry workflows, we create TS conformer ensembles for 20 model reactions, including transition-metal catalyzed reactions. We benchmark the created ensembles against CREST and GOAT as state-of-the-art methods, focusing on exhaustiveness, validity, accuracy in low-energy regions and speed.

We show that racer^{TS} samples the conformer space exhaustively, provides the same fraction of valid TS and is equally accurate in the low-energy regions as the state-of-the-art conformer ensemble generators. This performance is achieved with a speed-up of 4x and 400x compared to CREST and GOAT, respectively. Our results thus demonstrate that racer^{TS} is a highly efficient and accurate TS conformer ensemble generator. With its low computational cost and high usability, racer^{TS} facilitates the generation meaningful large TS datasets to advance machine learning methods in catalyst discovery.

[1] R. Laplaza, M.D. Wodrich, C. Corminboeuf, *The Journal of Physical Chemistry Letters*, 2024, 15, 7363-7370.

[2] P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett, S. Spicher, P. Steinbach, P.A. Wesółowski, F. Zeller, *The Journal of Chemical Physics*, 2024, 160, 114110.

[3]: B. de Souza, *Angewandte Chemie International Edition*, 2025, 64, e202500393.

[4]: S. Riniker, G.A. Landrum, *Journal of Chemical Information and Modeling*, 2015, 55, 2562-2574.

Efficient Atomic Feature Selection for Chemical ML Models with ECFP-Inspired FingerprintsG. Sulpizio¹, K. Jorner^{1*}¹Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich

Predictive machine learning in chemistry often grapples with limited datasets, underscoring the need for robust atomic and molecular descriptors. Although molecular descriptors are well established, identifying the optimal set of atomic descriptors for a given task remains challenging - and retraining large, state-of-the-art models such as graph neural networks (GNNs) to evaluate different combinations can be prohibitively expensive. Here, we present a lightweight, task-specific framework for atomic descriptor selection that can seamlessly integrate into larger, more accurate models. Our approach adapts the extended-connectivity fingerprint (ECFP) methodology to generate a molecular fingerprint that represents both the structure and atomic properties, then applies a linear regression to rank descriptor importance. By selecting descriptors based on their regression scores, we efficiently isolate the most predictive atomic features without repeated training of a larger model such as GNNs. This strategy not only reduces computational cost but also paves the way for faster development of high-accuracy models for molecular and reaction property prediction using atomic descriptors.

Atomic-Scale Insights into Water Arrangement and Oxygen Evolution Reaction on LaTiO₂N (100) Surface

D. Tang¹, F. Creazzo^{1*}, S. Lubert^{1*}

¹Department of chemistry, University of Zurich, CH-8057 Zurich, Switzerland

The perovskite oxynitride LaTiO₂N exhibits significant potential for visible-light-driven photocatalytic water splitting. However, the atomic-scale dynamics governing its catalyst-water interface remain poorly understood. Employing state-of-the-art density functional theory molecular dynamics (DFT-MD), the present research elucidates the complex interplay among water arrangement, surface reconstruction, and oxygen evolution reaction (OER) performance on the LaTiO₂N (100) surface¹. By considering explicit solvent effects, we reveal a pronounced hydrophilic character, with spontaneous water dissociation forming hydroxyl groups at undercoordinated Ti sites at the surface, stabilizing the latter and potentially enhancing the OER activity¹. Our simulations identify the thermodynamically stable XX/2 termination, exhibiting the lowest surface energy of 0.09 eV/Å², and demonstrate its role in fostering robust hydrogen-bond networks that facilitate proton-coupled electron transfer. Comparative analyses using PBE, PBE+U, and HSE06 functionals validate bulk properties, with PBE+U yielding a bandgap of 1.93 eV, closely aligning with experimental data. The surface Pourbaix diagram underscores hydroxylated configurations across diverse electrochemical conditions, corroborated by DFT-MD insights. Free energy calculations pinpoint the *OH to *O step as the OER bottleneck, with low overpotentials (0.11–0.94 V) evidencing superior catalytic efficiency, especially under neutral to alkaline conditions¹. The findings demonstrate the pivotal importance of accounting for an explicit solvent environment and its dynamics at a given temperature in the computational modelling of LaTiO₂N, providing a strategic foundation for developing high-efficiency, sustainable catalysts for photocatalytic water splitting.

[1] Tang, F. Creazzo, S. Lubert, To be submitted.

Bandgap and Structural Investigation of α -FASnI₃ Perovskite

M. Venkatanarayanan¹, V. Carnevali^{1*}, A. Vezzosi^{1*}, V. Slama^{1*}, U. Röthlisberger^{1*}

¹Laboratory of Computational Chemistry and Biochemistry, EPFL

Formamidinium Tin Triiodide (FASnI₃) is a promising lead-free perovskite for photovoltaics, but its limited thermal stability and lower efficiency compared to lead-based counterparts remain significant challenges. Accurate theoretical modelling of FASnI₃ is essential for understanding its structural and electronic properties. While hybrid functionals and spin-orbit coupling (SOC) are commonly employed to match experimental band gaps, they are often applied only to minimal unit cells or small supercells that lack structural realism [1, 2, 3].

Here, we investigate how supercell size, Formamidinium (FA) orientation, and the level of theory influence the predicted band gap and structure of α -FASnI₃. We find that when sufficiently large supercells (at least 6×6×6) are relaxed along all degrees of freedom, the system naturally converges toward a cubic average structure. This relaxation preserves local octahedral distortions, minimizes residual dipoles, and recovers threefold symmetry. Band gap fluctuations at 300 K also decrease with increasing system size, indicating improved structural stability. These findings emphasize that structural convergence must come before theoretical refinement, as accurate functionals cannot compensate for unphysical geometries.

[\[1\] Das, T.; Di Liberto, G.; Pacchioni, G. Density Functional Theory Estimate of Halide Perovskite Band Gap: When Spin Orbit Coupling Helps. J. Phys. Chem. C 2022, 126 \(4\), 2184–2198.](#)

[\[2\] Xie, W.; Peng, L.; Li, N. An Insight into the Effects of SnF₄ Assisting the Performance of Lead-Free Perovskite of FASnI₃: A First-Principles Calculations. ACS Omega 2021, 6 \(23\), 14938–14951.](#)

[\[3\] Shi, T.; Zhang, H.-S.; Meng, W.; Teng, Q.; Liu, M.; Yang, X.; Yan, Y.; Yip, H.-L.; Zhao, Y.-J. Effects of Organic Cations on the Defect Physics of Tin Halide Perovskites. J. Mater. Chem. A 2017, 5 \(29\), 15124–15129.](#)

Development of a phase-transferable machine learning potential for FAPbI₃

A. Vezzosi¹, E. Vasey^{1*}, V. Carnevali^{1*}, V. Slama^{1*}, U. Röthlisberger^{1*}

¹Laboratory of Computational Chemistry and Biochemistry, EPFL, Switzerland

In this contribution, I will present a fast, highly accurate and phase-transferable artificial neural network potential (ANNP) [1] for FAPbI₃, a hybrid organic-inorganic perovskite material holding enormous potential for next-generation photovoltaic devices. The force field is trained on energies and atomic forces obtained through density functional theory-based ab-initio molecular dynamics sampling at various temperatures for both the perovskite, photoactive α -phase and the non-photoactive δ -phase of FAPbI₃. [2] As stability issues represent the major challenge for the commercialization of FAPbI₃-based solar cells, the ANNP is validated to predict the δ -to- α phase transition temperature [3] as well as by comparing structural and dynamical properties with ab-initio calculations. [4] The presented MLP will enable extensive (both in simulation time and system size) and high-fidelity classical molecular dynamics simulations, providing mechanistic insights into critical factors influencing the relative α/δ phase stability, and thus represent a fundamental tool to help and guide the design of strategies for producing low temperature and pure α -phase FAPbI₃ materials, such as compositional engineering and the introduction of suitably chosen additives. [5]

- [1] Zeng J., et al. J. Chem. Phys. 7 August 2023; 159 (5): 054801
- [2] Vona C., et al., J. Phys. Chem. C2025, 129, 3040–3053
- [3] Braeckvelt T., et al., Chemistry of Materials 2022 34 (19), 8561-8576
- [4] Carnevali V., et al. <https://doi.org/10.21203/rs.3.rs-5730287/v1>
- [5] G. AlSabeh, et al. Angew. Chem. Int. Ed. 2025, 64, e202417432.

Quantum Energy Level Splittings from Path IntegralY. Wang¹, L. Raso^{2*}, J. O. Richardson^{1*}¹Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland, ²Department of Chemistry, University of Milan, Via Golgi 19, 20133 Milan, Italy

Understanding tunneling splittings and vibrational-rotational level structures in molecules is essential for deciphering nuclear quantum effects and fluxional behavior in systems with highly delocalized hydrogen atoms. Among available theoretical methods, path-integral molecular dynamics (PIMD) offers a powerful and rigorously quantum-mechanical framework to compute such quantities, capturing nuclear delocalization and zero-point motion. However, conventional approaches typically rely on thermodynamic integration (TI) across a discretized λ -path connecting symmetry-related wells. This procedure becomes increasingly costly and operationally complex for multi-well systems such as CH₅, requiring multiple long-time simulations and careful control of discretization errors.

Here we present a new methodology that integrates PIMD with Enveloping Distribution Sampling (EDS) to bypass the need for TI. Specifically, we construct a smooth, barrier-free enveloping potential that connects all relevant end states by combining a set of λ -dependent bias-corrected bridging potentials within the EDS formalism. Optimal λ spacing and bias values are efficiently determined using ring-polymer instanton calculations. To enhance sampling efficiency of the zero mode naturally present in the ring polymer, we introduce ring-polymer permutation moves coupled with Metropolis Monte Carlo, enabling efficient configuration sampling and significantly improving convergence. We demonstrate the effectiveness of this scheme on one-dimensional double-well models across a wide range of barrier heights, showing accurate and efficient reproduction of tunneling splittings. Ongoing work focuses on extending the method to realistic molecular systems, with CH₅ as a target application.

Graph Theory-Based Analysis of Connectivity and Chirality in Silver and Gold Nanoclusters

E. Zerbato¹, K. Ahmed Ethmane^{1*}, C. Besnard^{2*}, A. F. Perez Mellor^{1*}, T. Burgi^{1*}

¹Department of Physical Chemistry, University of Geneva, 30 Quai Ernest-Ansermet, 1211 Geneva, Switzerland, ²Laboratory of Crystallography, 24 Quai E. Ansermet, CH-1211 Geneva 4, Switzerland

Graph theory offers a flexible framework for analyzing complex systems and has found applications in fields ranging from biology and computer science to chemistry and materials science. By representing entities as nodes and their interactions as edges, it becomes possible to explore underlying patterns and connectivity in a systematic way. These methods have been applied to a diverse range of applications from modeling the evolution of social networks to solving urban planning problems, such as optimizing the placement of metro stops in response to demographic changes.

In this study, we apply graph-theoretical methods to investigate connectivity patterns in a series of silver and gold nanoclusters obtained from the Cambridge Crystallographic Data Centre (CCDC). Each nanocluster is modeled as a graph, where atoms are represented as nodes and chemical bonds as edges [1]. Within this framework, we compute various graph-based properties, including node degree and canonical labeling. Additionally, we introduce a novel graph-based chirality descriptor that enables quantitative analysis of chirality within the nanoclusters. This new descriptor also allows us to localize chiral features to specific regions of the cluster—namely, the metallic core, the metal–ligand interface, and the outer ligands.

We present a systematic comparison of such descriptors for silver and gold nanoclusters, highlighting key similarities and differences in their structural connectivity and chiral properties. These insights contribute to a deeper understanding of the topological organization of metal nanoclusters and their chemical behavior.

[1] M. Cha, J. Ma, J-K. Kim, E. S. T. Emre, N. A. Kotov, *Chirality*, 2024, 36(6), e23678.

Ab Initio MASH with Non-Aufbau ROKS method: Robust Excited States and Analytical CouplingsR. Zhou¹, J. O. Richardson^{1*}¹Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland

Efficiently and accurately determining electronic structure states remains a significant bottleneck for performing nonadiabatic molecular dynamics simulations on larger systems. This work presents recent advancements in the *ab initio* Mapping Approach to Surface Hopping (MASH) methodology. We utilize the single-reference non-aufbau method based on the Restricted Open-shell Kohn-Sham (ROKS) framework for excited-state calculations, addressing two key challenges. First, we demonstrate the application of a time-evolution algorithm to effectively manage and mitigate problematic near-degeneracies or discontinuities encountered during excited-state calculations. Second, we report an efficient implementation for the analytical calculation of nonadiabatic coupling vectors between different electronic states. We present proof-of-principle examples that validate these developments and highlight the potential of our MASH approach to be extended to complex, realistic chemical systems, thereby broadening the scope of nonadiabatic dynamics simulations.

Efficient Hit Expansion and Hit-to-Lead Optimization in Ultra-Large Combinatorial LibrariesC. Loeffeld¹, M. von Korff^{1*}, T. Sander^{1*}¹Alipheron AG, Dornach, Switzerland

In drug discovery, hit expansion, hit-to-lead optimization and substructure searches in ultra-large combinatorial libraries are increasingly vital for identifying potential candidates for further development. Discovered analogues should exhibit significant chemical dissimilarity from the query molecule and can thus serve as starting points for new medicinal chemistry programs aimed at addressing issues related to ADMET or intellectual property (IP). Ultra Large Combinatorial Libraries (ULCLs) currently ranging into the trillions of readily synthesizable compounds present a viable source for successful hit expansion and lead optimization campaigns. However, efficiently harvesting this information using traditional tools is costly and time-consuming. Here, we present two novel and highly efficient, complimentary pharmacophore-based search methods, Pharos3D and Hyperspace Search, optimized for hit expansion and substructure search, respectively. Both methods are applicable to a steadily increasing number of prominent Ultra Large Combinatorial Libraries on the versatile and widely used open-source cheminformatics platform, DataWarrior.

Focused Sub-Libraries from Ultra-Large Combinatorial Spaces with the Hyperspace ExplorerM. v. Korff¹, C. Loeffeld^{1*}, T. Sander^{1*}¹Alipheron AG, Schlossweg 63, 4143 Dornach, Switzerland

In drug discovery, the ability to search ultra-large combinatorial libraries (ULCLs) becomes increasingly essential for hit expansion. We present a new software tool—the **Hyperspace Explorer**—designed to systematically search combinatorial spaces for derivatives of a given molecule. Developed using the code base of the **OpenChemLib**, an open-source cheminformatics library, this tool enables highly flexible substructure searches through the use of wildcard-based structure encoding. By replacing defined structural features of the query molecule with wildcards—guided by its pharmacophoric substructures—the Hyperspace Explorer reduces the molecule into a hierarchy of topological pharmacophore patterns. These reduced patterns are then used to explore ULCLs for structurally and pharmacologically relevant derivatives. For the first time, we will present results of chemical space expansion around recently approved drugs, including **Atrasentan**, **Gepotidacin**, and **Vimseltinib**. Our method yields **focused sub-libraries** containing compounds that retain key pharmacophore features of the original molecule while omitting or replacing less relevant moieties. These focused libraries offer a promising starting point for further refinement and optimization in hit expansion campaigns.

Machine Learning-Guided Discovery of Electron Donor-Acceptor Complex Reactions

P. L. Türtcher¹, A. F. Zahrt^{1*}

¹University of Pennsylvania, 231 S. 34 Street, Philadelphia, PA 19104, United States of America

Photocatalysts are capable of generating radical intermediates in a controlled manner and under mild conditions.[1] The subsequent challenge is the vast combinatorial space of possibilities of reagents - such as photoredox catalysts, transition metal catalysts, ligands, and additives - as this complexity creates a large and diffuse chemical space. Until now, the exploration typically relies heavily on existing domain expertise. A simpler subclass of photocatalyzed reactions is based on the formation of electron donor-acceptor (EDA) complexes,[2] where a reaction only occurs upon formation of a visible light absorbing complex. Here, the parameter space of reactants can, in a first step, be reduced to donor and acceptor molecules. Similar approaches have already used machine learning (ML) to uncover new reactivity for completely out-of-sample molecules in electrochemical reactions.[3]

In this work, we apply a generally applicable encoding for donor and acceptor molecules and, based on this, derive an intuitive reaction descriptor for EDA reactions. Our approach prioritizes generality, enabling the representation of currently challenging molecules such as transition metal catalysts. Additionally, we ensure sufficient conformational sampling to capture the flexibility of molecules. With a corresponding curated dataset, we investigate how ML models can predict novel, feasible reactions that share similar bonding patterns but involve distinct reactants. To discover new reactivity, we compile a large dataset of commercially available compounds and screen potential reaction pairs based on a few key ML-predicted quantum mechanical properties, such as the adiabatic reduction-oxidation potential. Promising donor-acceptor pairs filtered out through these predictions are then compared to the curated literature data to detect the most similar reactions based on our reaction descriptor. This two-stage approach enables a rapid yet approximate initial evaluation, followed by a more fine-tuned and accurate refinement leading to promising reaction candidates.

[1] C. McAtee, E. J. McClain, C. R. J. Stephenson, *Trends in Chemistry* **2019**, *1* (1), 111–125.

[2] E. M. Crisenza, D. Mazzarella, P. Melchiorre, *J. Am. Chem. Soc.* **2020**, *142* (12), 5461–5476.

[3] F. Zahrt, Y. Mo, K. Y. Nandiwale, R. Shprints, E. Heid, K. F. Jensen, *J. Am. Chem. Soc.* **2022**, *144* (49), 22599–22610.