

# CHIMIA

CHIMIA 2022, Volume 76  
ISSN 0009-4293  
www.chimia.ch  
Supplementa to Issue 7-8/2022



**SCS**  
Swiss Chemical  
Society

SCS Fall Meeting 2022  
Lecture, Short Talk and Poster Abstracts

## Session of Computational Chemistry

September 8, 2022  
University of Zurich, Irchel Campus  
<https://fm22.scg.ch>

Swiss Chemical Society  
Haus der Akademien  
Postfach  
3001 Bern  
Switzerland  
[info@scg.ch](mailto:info@scg.ch)  
[www.scg.ch](http://www.scg.ch)

**Fluctuations Beyond Chemical Imagination in Liquid Water**A. Hassanali<sup>1</sup>

<sup>1</sup>The International Center for Theoretical Physics (ICTP), Condensed Matter and Statistical Physics section, Trieste, Italy

Water is considered the solvent for sustaining life as we know it on planet Earth. Despite long study, very basic notions in physical chemistry such as hydrophobicity, continue to be the source of numerous open questions. In this talk, I will give an overview of how the use of atomistic modeling with various degrees of complexity provides a powerful framework that allows for developing a detailed microscopic understanding of the wide variety of fluctuations that exist in liquid water. These fluctuations occur on different length and timescales and involve the complex coupling of density, topology and dielectrics. Within this context, I will discuss problems ranging from the complex shapes of empty spaces in liquid water and their role in solvation, to how oil and water can counterintuitively mix and finally, why soap bubbles might be used for artificial photosynthesis, to name a few. Building a microscopic understanding of the collective nature of the fluctuations in these systems requires us to move beyond our chemical imagination. In this spirit, I will present how data-science and chemical intuition can be combined to provide a more nuanced way to study the chemical physics of aqueous systems.

**Enzeptional: enzyme optimization via a generative language modeling-based evolutionary algorithm**

Y. Nana Teukam<sup>1,2</sup>, M. Manica<sup>1</sup>, F. Grisoni<sup>2</sup>, T. Laino<sup>1\*</sup>

<sup>1</sup>IBM Research Europe, Zürich, <sup>2</sup>Eindhoven University of Technology

Enzymes are molecular machines optimized by nature to allow otherwise impossible chemical processes to occur. Besides the increased reaction rates, they present remarkable characteristics to enable more sustainable reactions: mild conditions, less toxic solvents, and reduced waste. Billion years of evolution have made enzymes extremely efficient. However wide adoption in industrial processes requires faster design using in-silico methodologies, a daunting task far from being solved. The majority of methods operate by introducing mutations in an existing amino acid (AA) sequence using a variety of assumptions and strategies to introduce variants in the original sequence. More recently, machine learning and deep generative networks have gained popularity in the field of protein engineering by leveraging prior knowledge on protein binders, their physicochemical properties, or the 3D structure. Here, we cast the problem of enzyme optimization as an evolutionary algorithm where mutations are modeled via generative language modeling. Relying on pretrained language models trained on AA sequences, we apply transfer learning and train a scoring model on a dataset of biocatalysed chemical reactions that is used to drive the optimization process. Our methodology allows designing enzymes with higher biocatalytic activity, emulating the evolutionary process occurring in nature by sampling optimal sequences modeling the underlying proteomic language.

## Predicting chemical toxicity through machine learning

C. Schür<sup>1</sup>, L. Gasser<sup>2</sup>, J. Wu<sup>3</sup>, F. Perez-Cruz<sup>2</sup>, K. Schirmer<sup>1\*</sup>, M. Baity-Jesi<sup>3\*</sup>

<sup>1</sup>Department of Environmental Toxicology, Eawag, Swiss Federal Institute of Aquatic Science and Technology, <sup>2</sup>Swiss Data Science Center, <sup>3</sup>Department of Systems Analysis, Integrated Assessment and Modelling, Eawag, Swiss Federal Institute of Aquatic Science and Technology

The global landscape of registered chemicals on the market is growing at a fast pace. Meanwhile, the regulation of compounds requires extensive animal testing to ensure their safety for both the public and the environment. Beyond its ethical implications, animal testing is labor-, cost-, and time-intensive. Here, new approach methods can be a remedy by reducing, if not ultimately replacing, conventional animal tests. An increase in computational power and accessibility of machine learning methods enable the use of *in silico* methods for ecotoxicological questions such as the hazard assessment of chemicals. Our research is aimed at accurately predicting the toxicity of chemicals to a wide range of taxa beyond model organisms and to explore its potential and limitations. For this, we assembled an extensive dataset of ecotoxicity data across three trophic levels (fish, crustaceans, and algae) with over 1000 species. This core dataset was enhanced by adding species-specific and phylogenetic information and by comparing several ways to represent chemical structure and properties (molecular fingerprints, mol2vec). We explore the importance of adding certain feature groups and put their individual relevance in the context of biology to better understand the drivers of toxicity as well as what potential supplementary experiments could help reduce cost and harm.

## The Spectrum of Approximated Hamiltonian Matrices Representations (SPA<sup>H</sup>M)

K. R. Briling<sup>1</sup>, Y. Calvino Alonso<sup>1</sup>, A. Fabrizio<sup>1,2</sup>, C. Corminboeuf<sup>1,2\*</sup>

<sup>1</sup>Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland, <sup>2</sup>National Center for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland

Recently we proposed a class of molecular representations for machine learning (ML) — Spectrum of approximated Hamiltonian matrices (SPA<sup>H</sup>M) [1] based on lightweight one-electron Hamiltonians, widely used as an SCF initial guess in computational chemistry. SPA<sup>H</sup>M naturally contains the information about the nuclear charges and positions as well as obeys the symmetries required for an effective ML representation.

The first variant is a compact global representation consisting of occupied-orbital eigenvalues of the approximated Hamiltonian. Owing to a seamless generalization to open-shell systems, SPA<sup>H</sup>M performs well on datasets characterized by a wide variation of charge and spin, for which the traditional structure-based representations commonly fail.

Using our first eigenvalue-variant as a starting point, we then focus on designing local and transferable representations, through exploiting the occupied eigenvectors and associated density matrix. Bridging the advantages of both the SOAP [2] and SLATM [3], we construct similarity measures from density-based fingerprints expanded into atoms and bond contributions. The atomic (a) or bond (b) density overlap kernel is defined as a dot product of power spectra, which themselves are used as representation vectors [SPA<sup>H</sup>M(a) and SPA<sup>H</sup>M(b)] of an atomic environment.

Preliminary results indicate that the SPA<sup>H</sup>M(a) is already competing with SLATM [3] for predicting atomic properties. Current efforts are placed into leveraging SPA<sup>H</sup>M(a,b) to describe charged species, open-shell systems and properties derived from them.

[1] Alberto Fabrizio, Ksenia R. Briling, Clemence Corminboeuf, *Digital Discovery*, **2022**, Advance Article (doi:10.1039/D1DD00050K).

[2] Albert P. Bartók, Risi Kondor, Gábor Csányi, *Physical Review B*, **2013**, 87, 184115.

[3] Bing Huang, O. Anatole von Lilienfeld, *Nature Chemistry*, **2020**, 12, 945–951.

## Pathfinder - A Route Planner for Chemical Reaction Networks

P. L. Türtcher<sup>1</sup>, M. Reiher<sup>1\*</sup>

<sup>1</sup>ETH Zürich, Laboratory of Physical Chemistry, Vladimir-Prelog-Weg 2, CH-8093 Zürich, Switzerland

Automated reaction network exploration enables the discovery of vast chemical space[1, 2, 3], consisting of many compounds highly interconnected via reactions. As a consequence, tracking what compounds are formed via which reaction sequence is a non-trivial task. While kinetic modeling provides the concentration or population development of a compound over time, a specific sequence of reactions resulting, for instance, in the most probable path towards a compound is not directly accessible. Furthermore, in a large and dense chemical reaction network containing a wide range of reaction barriers, kinetic modeling on time scales of interest might be too costly for the evaluation of populations in a network.[4]

This work aims to represent any chemical reaction network as a graph and analyze this graph to find paths, meaning sequences of reactions, connecting any two compounds in the network. An identified path should provide information on its feasibility of the reactions along the path with regard to availability of reagents and kinetic barriers. This yields a ranking of accessibility for all compounds in the network which can be employed to guide further explorations by focusing on compounds with high accessibility according to *Pathfinder*. Such a guiding approach tames the combinatorial explosion of an exploration and gives the explorer more control over a running exploration. In addition, for economic kinetic modeling, the network can be pruned based on the ranking of compounds, again starting by only including compounds with high accessibility.

We employ such a graph-based approach to build and analyze a chemical reaction network of the disproportionation of iodine with water and improve explorations based on the results within our SCINE framework.[5]

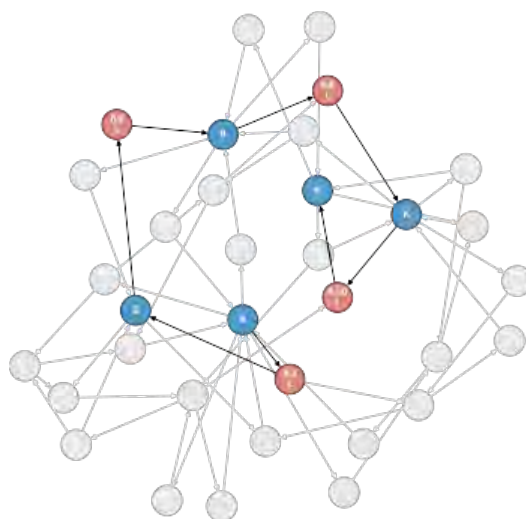


Figure 1: Schematic representation of a path from compound **A** to compound **H** in an arbitrary chemical reaction network.

[1] G. N. Simm, A. C. Vaucher, M. Reiher, *J. Phys. Chem. A*, **2019**, 123 (2), 385–399.

[2] J. P. Unsleber, M. Reiher, *Annu. Rev. Phys. Chem.*, **2020**, 71 (1), 121–142.

[3] J. P. Unsleber, S. A. Grimm, M. Reiher, *arXiv:2202.13011 [physics.chem-ph]*, **2022**.

[4] J. Proppe, M. Reiher, *J. Chem. Theory Comput.*, **2019**, 15 (1), 357–370.

[5] SCINE - Software for Chemical Interaction Networks. <https://scine.ethz.ch/> (accessed February 24, 2022).

## Mechanistic investigations at the alumina/water interface

C. Michel<sup>1</sup>

<sup>1</sup>Laboratoire de Chimie, Ens de Lyon, CNRS UMR 5182, Lyon 1, 69342 Lyon, France

$\gamma$ -Al<sub>2</sub>O<sub>3</sub> has remarkable properties as a support, but it is very sensitive to water, either as a liquid[1] or even as steam[2]. These structural changes can be retarded by the presence of polyols.[3] The mechanism of action of these additives is believed to lie in their chemisorption on the surface that would make  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> water-resistant[3]. However, little is known on which exposed facet(s) and what site(s) need to be particularly targeted for protection. The optimization of the structure of inhibitors is therefore a challenging task since the exact mechanism of the decomposition of  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> in liquid water remains unknown. We combine here experiments and theory to provide the atomistic mechanism for the early-stage decomposition of  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> in liquid water,[4] but also for the adsorption/desorption mechanism of (poly)-alcohols at the  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> / water interface.[5]

Alumina samples containing particles with different shapes were exposed to a hydrothermal treatment in presence of various concentrations of polyol (sorbitol and xylitol, Figure 1b and 1c). We determined the minimum surface coverage of polyol at which the decomposition of alumina gets inhibited, a quantity referred to as *inhibiting coverage*. Since this inhibiting coverage strongly correlates with the fractional area of the (110) facet of the alumina sample and only with this area, we concluded that the decomposition is initiated at the (110) facet (Figure 1d). Then, performing *ab initio* metadynamics simulations using cp2k, we probed the reactivity of this Al<sub>2</sub>O<sub>3</sub>(110)/water interface and identified specific aluminum tetrahedral centers that are particularly reactive with water. We showed that interfacial water molecules are involved in the mechanism, both as reactants for the hydration of aluminum and as intermediates for the proton reshuffling required by the decomposition mechanism. We managed to determine the barriers (all lower than 80kJ.mol<sup>-1</sup>) for the 6 successive addition/elimination steps in a single simulation and the resulting final structure is shown in Figure 1e. Furthermore, the effect of xylitol was rationalized: its adsorption locally rendered the surface more hydrophobic and pushes water molecules away from the Al sites identified as water-sensitive.[4] Then, we investigated in details the adsorption/desorption mechanism of (poly)-alcohols at the  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> / water interface and found that liquid water induces a barrier of at least 20kJ.mol<sup>-1</sup> due to the substitution of chemisorbed water and the solvent reorganization.[5] Beyond its impact on the stability of this important support, the present achievement constitutes an unprecedented milestone in the understanding of solid/liquid interface transformation in term of methodology.

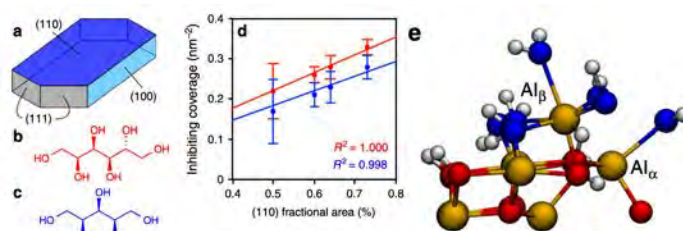


Figure 1. **a.** Typical  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> platelet; Inhibitors: **b.** sorbitol **c.** xylitol; **d.** Correlation between the inhibiting coverage and the fraction of exposed (110) ; **e.** Side view of the last intermediate obtained from the hydration of tetrahedral Al <sub>$\beta$</sub>  exposed on (110) during an *ab initio* metadynamics of the  $\gamma$ -Al<sub>2</sub>O<sub>3</sub>(110)/water interface. O from  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> in red, O from water in blue, H in white, Al in yellow.

[1] Ravenelle, R. M., Copeland, J. R., Kim, W. -G., Crittenden, J. C. and Sievers, C. *ACS Catal.* 1, 552 (2011). [1] Sakwa-Novak, M. A. and Jones, C. W. *ACS Appl. Mater. Interfaces* 6, 9245 (2014). [3] Ravenelle, R. M., Copeland, J. R., Van Pelt, A. H., Crittenden, J. C. and Sievers, C. *Catal.* 55, 162 (2012). [4] Réocreux, R., Girel, É., Clabaut, P., Tuel, A., Besson, M., Chaumonnot, A., Cabiac, A., Sautet, P. and Michel, C. *Comm.* 10, 3139 (2019). [5] Clabaut, P., Rey, J., Réocreux, R., Steinmann, S. N., and Michel, C., *ChemRxiv*, 10.33774/chemrxiv-2021-n77jn ; (2021). [6] Réocreux, R., Jiang, T., Iannuzzi, M., Michel, C. and Sautet, P. *ACS App. Nano Mat.*, 1, 191, (2018)

**Digital Chemistry at Syngenta: from academic labs to industrial applications**A. R. Finkelmann<sup>1</sup><sup>1</sup>Syngenta Crop Protection, Stein, Switzerland - arndt.finkelmann@syngenta.com

Modern agrochemicals must strike the right balance across a large panel of target properties from biological efficacy, environmental impact, resistance management, and cost of goods. This is arguably one of the most complex optimization tasks in the chemical industry. Recent breakthroughs in inverse design and generative chemistry enable to rethink this optimization approach.<sup>1,2</sup> Successful adoption of inverse design as research strategy, requires high quality data to build accurate models for relevant target properties. Most importantly, compounds need to be designed that can be readily synthesized. To address these challenges, Syngenta Crop Protection Research Chemistry has initiated an ambitious program to overhaul the whole software infrastructure that supports chemical synthesis from idea to physical sample.



In this presentation we will describe the main concepts and philosophy that went into the design of the platform and how it enables to integrate recent cutting-edge technology in a production environment that will ultimately serve hundreds of chemists worldwide. We will highlight the underlying modeling of chemical information and incorporation of large-scale reaction data for reaction prediction and mapping of synthesis targets and routes against the network of known organic reactions. Several challenges that are subject to current research will be touched.

[1] Vanhaelen, Q.; Lin, Y.-C.; Zhavoronkov, A. The Advent of Generative Chemistry. *ACS Med. Chem. Lett.* **2020**, *11* (8), 1496–1505. <https://doi.org/10.1021/acsmchemlett.0c00088>

[2] Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* (80). **2018**, *361* (6400), 360–365. <https://doi.org/10.1126/science.aat2663>

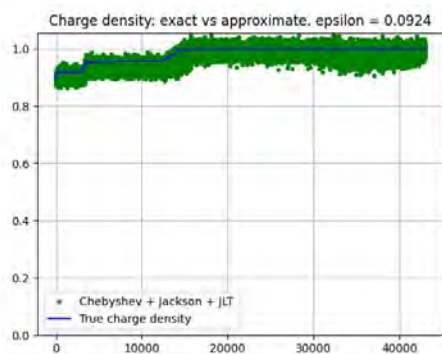
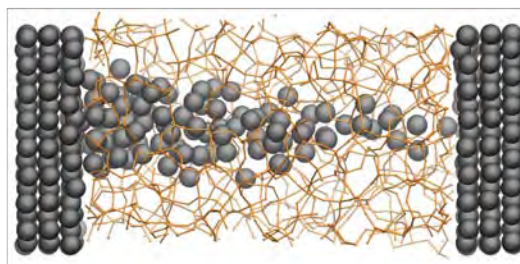


## Random projections in computational quantum mechanics

A. Sobczyk<sup>1,2</sup>, T. Laino<sup>1</sup>, M. Luisier<sup>2</sup>

<sup>1</sup>IBM Research Europe, <sup>2</sup>ETH Zürich

Computing the electronic properties of solid-states at the quantum mechanical level can be time-consuming, especially for realistic systems, e.g. semiconducting nanowires or van der Waals heterostructures, as large matrices must be diagonalized. Different numerical methods have been developed over the past decades to speed up such computations. Recently, techniques using randomization have been proposed to achieve this goal, typically by projecting the matrix of interest on a randomly chosen low-dimensional subspace. By doing so, the problem size is reduced and consequently, the computational overhead as well. In this presentation, we revisit such randomization methods and discuss their convergence and approximation guarantees. Improvements relying on recent advances in the random projections literature are also suggested. In addition to the theoretical discussion, we present numerical experiments for a so-called memristor structure whose charge density is approximated with the proposed technique. We show convergence and approximation errors and discuss the general applicability and limitations of randomization methods.

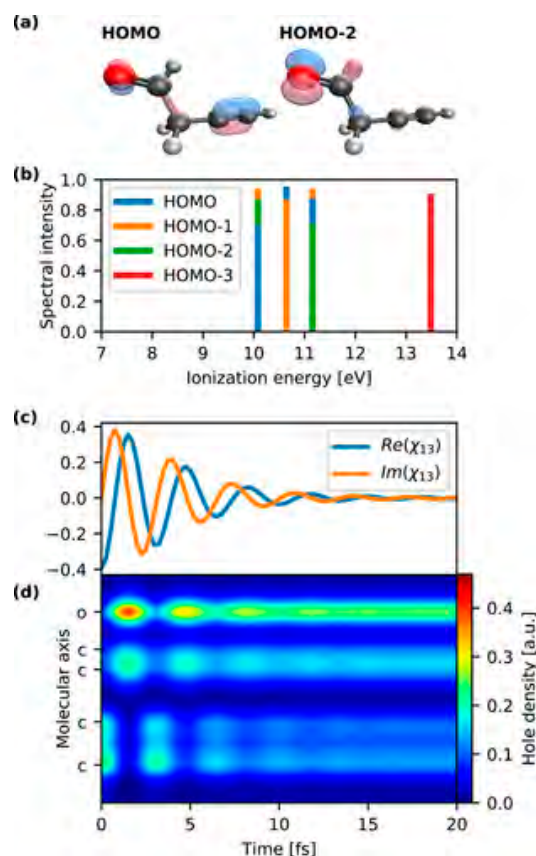


## Search for long-lasting electronic coherence using on-the-fly ab initio semiclassical dynamics

A. Scheidegger<sup>1</sup>, J. Vaníček<sup>1\*</sup>, N. Golubev<sup>1\*</sup>

<sup>1</sup>Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques

Using a combination of high-level ab initio electronic structure methods with efficient on-the-fly semiclassical evaluation of nuclear dynamics, we performed a massive scan of small polyatomic molecules searching for a long-lasting oscillatory dynamics of the electron density triggered by the outer-valence ionization. We observed that in most of the studied molecules, either the sudden removal of an electron from the system does not lead to the appearance of the electronic coherence or the created coherences become damped by the nuclear rearrangement on a time scale of a few femtoseconds. However, we report several so far unexplored molecules with the electronic coherences lasting up to 10 fs, which can be good candidates for experimental studies. In addition, we present the full-dimensional simulations of the electronic coherences coupled to nuclear motion in several molecules which were studied previously only in the fixed nuclei approximation.



**FIG. 1.** Ionization spectrum and the coupled electron-nuclear dynamics triggered by the ionization out of the HOMO of the but-3-ynal molecule. (a) Molecular orbitals involved in the hole-mixing. (b) First four computed cationic states. (c) Time evolution of the electronic coherence between the first and the third cationic states created after removal of the HOMO electron. Dynamics was performed with the semiclassical on-the-fly thawed Gaussian approximation. (d) Time evolution of the hole density along the molecular axis. The charge initially localized in the HOMO orbital migrates back and forth between the alkyne and aldehyde moieties of the molecule before being trapped by the nuclear motion.

## Recent Advances in the Development of Accurate Forcefields for Modelling Phase Stabilities in Hybrid Halide Perovskites

M. Dankl<sup>1</sup>, U. Röthlisberger<sup>1\*</sup>

<sup>1</sup>École Polytechnique Fédérale de Lausanne, Institute of Chemical Sciences and Engineering, Laboratory of Computational Chemistry and Biochemistry, Lausanne, Switzerland

Organic inorganic hybrid halide perovskites (PSCs) have emerged as a promising class of materials for solar cells. Thanks to their remarkable optical properties such as high absorption coefficient, tunable bandgap, high charge carrier mobility and low exciton binding energy, PSCs facilitate high power conversion efficiencies (PCE). Indeed, the PCEs of PSCs have rapidly risen since 2009 from 3.8% to a certified 25.6% [1] in 2021. Nevertheless, the PSCs have not yet conquered the market due to their limited long-term stability of the perovskite phase. Simulation assisted methods which could reliably predict relative finite temperature stabilities of the competing phases (e.g.  $\delta$ -phase), as well as allow the treatment of sufficiently large samples (and time scales) would give important guidance in the design of long-term stable perovskite systems for PSCs. While full first-principle methods (e.g. density functional theory) are limited by the simulated system size and feasible timescale, classical molecular dynamics simulation overcomes these limitations. Not surprisingly considerable research efforts have been directed at the development of accurate forcefields (FFs) for these systems. One of the first FFs for perovskites was developed by Mattoni et al. [2,3] for methylammonium-lead-iodide (MAPbI<sub>3</sub>) and later adapted in different flavors for describing the interaction of MAPbI<sub>3</sub> with water [4] and capturing MAPbBr<sub>3</sub> as well as mixed halide MAPbI<sub>3-x</sub>Br<sub>x</sub> compounds [5]. Recently, a polarizable FF has been introduced [6] that is able to describe the  $\delta$ -phase of CsPbI<sub>3</sub>, however still not yet its transition to the (photovoltaically active) perovskite phase. Clearly, the simultaneous description of non-perovskite and perovskite phases along with their phase transition(s) is still nowadays a challenge for which hardly any FFs have been developed so far. This talk will present an overview and recent advances of the developments of such forcefields in that field.

[1] J. Jeong, M. Kim, J. Seo, H. Lu, P. Ahlawat et al., *Nature*. **2021**, 592, 7854, 381-385.

[2] A. Mattoni, A. Filippetti, M. I. Saba, P. Delugas, *The Journal of Physical Chemistry C*, **2015**, 119, 17421-17428.

[3] A. Mattoni, A. Filippetti, C. Caddeo, *Journal of Physics: Condensed Matter*, **2017**, 29, 043001.

[4] C. Caddeo, M. I. Saba, S. Meloni, A. Filippetti, A. Mattoni, *ACS Nano*, **2017**, 11, 9183-9190.

[5] T. Hata, G. Giorgi, K. Yamashita, C. Caddeo, A. Mattoni, *The Journal of Physical Chemistry C*, **2017**, 121, 3724-3733.

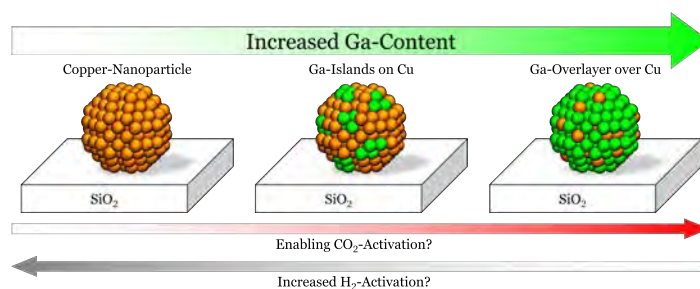
[6] P. V. G. M. Rathnayake, S. Bernardi, A. Widmer-Cooper, *The Journal of Chemical Physics*, **2020**, 152, 024117.

## Copper-Based Catalysts for CO<sub>x</sub> Hydrogenation To Methanol - Metadynamics Bridges the Pressure and Material Gap

A. Müller<sup>1</sup>, J. L. Alfke<sup>1,3</sup>, A. Comas-Vives<sup>2,4</sup>, O. Safonova<sup>3</sup>, C. Copéret<sup>1\*</sup>

<sup>1</sup>Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zurich, Switzerland, <sup>2</sup>Institute of Materials Chemistry, TU Wien, 1060 Vienna, Austria, <sup>3</sup>Paul Scherrer Institut, 5232 Villigen, Switzerland, <sup>4</sup>Departament de Química, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Catalonia, Spain

Among multiple alternatives, Cu has emerged as the most promising candidate for the hydrogenation of CO<sub>x</sub> (CO<sub>2</sub>, CO, etc.) to methanol, mainly due to the absence of the Sabatier reaction (thermodynamically favoured overreduction to methane). But pure Cu-catalysts show low activity and selectivity, and this can only be improved by using Lewis acidic supports (Al<sub>2</sub>O<sub>3</sub>, ZrO<sub>2</sub>), or by adding promoters such as Ga or Zn. Especially the latter are of high interest, since these systems consistently outperform the catalysts with an irreducible support (higher activity and selectivity). [1] But understanding the active structure of these systems proved to be extremely challenging: The gas atmosphere strongly interacts with the catalyst surface, forming a dynamic equilibrium, which is impossible to understand using classical calculations, or *state-of-the-art in-situ* spectroscopy. [2] This requires us to look beyond static calculations with slab models and understand the dynamics of the system on a microscopic, fundamental level. Herein, we used *ab initio* molecular dynamics (AIMD) in combination with Metadynamics (MTD) [3] to explore structurally distinct configuration of CuGa nanoparticles (NPs) and evaluate their free energy. With this approach, we could rationalise the structure of these NPs with different concentrations of Ga (~10% to ~50 %), both under oxidising conditions (CO<sub>2</sub> hydrogenation or air) and reducing conditions (CO hydrogenation or H<sub>2</sub>). For low Ga-loadings in vacuum, Ga-islands are formed on the particle surface, which grow when the Ga-content increases until a full overlayer is obtained. If these particle are exposed to a more oxidising atmosphere, the Ga-islands/overlayers are oxidised to a suboxide, partially or fully covering the Cu NPs.



These findings allowed us to re-examine previously measured XAS data of CuGa NPs with different loadings of Ga, which showed some irregularities with regards to the EXAFS coordination number, especially when considering the observed particle size. Using the data obtained by MTD, we could construct a new model with a partial core-shell structure, which not only resolves the conflict from the XAS data, but also rationalises why Ga acts as a promoter for low concentrations while becoming a poison at higher concentrations. This shows that by combining *in-situ* XAS and MTD, the limitation of pure experimental setups as well as classical slab calculations can be overcome and the catalytic systems can be understood on a more fundamental level.

[1] E. Lam, et al., *J. Catal.*, **2021**, 394, 266-272.

[2] A. Müller, et al., *submitted manuscript*, **2022**.

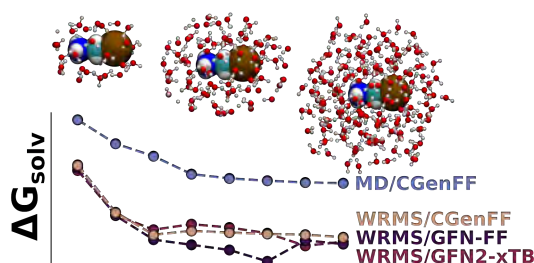
[3] A. Barducci, et al., *WIREs Comput Mol Sci*, **2011**, 1, 826-843.

## Solvation Free Energies in Subsystem Density Functional Theory

M. Bensberg<sup>1</sup>, P. L. Türtcher<sup>1</sup>, J. P. Unsleber<sup>1</sup>, M. Reiher<sup>1\*</sup>, J. Neugebauer<sup>2\*</sup>

<sup>1</sup>ETH Zürich, Laboratorium für Physikalische Chemie, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland, <sup>2</sup>Theoretische Organische Chemie, Organisch-Chemisches Institut, Center for Multiscale Theory and Computation, Westfälische Wilhelms-Universität Münster, Corrensstraße 36, 48149 Münster, Germany

The accurate modeling of solvation effects on chemical reactions in solution is key for reliable prediction of reaction mechanisms and molecular properties in general. However, the steep increase in computational cost with the number of molecules in quantum chemical calculations limits the number of explicit solvent molecules that can be included significantly. In this work [1], we propose a micro-solvation ansatz for the quantum mechanical description of solvent-solvent and solvent-solute interactions based on subsystem density functional theory (sDFT) and the conductor-like polarizable continuum model.



Convergence of the solvation free energy with the number of solvent molecules.

This new hybrid model is able to include more than one hundred solvent molecules at moderate computational cost, including the sampling of the solvent configurations. Furthermore, the sDFT-based description of the solvent-solute interaction naturally decomposes the electronic structure in terms of solute and solvent, which we use to decompose the interaction energy and directly calculate the free energy of solvation. We demonstrate the accuracy of our hybrid model by investigating the reaction energies and barriers of the Menshutkin reaction of chloromethane with ammonia as well as the intramolecular Claisen rearrangement of p-tolyl ether and compare different sampling approaches for the solvent configurations.

[1] Moritz Bensberg, Paul L. Türtcher, Jan P. Unsleber, Markus Reiher, Johannes Neugebauer, *J. Chem. Theory Comput.*, **2022** 18, 723–740

## Regression Transformer: Concurrent sequence regression and generation for molecular language modeling

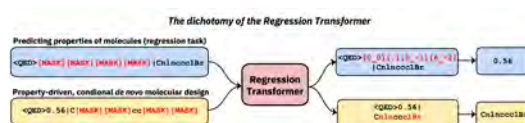
J. Born<sup>1,2</sup>, M. Manica<sup>1\*</sup>

<sup>1</sup>IBM Research Europe, <sup>2</sup>ETH Zurich

Despite significant progress of generative models in the natural sciences, their controllability remains a key challenge. One fundamentally missing aspect of molecular or protein generative models is an inductive bias that can reflect continuous properties of interest. To that end, we propose the Regression Transformer (RT), a novel method that abstracts regression as a conditional sequence modeling problem. Thus, the RT introduces a new paradigm of multitask masked language models that seamlessly bridge sequence regression and conditional sequence generation tasks.

First, we thoroughly demonstrate that, despite using a nominal-scale training objective, the RT matches or surpasses the performance of conventional regression models in small molecule as well as protein property prediction task. Critically, priming the same model with continuous properties yields a highly competitive conditional generative model that outperforms specialized approaches in a substructure-constrained, property-driven molecule generation benchmark by > 100%. Our dichotomous approach is facilitated with a novel, alternating training scheme that is even shown to benefit text generation in natural language tasks.

In sum, the Regression Transformer is the first report of a “swiss army knife” model that concurrently excels at predictive and generative tasks in biochemistry. This finds particular application in property-driven, local exploration of the chemical or protein space.



## Rationalizing Stereoselective Membrane Permeability of Mutanobactin D and Synthetic Analogues with Molecular Dynamic Simulations

P. Brandl<sup>1</sup>, A. Albertini<sup>1</sup>, S. Linker<sup>1</sup>, F. Pultar<sup>1</sup>, S. Riniker<sup>1\*</sup>

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

Mutanobactins are semipeptidic macrocycles produced by *Streptococcus mutans* that impact the yeast-mycelium transition of *Candida albicans*. [1,2] Among various mutanobactin metabolites, mutanobactin D has the strongest inhibitory effect on *C. albicans* biofilm formation. [3] While its mechanism of action is not yet fully resolved, that same work suggests that the conformational ensembles, which are accessible to the compounds depending on their specific stereochemistry, are decisive for their activity.

This work aims to investigate conformational changes in Mutanobactin D, its stereoisomers and other synthetic analogues upon interaction with the fungal cell membrane. Previous works have successfully linked improved cell permeation, which is a prerequisite for bioactivity, to the existence of congruent conformational states between the macrocycles' conformational ensembles in polar and apolar environments. [4,5] We hence explore the conformational space of these macrocycles in polar and apolar environments, namely water and chloroform, with the help of molecular dynamic simulations. To capture differences in the ensemble populations and kinetics between different solvents we perform a thorough statistical analysis based on a Markov-State-Models framework.

Furthermore, we construct a model system consisting of the macrocycles at an H<sub>2</sub>O/CHCl<sub>3</sub> interface to elucidate the conformational mechanism which facilitate a membrane crossing. Reference [6] showed that modeling the interactions at this simpler polar/apolar interface serves as a good proxy for simulating the more detailed model using a POPC lipid bilayer. As a next step we build the latter more advanced model based on the structural preferences identified at the H<sub>2</sub>O/CHCl<sub>3</sub> interface. With these computational setups of increasing size and complexity we rationalize the stereo-selective permeability and activity observed in the Mutanobactin D analogues.

[1] Chenggang Wu, Robert Cichewicz, Yihong Li, Jinman Liu, Bruce Roe, Joseph Ferretti, Justin Merritt, Fengxia Qi, *Applied and environmental microbiology*, **2010**, 76, 5815–5826.

[2] P. Matthew Joyner, Jinman Liu, Zhijun Zhang, Justin Merritt, Fengxia Qi, Robert H. Cichewicz, *Organic & biomolecular chemistry*, **2010**, 8, 5486–5489.

[3] Felix Pultar, Moritz E. Hansen, Susanne Wolfrum, Lennard Bösel, Ricardo Fróis-Martins, Susanne Bloch, Alberto G. Kravina, Deren Pehlivanoglu, Christina Schäffer, Salomé Leibundgut-Landmann, Sereina Riniker, Erick M. Carreira, *Journal of the American Chemical Society*, **2021**, 143, 10389–10402.

[4] Jagna Witek, Shuzhe Wang, Benjamin Schroeder, Robin Lingwood, Andreas Dounas, Hans-Jörg Roth, Marianne Fouché, Markus Blatter, Oliver Lemke, Bettina Keller, Sereina Riniker, *Journal of chemical information and modeling*, **2019**, 59, 294–308.

[5] Shuzhe Wang, Gerhard König, Hans-Jörg Roth, Marianne Fouché, Stephane Rodde, Sereina Riniker, *Journal of medicinal chemistry*, **2021**, 64, 12761–12773.

[6] Stephanie M. Linker, Christian Schellhaas, Benjamin Ries, Hans-Jörg Roth, Marianne Fouché, Stephane Rodde, Sereina Riniker, *RSC Advances*, **2022**, 12, 5782–5796.

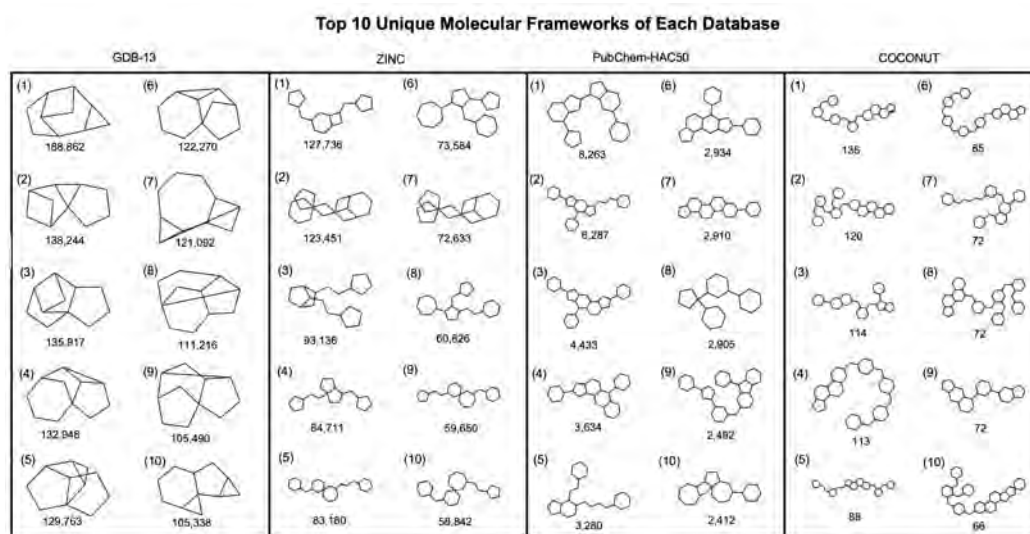
## Molecular Framework Analysis of the Drug-like Chemical Space

Y. Buehler<sup>1</sup>, J. L. Reymond<sup>1\*</sup>

<sup>1</sup>University of Bern, Department of Chemistry, Biochemistry and Pharmacy, Freiestrasse 3, 3012 Bern

**Abstract:** Millions of organic molecules of interest for drug discovery are available in public databases spanning from small fragments to screening compounds and natural products. In 1996 Bemis and Murcko proposed to classify drug molecules according to their molecular frameworks,<sup>1</sup> which are what remains after iteratively removing all terminal atoms from a molecule and converting all atoms to carbon and all bonds to single bonds. Here we present a large scale comparison of the molecular frameworks occurring in molecules from PubChem (101 million molecules in the public domain, with heavy atom count not greater than 50),<sup>2</sup> ZINC (886 million commercially available screening compounds),<sup>3</sup> COCONUT (401,624 natural products and natural product-like molecules)<sup>4</sup> and GDB13 (977 million possible organic molecules up to 13 atoms).<sup>5</sup> We identify the most frequent molecular frameworks shared by all four databases, as well as molecular frameworks unique to each of the four databases (Figure 1). This comparative analysis shows that a very vast chemical space remains entirely unexplored even at the level of very small organic fragments such as those in GDB13.

**Keywords:** Molecular framework; Chemical space; Structural diversity; Drug discovery



[1] Guy W. Bemis, Mark A. Murcko, *J. Med. Chem.*, **1996**, 39, 2887-2893.

[2] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, Evan E Bolton, *Nucleic Acids Research*, **2019**, 47, D1102-D1109.

[3] John J. Irwin, Khanh G. Tang, Jennifer Young, Chinzorig Dandarchuluun, Benjamin R. Wong, Munkhzul Khurelbaatar, Yurii S. Moroz, John Mayfield, and Roger A. Sayle, *J. Chem. Inf. Model.*, **2020**, 60, 6065-6073.

[4] Maria Sorokina, Peter Merseburger, Kohulan Rajan, Mehmet Aziz Yirik and Christoph Steinbeck, *J. Cheminf.*, **2021**, 13, 2.

[5] Lorenz C. Blum, Jean-Louis Reymond, *J. Am. Chem. Soc.*, **2009**, 131, 8732-8733.



## Efficient linear-response TDDFT implementation for core-level spectroscopy of large and periodic systems

A. Bussy<sup>1</sup>

<sup>1</sup>University of Zurich

X-ray absorption spectroscopy (XAS) is a popular tool to probe the local geometric and electronic structure of matter. To complement measurements, calculations based on excited state methods such as time-dependent density functional theory (TDDFT) are usually needed.

Therefore, we recently implemented a new TDDFT based XAS simulation method in the CP2K software package. The method exploits the localized nature of core electronic states and Gaussian basis sets by using a restricted resolution of the identity (RI) scheme, allowing for efficient calculation of large and periodic systems.

XAS TDDFT correctly reproduces the spacing and relative intensities of experimental spectral features, but absolute energies are off. To match experiments, the spectrum is usually rigidly shifted, either empirically or based on some separate  $\Delta$ SCF calculation. We tackle this problem differently and propose an integrated fully ab-initio GW correction scheme that leverages the same efficient RI approach as the base TDDFT calculation.

## Accessing the complex free energy landscape of any (photoswitchable) organocatalysts with a quantum chemical accuracy

F. Celerse<sup>1</sup>, R. Fabregat<sup>1</sup>, S. Vela<sup>1</sup>, S. Gallarati<sup>1</sup>, C. Corminboeuf<sup>1,2\*</sup>

<sup>1</sup>EPFL, Laboratory of Computational Molecular Design, Institute of Chemical Sciences and Engineering, <sup>2</sup>National Center of Competence in Research-Catalysis (NCCR-catalysis)

Photoswitchable organocatalysts constitute an appealing alternative to standard catalysts, which afford enhanced control of the spatial orientation of the reagents through the use of a molecular photoswitching unit. Yet, many technical and scientific questions remain to ensure that the different configurational states (ON/OFF, i.e., active/inactive) are indeed capable of controlling catalytic activity and selectivity. Some of these challenges are ideally suited to be resolved *in silico* under the conditions that new quantum chemistry approaches are developed.

Within this context, one of the main objectives requires the development of methods that promote the thorough exploration of the free-energy landscapes of any photoswitchable systems. Within this work, we specifically advance state-of-the-art atomistic machine learning techniques to have access to the free energy landscapes of *any* fluxional organic molecules at a cost orders of magnitude smaller than with the explicit quantum chemical computations. Placing emphasis on the transferability of the model, we achieve this goal by first building an extensive database representative of structurally and chemically diverse characteristic organic moieties. We then show how such a dataset can be used to train a sparse local kernel regression (LKR) model [1] to explore complex free energy landscapes of the configurational states of photoswitchable organocatalysts.

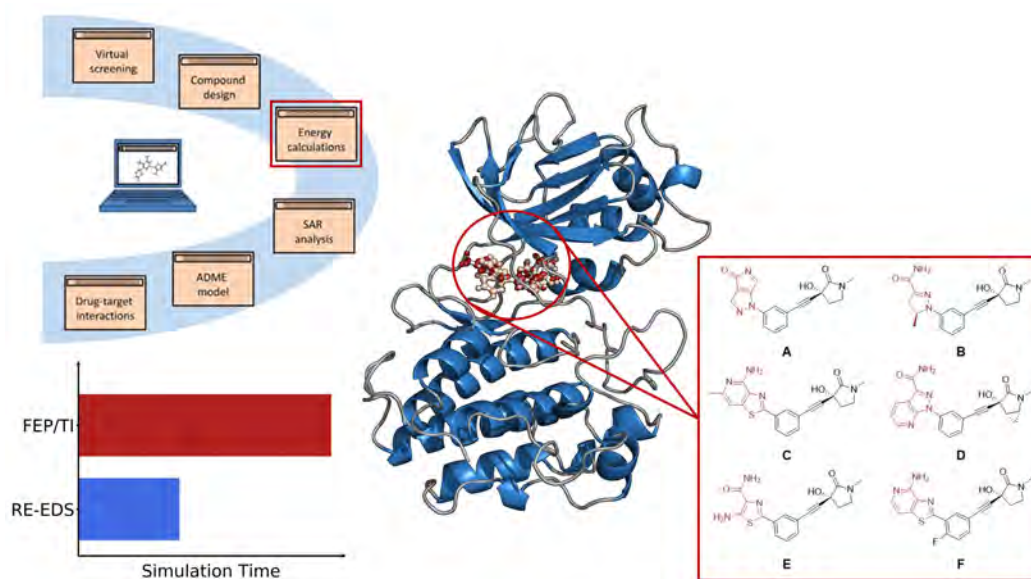
[1] Raimon Fabregat, Alberto Fabrizio, Edgar Engel, Benjamin Meyer, Veronika Juraskova, Michele Ceriotti, & Clemence Corminboeuf, *Journal of Chemical Theory and Computation*, **2022**, 18(3), 1467-1479.

## Predicting Multiple Ligand Binding Affinities from Single Alchemical Free Energy Calculations

C. Champion<sup>1</sup>, R. Gall<sup>1</sup>, B. Ries<sup>1</sup>, S. Rieder<sup>1</sup>, E. P. Barros<sup>1</sup>, S. Riniker<sup>1\*</sup>

<sup>1</sup>ETH Zurich, Laboratory of Physical Chemistry, Vladimir Prelog Weg 2, 8093 Zurich

Computational chemistry has become an important tool in virtually all stages of the drug discovery process. The prediction of binding affinities *in silico* during the lead optimization phase reduces costs significantly allowing to focus synthetic efforts on the most promising compounds. The routine calculation of rigorous protein-ligand binding affinities from molecular dynamics (MD) simulations have only become feasible in industry in the last few years due to their comparatively high computational costs. Standard alchemical free-energy methods require the calculation of (nearly) all pairwise combinations of ligands to be ranked, resulting in a combinatorial problem when the number of ligands increases. To address this issue, our research group is developing a method termed replica-exchange enveloping distribution sampling (RE-EDS) [1–3], where all transformations are performed from the same simulation, thus decreasing the costs by an order of magnitude. These cost reductions will facilitate the increased uptake of rigorous free-energy calculations in industry, and contribute to the development of new small organic drugs. In this talk, we present our recent work with RE-EDS, applying the method to a series of kinases (NIK, PAK, CHK1) and predicting binding free energies with similar accuracy as state-of-the-art method (thermodynamic integration). We also discuss improvements made to our methodology and the fully automated open-source Python pipeline (<https://github.com/rinikerlab/reeds>).



- [1] Dominik Sidler, Arthur Schwaniger, Sereina Riniker, *J. Chem. Phys.*, **2016**, 145, 154114  
 [2] Dominik Sidler, Michael Cristòfol-Clough, Sereina Riniker, *J. Chem. Phys.* 2016, 145, 154114  
 [3] Benjamin Ries, Karl Normak, Richard Gregor Weiss, Salomé Rieder, Emília Pécora de Barros, Candide Champion, Gerhard König, Sereina Riniker, *J. Chem. Phys.* 2016, 145, 154114

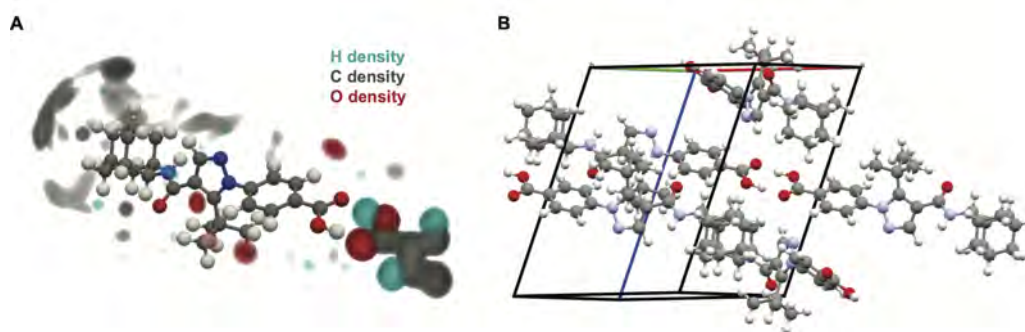
**PGT: A machine learning model to accelerate IP generation in chemistry**D. Christofidellis<sup>1,2</sup><sup>1</sup>IBM Research, <sup>2</sup>Queen's University Belfast

Patents are a valuable source of knowledge but drafting them is a time-consuming and expensive task. Methods that assist patent generation can provide a two-fold improvement as they can speed up the generation process and suggest to the inventor ideas and claims. Herein, influenced by recent advances in language modeling via multitask learning and prompt engineering, we present Patent Generative Transformer (PGT), a transformer-based language model trained to facilitate patent drafting. Specifically, the model supports three tasks: part-of-patent generation, text infilling, and patent coherence evaluation. PGT complements inventors and assures the fast and successful transition from their input to a coherent patent disclosure taking advantage of its multitasking nature. Focusing on the applications of the model in the chemistry field, we examine its performance to generate chemistry-related patent texts. We also examine to what extent the model can suggest ideas for the broader chemistry field in a zero-shot setting.

**Chemical shift-dependent interaction maps in molecular solids**M. Cordova<sup>1,2</sup>, L. Emsley<sup>1,2\*</sup>

<sup>1</sup>Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, <sup>2</sup>National Centre for Computational Design and Discovery of Novel Materials MARVEL, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Solid-state NMR, in combination with computational methods such as density functional theory chemical shift computation and crystal structure prediction (CSP) procedures, has proven able to determine the crystal structure of organic solids. A major bottleneck of NMR crystallography is the generation of candidate structures through CSP, which requires substantial computational resources. Experimentally determined interatomic distance constraints are widely used to drive structure determination of proteins by NMR. Here, we analyse a database of chemical shifts predicted by machine learning for over 200,000 molecular crystals in order to identify intermolecular interactions in crystalline solids directly from the observed chemical shifts and associated atomic site in a molecule. This analysis provides three-dimensional maps of expected atomic densities in the solid state around any conformer, enabling the construction of distance constraints that could accelerate structure determination by solid-state NMR.



## Implementation of Nuclear Velocity Perturbation and Magnetic Field Perturbation Theory in CP2K and Their Application to Vibrational Circular Dichroism

E. Ditler<sup>1</sup>

<sup>1</sup>University of Zurich

We present the implementation of the nuclear velocity perturbation theory (NVPT) and the magnetic field perturbation theory (MFPT), using velocity-dependent atomic orbitals and gauge-including atomic orbitals, respectively [1]. The theory was implemented in the CP2K code package [2] which is based on a combination of Gaussian-type basis functions and plane waves (GPW) for the description of the electron density and arising potentials. The presence of non-local pseudo-potentials has been considered, as well as contributions arising from the basis functions being centered on the atoms.

The atomic polar tensors (APTs) and atomic axial tensors (AATs) are calculated in the length and velocity representations by employing Density Functional Perturbation Theory (DFPT) [3]. In the case of the NVP, the complete adiabatic formalism was employed to calculate IR absorption and VCD spectra. For the MFP implementation, gauge-including atomic orbitals were used.

To our knowledge, this is the first implementation of the NVPT and MFPT in the same code. It is also the first implementation of the NVPT using velocity-including atomic orbitals and the MFPT in the GPW approach. We discuss the basis set convergence and compare both implementations.

[1] E. Ditler, T. Zimmermann, C. Kumar, S. Luber, *J. Chem. Theory Comput.* 2022, 18, 4, 2448–2461

[2] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, *Computer Physics Communications*, 2005, 167, 2, 103-128

[3] E. Ditler, C. Kumar, S. Luber, *J. Chem. Phys.*, 2021, 154, 10412

**On the quantum chemical roots of molecular similarity descriptors for machine learning**S. Gugler<sup>1</sup>, M. Reiher<sup>1\*</sup><sup>1</sup>Laboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

Molecular similarity is a key concept for understanding and rationalizing chemistry.[1,2,3] The importance of descriptors that measure molecular similarity has seen renewed interest in recent applications of machine learning approaches for the prediction of molecular properties, especially of energies. Whereas these approaches typically address vast and heterogeneous regions of chemical space, truly similar molecular structures – such as those found in sequences of related structure produced by molecular dynamics trajectories, during geometry optimizations, and in reaction networks of elementary steps – have hardly been the target of molecular similarity analyses. Accordingly, descriptors have not been scrutinized at this very fundamental level. We close this gap by providing a detailed investigation into molecular similarity at the level of Born-Oppenheimer surfaces and chemical reaction networks. We start out from basic quantum chemical expressions for the electronic energy in order to obtain similarity measures that are clearly related to the proper quantum chemical basis. We show how the then emerging features can be related to existing similarity measures.

[1] Markus Reiher, *Isr. J. Chem.* **2021**, 62, e202100101.

[2] Gregor N. Simm, Markus Reiher, *J. Chem. Theory Comput.* **2018**, 14, 5238-5248.

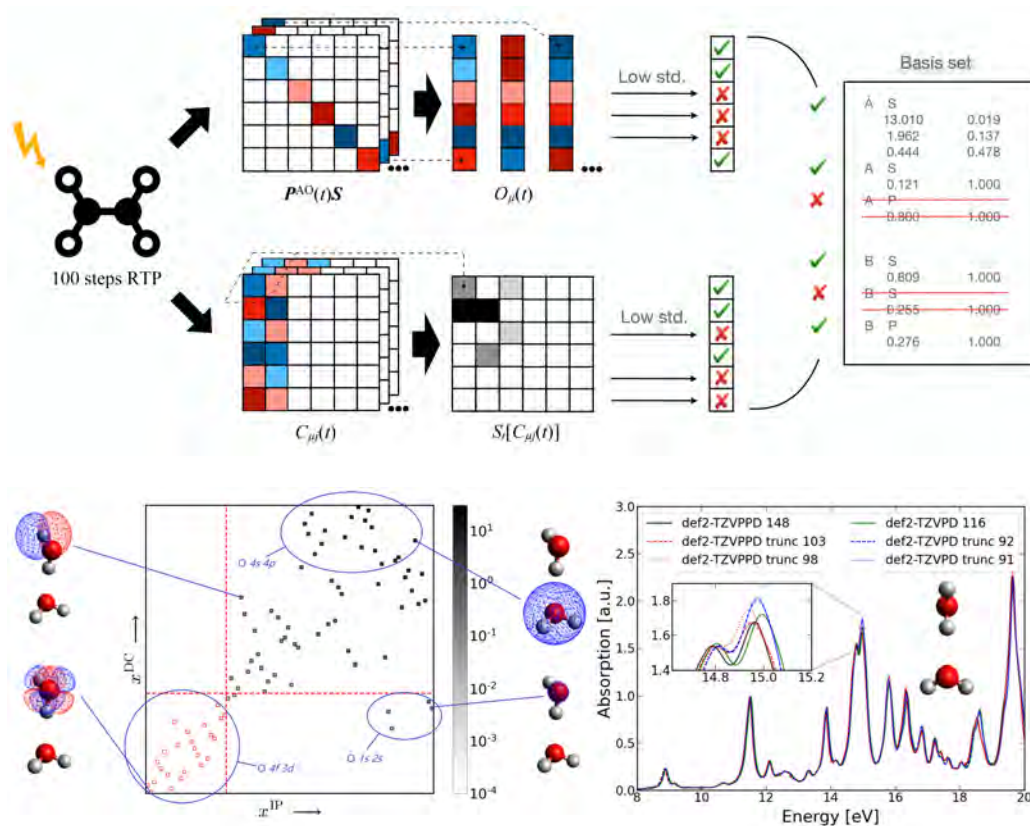
[3] Stefan Gugler, Markus Reiher, On the quantum chemical roots of molecular similarity descriptors for machine learning, *soon on arXiv*

## Automatic purpose-driven basis set truncation for time-dependent Hartree-Fock and time-dependent density-functional theory

R. Han<sup>1</sup>, J. Mattiat<sup>1</sup>, S. Luber<sup>1\*</sup>

<sup>1</sup>Department of Chemistry A, University of Zurich, 8057 Zurich, Switzerland

Real-time time-dependent density-functional theory (RT-TDDFT) and linear response time-dependent density-functional theory (LR-TDDFT) are two important approaches, in the context of simulating electronic spectra. For the sake of computational efficiency, we propose a robust scheme to truncate the atomic orbital (AO) basis set based on the analysis of the time evolution of the density matrix in the AO basis during the real-time propagation. The truncated bases are tested for electronic absorption spectra and electronic circular dichroism spectra with both LR and RT approaches, and provide an acceleration up to an order of magnitude while the shifts of excitation energies of interest are generally within 0.2 eV.



[1] Ruocheng Han, Johann Mattiat, Sandra Luber, in preparation.



**Scrutinizing Semi-empirical Methods for Excited Electronic States**

V. Klasovita<sup>1,2</sup>, T. Weymuth<sup>1,2</sup>, M. Reiher<sup>1,2\*</sup>

<sup>1</sup>Laboratory of Physical Chemistry, ETH Zurich, <sup>2</sup>National Centre of Competence in Research (NCCR) Catalysis, ETH Zurich, Zurich, Switzerland

Spectral signals can serve as diagnostic probes for identifying and characterizing molecular structures. In our group, we have implemented semiempirical methods, capable of calculating infrared (IR) and ultraviolet-visible (UV/Vis) spectra in an ultra-fast fashion, into our computer program, Sparrow [1]. To probe the performance of these methods, we calculate UV/Vis spectra with a series of first-principles reference methods, such as time-dependent density functional theory and approximate coupled cluster singles and doubles, for a set of test systems consisting of porphyrin-type structures. These calculations serve as a benchmark to which we compare the results obtained with semiempirical methods. The first test system is Tetraphenylporphyrin (TPP). The second system we investigate is Cobalt-Tetraphenylporphyrin (CoTPP). In a third test system, we introduce a ligand, N-Nitrosodimethylamine (NDMA), resulting in CoTPP + NDMA. We examine how the UV/Vis spectra calculated with ab initio methods change as we move from one test system to the next, and how the performance of semiempirical methods with respect to computation time and accuracy varies among them.

[1] Francesco Bosia, Thomas Weymuth, Markus Reiher, arXiv:2204.07534 [physics.chem-ph].

## NaviCat: Computational Tools to Navigate Molecular Catalyst Landscape

R. Laplaza<sup>1,2</sup>, M. D. Wodrich<sup>1,2</sup>, S. Das<sup>1</sup>, S. Gallarati<sup>1</sup>, C. Corminboeuf<sup>1,2\*</sup>

<sup>1</sup>Laboratory for Computational Molecular Design (LCMD), Institute of Chemical Sciences and Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), 1015 Lausanne, Switzerland, <sup>2</sup>National Center for Competence in Research-Catalysis (NCCR-Catalysis), Ecole Polytechnique Federale de Lausanne (EPFL), 1015 Lausanne, Switzerland

We introduce the NaviCat (**N**avigating **C**atalysis) platform, which assembles interoperable computational tools to accelerate catalyst discovery and optimization.

NaviCat currently includes: *volcanic*, an automated generator of volcano plot and activity/selectivity maps[1,2]; *NaviCatGA*, a genetic algorithm-based catalyst optimizer[3]; and *EPsim*, which measures similarity of computed energy profiles with respect to the ideal Sabatier catalyst[4]. These tools can be coupled to construct sophisticated pipelines, integrated with other python codes, or used separately as powerful standalone programs.

The *volcanic* module is, for instance, able to generate thermodynamic, kinetic and turn-over frequency volcano plots and activity maps readily from a handful of computed DFT energy profiles. *volcanic* automatically constructs robust linear scaling relationships, from which the performance can be predicted as a function of selected descriptor variables. The generated maps allow the user to rationalize catalytic performance by highlighting the limiting factors in the catalytic cycle, as well as ascertaining general trends.

Providing a natural access to the optimum region of catalytic activity, these maps can also be used as inexpensive fitness functions to drive catalyst optimization. Such optimizations can be done by directly interfacing *volcanic* with *NaviCatGA*.

Overall, the NaviCat platform aims at elaborating original data-driven tools and broadening their use in the field of computational homogeneous catalysis.

**NaviCat**  
volcAnic EPSSim NaviCatGA

[1] M. D. Wodrich, *et al. Accounts of Chemical Research*, **2021**, 54, 5, 1107-1117

[2] R. Laplaza, *et al. Nature Protocols*, **2022**, Accepted

[3] R. Laplaza, *et al. Chemistry - Methods*, **2022**, Accepted

[4] M. D. Wodrich, *et al. Chemical Science*, **2020**, 11, 12070-12080

## Quantifying Tunnelling in Molecules and Clusters with Instanton Theory

G. Laude<sup>1</sup>, J. O. Richardson<sup>1\*</sup>

<sup>1</sup>Laboratory of Physical Chemistry, ETH Zürich

Quantum tunnelling is a phenomenon which is ever-present in a large number of molecular systems. It can result in larger-than-expected reaction rates, and splitting of energy levels and delocalisation. However, unfavourable scaling of computational times with respect to system size hamper many attempts at quantifying tunnelling effects in molecular systems. With instanton theory,<sup>1,2</sup> one will only have to locate a dominant tunnelling pathway (called the instanton) to determine reaction rates and tunnelling splittings. In combination with machine-learning approaches<sup>3,4</sup> and a divide-and-conquer algorithm for evaluating fluctuations along the instanton pathway,<sup>7</sup> the approach becomes less computationally demanding and more numerically stable. We have now extended instanton theory such that we are able to evaluate tunnelling splittings in asymmetric systems,<sup>5,6</sup> and of vibrationally-excited states.<sup>8</sup> We have applied these new extensions to a set of one- and two-dimensional systems, for which our results compare favourably with numerically-exact quantum mechanics. Finally, we apply our method to some multidimensional systems, namely, malonaldehyde, and  $\alpha$ -fenchol, an asymmetric biomolecule.

- [1] J. O. Richardson, "Ring-polymer instanton theory", *Int. Rev. Phys. Chem.* **2018**, 37, 171-216.  
[2] J. O. Richardson, "Perspective: Ring-polymer instanton theory", *J. Chem. Phys.* **2018**, 148, 200901.  
[3] G. Laude, D. Calderini, D. P. Tew and J. O. Richardson, "Ab initio instanton rate theory made efficient using Gaussian process regression", *Faraday Discuss.* **2018**, 212, 237-258.  
[4] G. Laude, D. Calderini, R. Welsch and J. O. Richardson, "Calculations of quantum tunnelling rates for Muonium reactions with methane, ethane and propane", *Phys. Chem. Chem. Phys.* **2020**, 22, 16843.  
[5] E. Jahr\*, G. Laude\* and J. O. Richardson, "Instanton theory of tunneling in molecules with asymmetric isotopic substitutions", *J. Chem. Phys.* **2020**, 153, 094101.  
[6] M. Fiechter\*, G. Laude\* and J. O. Richardson, "Projected flux instanton theory for tunnelling between slightly asymmetric wells", in preparation.  
[7] P. Winter and J. O. Richardson, "Divide-and-conquer method for instanton rate theory", *J. Chem. Theory Comput.*, **2019**, 15, 2816-2825.  
[8] G. Laude and J. O. Richardson, "Tunnelling splittings of vibrationally-excited states from ring-polymer instanton theory", in preparation.

**Nonadiabatic processes in condensed phase systems with  $\Delta$ SCF**

M. Mališ<sup>1</sup>, E. Vandaele<sup>1</sup>, S. Lubert<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

The variational delta self-consistent field ( $\Delta$ SCF) density functional theory-based method enables full atomistic simulations of nonadiabatic processes in condensed phase systems by balancing computational costs and accuracy of the description of excited electronic state properties. We will discuss novel Ansätze in order to incorporate spin-orbit coupling and certain multi-reference characters into  $\Delta$ SCF calculations. We applied our methodology for the investigation of nonradiative deactivation mechanisms in a number of solvated systems, emphasizing the details of chromophore-environment interactions and their influence on photochemical processes in condensed phase systems.

[1] Momir Mališ, Sandra Lubert, *Journal of Chemical Theory and Computation*, **2020**, 16, 4071-4086.

[2] Momir Mališ, Sandra Lubert, *Journal of Chemical Theory and Computation*, **2021**, 17, 1653-1661.

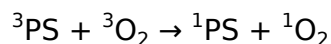
[3] Eva Vandaele, Momir Mališ, Sandra Lubert, *Physical Chemistry Chemical Physics*, **2022**, 24, 5669-5679.

[4] Eva Vandaele, Momir Mališ, Sandra Lubert, *The Journal of Chemical Physics*, **2022**, 156, 130901.

[5] Momir Mališ, Eva Vandaele, Sandra Lubert, *Accepted for publication in the Journal of Chemical Theory and Computation*.

**Quantum Effects in Photosensitization: Possibility of Tunneling?**M. A. Manae<sup>1</sup>, J. O. Richardson<sup>1\*</sup><sup>1</sup>Laboratory of Physical Chemistry, ETH Zurich

Photosensitization is the indirect electronic excitation of a molecule with the help of a photosensitizer (PS), and is a bimolecular nonradiative energy transfer process. It is employed when direct photo-excitation of the target molecule is not feasible. A prototypical example is O<sub>2</sub>, where the direct photo-excitation from its ground triplet state to its excited singlet state is spin forbidden, and its photosensitization can be represented as



Photosensitization of O<sub>2</sub> is widely made use of in organic synthesis, photovoltaic cells and is a key step in photodynamic therapy. While photosensitization has been widely studied experimentally, the complete mechanism of photosensitization is still unknown. A recent theoretical study [1] attempted to elucidate its mechanism by calculating rate constants of photosensitization of oxygen by thiothymines (2-thiothymine, 4-thiothymine and 2,4 dithiothymine), which are known to be excellent PSs. The study used two methods to calculate rate constants, a classical and a quantum method, and found quantum effects to play a significant role in photosensitization, with the classical method underestimating rate constants by several orders of magnitude. This study motivated investigating the role of tunneling in photosensitization. Calculations using Golden rule non-adiabatic instanton theory have found evidence of tunneling in the intersystem crossing pathways of these PSs - the thiothymines. Further, the nuclear mode that is responsible for intersystem crossing in an isolated PS is the same along which photosensitization occurs [1]. These results indicate that tunneling plays a crucial role in energy transfer from the PS to O<sub>2</sub> as well.

[1] Meghna A. Manae, Anirban Hazra, *Phys. Chem. Chem. Phys.*, **2022**, DOI: 10.1039/d2cp01337a

**DFT-Based Mechanistic Investigation Of Zinc(II) Catalyzed O-Arylation Reaction**L. E. Mathews<sup>1</sup>, C. Rajalakshmi<sup>1</sup>, P. Santhoshkumar<sup>1</sup>, V. I. Thomas<sup>1\*</sup><sup>1</sup>Department of Chemistry, CMS College Kottayam (Autonomous), Kottayam, Kerala, 686001, India

Zinc catalyzed cross-coupling reactions have captivated attention these days with regard to the low cost, mild reaction conditions, and good functional group tolerance, despite the non-redox character of the Zn metal<sup>1,2</sup>. Among the variants in Zn catalyzed cross-coupling reactions, Ullmann-type O-arylation reactions have garnered attention owing to their importance in preparing diaryl ethers<sup>3</sup>. Esteeming the notable advancement in the Zn catalyzed cross-coupling protocols, the mechanistic details of the former, however, remain scanty. In the present work, a detailed computational investigation into the molecular mechanism of the Zn (II) catalyzed C-O cross-coupling of phenols with activated iodoarenes is carried out using the Density Functional Theory method employing B3LYP-D3 functional [3] augmented with CPCM solvation model using acetonitrile as the solvent. The Zn and I atoms are described using (LANL2DZ) for the inner electrons and its associated double- $\zeta$  basis set for the outer electrons. The C, H, N, and O atoms were described by a 6-31+G(d) basis set. L-proline is the ancillary ligand used in this study. The active catalyst species is found to be trigonal, L-proline ligated Zn (II) phenoxide complex. The investigation revealed that the governing mechanism for this Ullmann-type O-arylation reaction proceeds through a  $\sigma$  - bond metathesis mechanism involving the concerted breaking of the Csp<sup>2</sup> - I bond and the formation of the Csp<sup>2</sup> - O bond proceeding through a four-centered transition state. Frontier Molecular Orbital (FMO) analysis was performed to investigate the effect of functional groups at the para position of the substrates. The substitution by electron-withdrawing groups (EWG) considerably decreases the energy of FMOs involved. The presence of EWG on aryl iodides tend to favor the reaction by reducing the energy of its LUMO leading to a decreased HOMO-LUMO gap. The mechanistic investigation into these non-redox transition metal-catalyzed cross-couplings could aid in designing better cross-coupling reaction protocols.

[1] S. Enthaler, *ACS Catal.*, 2013, **3**, 150-158.

[2] S. Tang, L. Zeng, Y. Liu and A. Lei, *Angew. Chemie*, 2015, **127**, 16076-16079.

[3] K. K. Krishnan, A. Harry and S. M. Ujwaldev, 2018, **4**, 3984-3988.

## Spectroscopic simulations with real-time TDDFT - linear response and gauge dependence

J. Mattiat<sup>1</sup>, S. Luber<sup>1\*</sup>

<sup>1</sup>University of Zurich

In this contribution the versatility of real-time time-dependent density functional theory (RT-TDDFT) is illustrated by showcasing its ability to simulate UV-VIS absorption, electric circular dichroism (ECD), (resonance) Raman and (resonance) Raman optical activity (ROA) spectra, the latter two within the short time approximation.

For the formulation of the spectroscopic response tensors a unified formalism in terms of linear response propagators is applied, allowing insights into how the perturbation and response operators are distinguishable in the real-time linear response protocol.

Special emphasis is on the choices of gauge and the coupling of the electro-magnetic fields to the non-local part of pseudo potentials, a proper handling of which proves to be vital for an adequate description of the chiral spectroscopies, ECD [1] and ROA [2].

For UV-VIS absorption and Raman spectroscopy results for non-periodic [3] and periodic [4] simulation cells are presented, drawing on the velocity gauge and the modern theory of polarization.

[1] Mattiat, J., and Luber, S., Chem. Phys., 2019, 527, 110464.

[2] Mattiat, J., and Luber, S., J. Chem. Phys., 2019, 151, 234110.

[3] Mattiat, J., and Luber, S., J. Chem. Phys., 2018, 149, 174108.

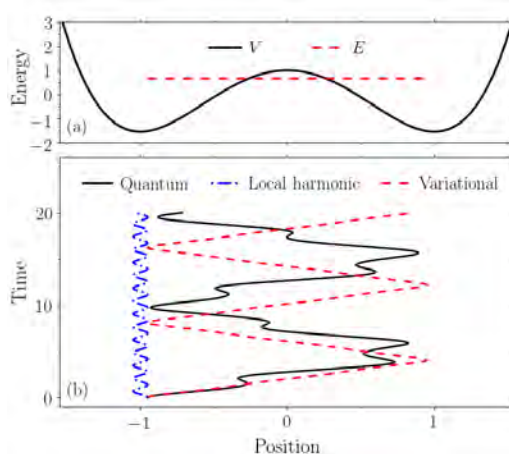
[4] Mattiat, J., and Luber, S., J. Chem. Theory Comput., 2021, 17, 1, 344-356.

## The variational Gaussian approximation combined with high-order geometric integrators with applications to quantum tunneling and vibronic spectra

R. Moghaddasi Fereidani<sup>1</sup>, J. Vanicek<sup>1\*</sup>

<sup>1</sup>Laboratory of Theoretical Physical Chemistry, EPFL, Lausanne, Switzerland

Among the single trajectory Gaussian-based methods for solving the time-dependent Schrödinger equation, the variational Gaussian approximation (VGA) [1,2] is the most accurate one. However, the equations of motion for the parameters of the Gaussian wavepacket require expectation values of the potential and its first two derivatives, making the method much more expensive than original Heller's thawed Gaussian approximation (TGA) [3], which requires these potential energy properties only at the center of the wavepacket. To improve the efficiency of the VGA, we describe geometric integrators, which can achieve an arbitrary even order of convergence in the time step and are obtained by symmetrically composing the second-order symplectic integrator of Faou and Lubich [4]. We demonstrate that the high-order integrators can drastically speed up convergence compared to the second-order algorithm and, in contrast to the Runge-Kutta method, are time-reversible and conserve the norm exactly. To avoid making further approximations, we demonstrate the properties of the VGA and of the geometric integrators on several systems, in which the expectation values of the potential energy can be evaluated analytically. We show that the VGA, in contrast to the TGA, conserves energy exactly and takes into account tunneling at least qualitatively (see figure) and, in calculation of vibronic spectra, it agrees better than the TGA and global harmonic approximation with benchmark exact quantum calculations. Finally, to show that the method is not restricted to low-dimensional systems, we also applied it to a nonseparable twenty-dimensional model of coupled Morse oscillators.



[1] E. Heller, *The Journal of Chemical Physics*, **1976**, 64, 63-73.

[2] R. Coalson, M. Karplus, *The Journal of Chemical Physics*, **1990**, 93, 3919-3930.

[3] E. Heller, *The Journal of Chemical Physics*, **1975**, 62, 1544-1555.

[4] E. Faou, C. Lubich, *Comput. Visual Sci.*, **2006**, 9, 45-55.



**Streamlined Automatic Active Space Calculations through AutoCAS**M. Mörchen<sup>1</sup>, C. J. Stein<sup>2</sup>, M. Reiher<sup>1\*</sup>

<sup>1</sup>ETH Zürich, Laboratorium für Physikalische Chemie, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland, <sup>2</sup>Universität Duisburg-Essen, Fakultät für Physik, Lotharstraße 1, 47057 Duisburg, Germany

We present our new Python based implementation of the AutoCAS<sup>1,2,3,4</sup> software, which enables black-box multi-configurational calculations. A great problem in multi-configurational quantum chemistry is the choice of the active space for strongly correlated systems, such as open-shell metal complexes or dissociating molecular systems. AutoCAS tackles these problems by exploiting quantum information theory and the Density Matrix Renormalization Group method provided by the QCMAquis<sup>5</sup> program package, resulting in an automatic choice of orbitals for the active space. The re-implementation allows to streamline active space calculations either via direct use of the Python package or through the command line interface. Furthermore, to recover the dynamic correlation, methods like Tailored Coupled Cluster<sup>6,7,8</sup> can be routinely applied to improve on the prior active space calculations.

- [1] C. J. Stein and M. Reiher, *J. Comput. Chem.*, **2019**, 40, 2216-2226.
- [2] C. J. Stein and M. Reiher, *J. Chem. Theory Comput.*, **2016**, 12, 1760-1771.
- [3] C. J. Stein, V. von Burg and M. Reiher, *J. Chem. Theory Comput.*, **2016**, 12, 3764-3773.
- [4] C. J. Stein and M. Reiher, *Chimia*, **2017**, 71, 170-176.
- [5] S. Keller, M. Dolfi, M. Troyer, and M. Reiher, *J. Chem. Phys.*, **2015**, 143, 244118.
- [6] T. Kinoshita, O. Hino and R. J. Bartlett, *J. Chem. Phys.*, **2005**, 123, 074106.
- [7] L. Veis et al., *J. Phys. Chem. Lett.*, **2016**, 7, 4072-4078.
- [8] M. Mörchen, L. Freitag and M. Reiher, *J. Chem. Phys.*, **2020**, 153, 24413.

**Bachelor Students' Understanding of Basic Quantum Chemical Concepts**C. H. Müller<sup>1</sup>, M. Kapur<sup>2\*</sup>, M. Reiher<sup>1\*</sup>

<sup>1</sup>Laboratory of Physical Chemistry, ETH Zurich, <sup>2</sup>Professorship of Learning Sciences and Higher Education, ETH Zurich

Chemical knowledge has a unique structure in the sense that every information should be understood on multiple levels which might seem – to the untrained eye – independent. In other words, what we see in the flask before us must be understood on a sub-microscopic, invisible level that additionally must be described in terms of symbolic, mathematical characteristics such as energy [1].

In two studies with bachelor chemistry students, we have explored what their conceptions of selected target concepts are and how we can support their learning through real-time interaction with quantum chemical simulations. The learning environment was embedded in the graphical user interface SCINE Heron and allowed to interact with semi-empirical calculations in real-time either with a computer mouse or with a haptic device [2-6].

In the studies, we focused on the target concepts potential energy, activation barrier and how these concepts are connected. While they are introduced as early as high school, they do not seem to be well understood. The results indicate that there was no significant development in the conceptions around these concepts between the second and fourth semester of chemistry students. If we can avoid or counteract misconceptions at this fundamental level, we can prepare the students better for more advanced concepts and ergo for their professional career.

Here, we present the analysis of the learning trajectories in more detail and explore, how we might be able to support the students' learning process.

[1] H. Tümay, "Reconsidering learning difficulties and misconceptions in chemistry: Emergence in chemistry and its implications for chemical education," *Chemistry Education Research and Practice*, **2016**, 17(2), 229-245.

[2] M. P. Haag and M. Reiher, "Real-time quantum chemistry," *International Journal of Quantum Chemistry*, **2013**, 113(1), 8-20.

[3] A. C. Vaucher, M. P. Haag and M. Reiher, "Real-time feedback from iterative electronic structure calculations," *Journal of Computational Chemistry*, **2016**, 37(9), 805 - 812.

[4] M. P. Haag, A. C. Vaucher, M. Bosson, S. Redon and M. Reiher, "Interactive Chemical Reactivity Exploration," *ChemPhysChem*, **2014**, 15, 3301-3319.

[5] M. P. Haag and M. Reiher, "Studying chemical reactivity in a virtual environment," *Faraday Discuss.*, **2014**, 169, 89-118.

[6] T. Weymuth and M. Reiher, "Immersive Interactive Quantum Mechanics for Teaching and Learning Chemistry," *Chimia*, **2021**, 75(1-2), 45-49.

## Bayesian Optimisation-accelerated Additives Screening and Yield Improvements in Chemical Reactions

B. Ranković<sup>1,4</sup>, R. Griffiths<sup>2,5</sup>, H. Moss<sup>3,2</sup>, P. Schwaller<sup>1,4\*</sup>

<sup>1</sup>EPFL SB-ISIC LIAC, <sup>2</sup>University of Cambridge, <sup>3</sup>Secondmind.ai, <sup>4</sup>NCCR Catalysis, <sup>5</sup>Mila - Quebec AI Institute

Reaction additives have been shown to be key to altering the reactivity and outcome of chemical reactions [1,2]. A recent study [3] reported the discovery of a phthalimide ligand additive for Ni-catalyzed photoredox decarboxylative arylations through high-throughput experimentation additives screening, which led to a 5-fold yield improvement by stabilising oxidative addition complexes and breaking up deactivated catalyst aggregates. As such large-scale screenings remain inaccessible for most research groups, we investigate how similar discoveries and yield improvements could be done efficiently with the lowest number of experiments possible. We use a Bayesian optimisation approach [4] and explore different initialisations, Gaussian process kernels [5] and acquisition functions. The unique reaction screening setting with 720 additives forces us to go beyond simple one-hot encoding of the reaction components. We investigate diverse molecular representations and demonstrate convincing results when compared to the baselines. Our approach is not limited to Ni-catalysed reactions but could generally be applied. Hence, it could lead to yield improvements in diverse cross-coupling reactions and unlock access to new chemical spaces of interest for the chemical and pharmaceutical industries.

[1] Bellomo, Ana, et al. "Additive effects on palladium-catalyzed deprotonative-cross-coupling processes (DCCP) of sp<sup>3</sup> C-H bonds in diarylmethanes." *Chemical Science* 4.2 (2013): 849-857.

[2] Vantourout, Julien C., et al. "Spectroscopic studies of the Chan-Lam amination: A mechanism-inspired solution to boronic ester reactivity." *Journal of the American Chemical Society* 139.13 (2017): 4769-4779.

[3] Prieto Kullmer, Cesar N., et al. "Accelerating reaction generality and mechanistic insight through additive mapping." *Science* 376.6592 (2022): 532-539.

[4] Shields, Benjamin J., et al. "Bayesian reaction optimization as a tool for chemical synthesis." *Nature* 590.7844 (2021): 89-96.

[5] GAUCHE: A Gaussian Process Library for Molecules, Proteins and Reactions.

## A Density Functional Theory Investigation on the Mechanism of Ligand-Free Manganese Catalyzed Ullmann Reaction

P. Santhoshkumar<sup>1</sup>, R. Sulay<sup>1</sup>, V. I. Thomas<sup>1\*</sup>

<sup>1</sup>CMS College, Kottayam, Kerala, India

Transition metal-catalyzed reactions in synthetic organic chemistry have attracted much attention due to the formation of C-heteroatom bonds. Ullmann couplings have risen in prominence in recent decades as a result of their use in the synthesis of biaryl ethers, which have been found in a wide range of natural products. Palladium and other heavy metals have exhibited their exceptional ability in various drives pointed towards amended coupling reactions. Regardless, palladium-catalyzed Ullmann coupling has a few impediments owing to the expensive nature and toxicity matters. Manganese as a catalyst has rarely been studied in the Ullmann reaction. In the present study, we have elucidated the mechanism governing the ligand-free manganese-catalyzed Ullmann coupling reaction between 4-iodonitrobenzene and phenol with the aid of the Density Functional Theory (DFT) method. The O-arylation in this manganese catalyzed Ullmann reaction occurs via the sigma bond metathesis mechanism. Oxidative addition- Reductive elimination is unsuitable in this procedure due to its reluctance to facilitate the oxidative addition of aryl halide on the active catalyst and the transition from the stable +2 to +4 oxidation state. Other reaction pathways such as single electron transfer and halogen atom transfer mechanisms possess high activation barriers making them kinetically inappropriate. The activation barrier obtained for the sigma bond metathesis mechanism is consistent with the experimental results. Furthermore, we have investigated the effect of the substituent on the aryl halide revealing that the electron-withdrawing nature of the substituent at the aryl halide leads to a decrease in the activation barrier, which is in good agreement with the experimental results and thus provides additional support for such a sigma bond metathesis mechanism.

**Deep learning assisted Suzuki cross coupling catalyst design**O. Schilter<sup>1,2</sup>, A. Vaucher<sup>1</sup>, T. Laino<sup>1\*</sup><sup>1</sup>IBM Research Europe, <sup>2</sup>EPFL

The need for more efficient and sustainable catalysts is ever-growing, and so are the costs associated with experimentally searching the chemical space to find a new promising catalyst. Computational methods, such as DFT, allows for screening molecules virtually for their suitability. B. Meyer et al. showed that the binding energy calculation between a ligand molecule and a metal center can be used as an indicator for testing a ligand suitability for the Suzuki cross-coupling reaction [1]. We present a state-of-the-art deep learning model able to predict the binding energy solely from string representations of catalysts and to generate new potential suitable ligand candidates, by self-learning meaningful features of these catalysts.

In a first step, inspired by the work of Gomez-Bombarelli et al [2], an RNN-based Variational Autoencoder (VAE) is trained, by teaching an encoder neural network to compress the catalyst representation (SMILES or SELFIES) into a continuous latent space while a second neural network, the decoder, learns to reconstructs the original representation from the latent space. Additionally, a feed forward neural network is trained to predict the associated binding energy from the latent space, which achieves similar predictive performance regarding the predicted energy (MAE = 2.41 kcal mol<sup>-1</sup>) compared to original approach which relies on the 3D geometry of the catalyst. This predictive neural network as well as the compression by the encoder and reconstruction by decoder leads to a meaningful and organized condensed representation of the inputted molecules in the lower-dimensional latent space.

The main advantage of our approach is the following: in addition to the possibility to predict the binding energy for catalysts assembled in a combinatorial fashion, this continuous and low-dimensional self-learned latent representation can be searched by gradient-based optimization to generate new molecules. After the optimization finds a latent vector associated with promising binding energy, one can decode the latent space back to a molecule representation using the trained decoder. Overall, this approach shows a promising new way to design new catalysts and can be adapted to different reaction classes.

[1] B. Meyer, B. Sawatlon, S. Heinen, A. Lilienfeld, C. Corminboeuf, *Chem. Sci.*, **2018**, *9*, 7069-7077

[2] R. Gómez-Bombarelli, J. Wei, D. Duvenaud, J. Lobato, B. Lengleng, D. Sheberla, J. Iparraguirre, T. Hirzel, R. Adams, A. Aspuru-Guzik, *ACS Cent. Sci.*, **2018**, *4*, 268-276

**Novel approaches for chiral spectroscopy of periodic systems**

L. Schreder<sup>1</sup>, S. Lubert<sup>1\*</sup>

<sup>1</sup>University of Zurich, Department of Chemistry, Winterthurerstrasse 190, 8057 Zurich, Switzerland

Computational studies of bulk materials are routinely done by imposing periodic boundary conditions (PBC) in the simulations leading, however, to an ill-defined position operator. This poses challenges to the calculations, especially for electric circular dichroism (ECD) and Raman optical activity (ROA) spectroscopy where magnetic dipole and electric quadrupole contributions are required. We show how maximally localized Wannier functions can be used to obtain ECD and ROA spectra for periodic systems, considering a careful choice of gauge, in this way also providing a valuable way for analysis of those spectra.

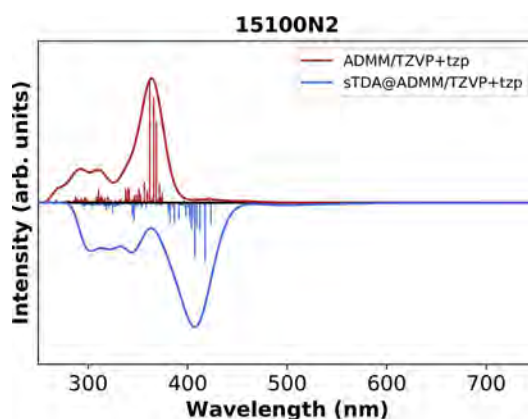
## Excited-State Properties for Extended Systems: Efficient Hybrid Density Functional Methods

B. Sertcan<sup>1</sup>, A. Hehn<sup>1</sup>, F. Belleflamme<sup>1</sup>, S. Chulkov<sup>2</sup>, M. Watkins<sup>2</sup>, J. Hutter<sup>1\*</sup>

<sup>1</sup>University of Zurich, Department of Chemistry, <sup>2</sup>University of Lincoln, School of Mathematics and Physics

Time-dependent density functional theory has become state-of-the-art for describing photophysical and photochemical processes in extended materials due to its affordable cost [1,2]. The inclusion of exact exchange was shown to be essential for the correct description of the long-range asymptotics of electronic interactions in order to achieve accurate descriptions of valence, Rydberg and charge-transfer excitations. While several approaches for an efficient treatment of exact exchange have been established for the ground state, the implementations for excited-state (ES) properties are rare. Furthermore, the high computational costs required for ES properties in comparison to ground-state computations often hinder large-scale applications on periodic systems with hybrid functional accuracy [3].

In this work [4], we propose two approximate schemes for improving computational efficiency for the treatment of exact exchange. Within the auxiliary density matrix method (ADMM), exact exchange is estimated using a relatively small auxiliary basis and the introduced basis-set incompleteness error is compensated by an exchange density functional correction term, and computational efficiency are improved by at least one order of magnitude [5,6]. Benchmark results for a test set of 35 molecules demonstrate that the mean absolute error introduced by ADMM is smaller than 0.3 pm for ES bond lengths and 0.07 eV for vertical excitation, adiabatic excitation and fluorescence energies. Computational timings for a series of covalent-organic frameworks taken from Curated COFs database [7,8] demonstrate that a speed-up of at least one order of magnitude can be achieved for ES geometry optimizations in comparison to conventional hybrid functionals.



The second method is to use a semi-empirical tight-binding approximation for both Coulomb and exchange contributions to the ES kernel. This simplified Tamm-Dancoff approximation (sTDA) achieves an accuracy comparable to approximated hybrid density functional theory when referring to highly accurate coupled-cluster reference data. We find that ES bond lengths deviate by 1.1 pm on average and mean absolute errors in vertical excitation, adiabatic excitation and fluorescence energies are smaller than 0.5 eV. In comparison to ADMM-approximated hybrid functional theory, sTDA accelerates the computation of broad-band excitation spectra by one order of magnitude, suggesting its potential use for large-scale screening purposes.

- [1] Andreas Dreuw, Martin Head-Gordon, *Chemical Reviews*, **2005**, 105, 4009-4037.  
 [2] Adèle D. Laurent, Denis Jacquemin, *International Journal of Quantum Chemistry*, **2013**, 113, 2019-2039.  
 [3] Marc de Wergifosse, Stefan Grimme, *Journal of Physical Chemistry A*, **2021**, 125, 3841-3851.  
 [4] Anna-Sophia Hehn, Beliz Sertcan, et al., *ChemRxiv* **2022**, <https://doi.org/10.26434/chemrxiv-2022-swkw9>  
 [5] Patrick Merlot, Róbert Izsák, et al., *Journal of Chemical Physics*, **2014**, 141, 094104  
 [6] Manuel Guidon, Jürg Hutter, et al., *Journal of Chemical Theory and Computation*, **2010**, 6, 2348-2364.  
 [7] Daniele Ongari, Aliaksandr V. Yakutovich et al., *ACS Central Science*, **2019**, 5, 1663-1675.  
 [8] Daniele Ongari, Leopold Talirz, et al., *ACS Central Science*, **2020**, 6, 1890-1900.

**Double-hybrid density functionals for the condensed phase: Gradients, stress tensor, and auxiliary-density matrix method acceleration**F. Stein<sup>1</sup><sup>1</sup>Universität Zürich

Due to their improved accuracy, double-hybrid density functionals emerged as an important method for molecular electronic-structure calculations. The high computational costs of double-hybrid calculations in the condensed phase and the lack of efficient gradient implementations thereof inhibit a wide applicability for periodic systems. We present an implementation of forces and stress tensors for double-hybrid density functionals within the Gaussian and plane-waves electronic structure framework. The auxiliary density matrix method is used to reduce the overhead of the Hartree-Fock kernel providing an efficient and accurate methodology to tackle condensed phase systems. First applications to water systems of different densities and molecular crystals show the efficiency of the implementation and pave the way for advanced studies. Finally, we present large benchmark systems to discuss the performance of our implementation on modern large-scale computers.

[1] Frederick Stein, Jürg Hutter, *J. Chem. Phys.*, **2022**, 156, 074107.

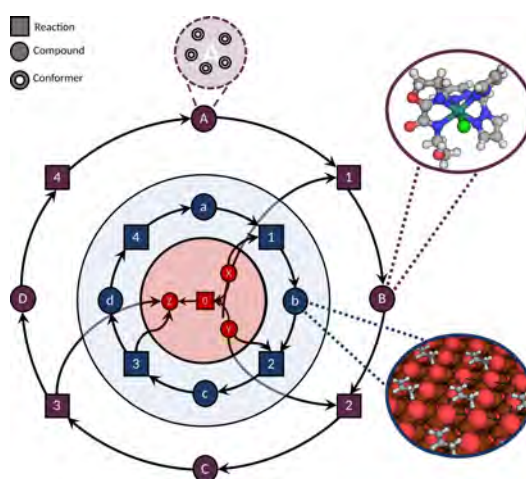


## Autonomous Reaction Network Exploration in Homogeneous and Heterogeneous Catalysis

M. Steiner<sup>1</sup>, M. Reiher<sup>1\*</sup>

<sup>1</sup>Laboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

Autonomous computations that rely on automated reaction network exploration algorithms [1, 2] facilitate systematic mechanism elucidations of catalytic reactions. They access several orders of magnitude more structures than what would be possible by manual inspection. This provides a high depth in terms of relevant species (including conformer resolution) and competing elementary reaction steps (including side, decomposition, and other deactivation reactions). The high degree of autonomy in automated exploration further reduces the amount of manual human work, processing errors, and human bias, which allows for a high fidelity of the formalization of catalytic processes and for surprising *in silico* discoveries. Hence, automated explorations offer key advantages for catalysis research. We have carried out automated explorations for both homogeneous [3, 4] and heterogeneous catalysis [5] and have analyzed the associated resource requirements for such computational campaigns [5]. Here, we present this work and outline how the costs can be reduced by fast electronic structure methods with uncertainty quantification and by allowing the exploration to be steerable with respect to specific regions of an emerging network and with respect to the addition of new reactant species.



[1] Gregor N. Simm, Alain C. Vaucher, Markus Reiher, *J. Phys. Chem. A*, **2019**, 123, 385-399.

[2] Jan P. Unsleber, Markus Reiher, *Annu. Rev. Phys. Chem.*, **2020**, 71, 121-142.

[3] Maike Bergeler, Gregor N. Simm, Jonny Proppe, Markus Reiher, *J. Chem. Theory Comput.*, **2015**, 11, 5712-5722.

[4] Stephanie A. Grimmel, Markus Reiher, *Faraday Discuss.*, **2019**, 220, 443-463.

[5] Miguel Steiner, Markus Reiher, *Top. Catal.*, **2022**, 65, 6-39.

**Computational Study on Multicopper Oxidase-mediated Biotransformations**V. J. Steiner<sup>1</sup>, H. Satoh<sup>1</sup>, J. Hutter<sup>1\*</sup>, K. Fenner<sup>1,2</sup>, A. Athanasakoglou<sup>2</sup><sup>1</sup>University of Zurich, 8057 Zurich, <sup>2</sup>Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600 Dübendorf

Substances found in municipal wastewater can be degraded by various enzymes during activated sludge treatment. One class of such enzymes are the laccases. The substrate range of these multicopper enzymes can be increased through mediators such as 4-hydroxybenzotriazole and 2,2'-azino-bis-3-ethylbenzthiazoline-6-sulphonic acid.<sup>[1]</sup>

21 wastewater-relevant substances that showed different degrees of transformation in such laccase-mediator systems<sup>[1]</sup> were examined on a theoretical basis through quantum chemical calculations. Several physico-chemical indices were analysed aiming to find an index that correlates strongly with observed reactivities, i.e., one that would be distinctly different in value for the reactive and non-reactive compounds. Such an index would potentially allow predicting the reactivity in laccase-mediator systems for substances yet to be examined. To this end, a qualitative analysis of individual indices as well as index combinations was conducted with regard to said reactivity. The second aspiration was to find indices with little computational time.

The combination of the one electron oxidation potential with  $\Delta E_{\text{HOMO-LUMO}}$  was found to be a good descriptor able to distinguish reactive from non-reactive substances. A machine learning model constructed using these descriptors with a support vector machine (SVM) showed a good performance for the classification of the molecules into reactive and non-reactive groups. Regarding low-cost computations, the analysis indicated that the electronegativity, the ionization potential and the HOMO energy can be replacements for the one electron oxidation potential, all of which require less computational time.

---

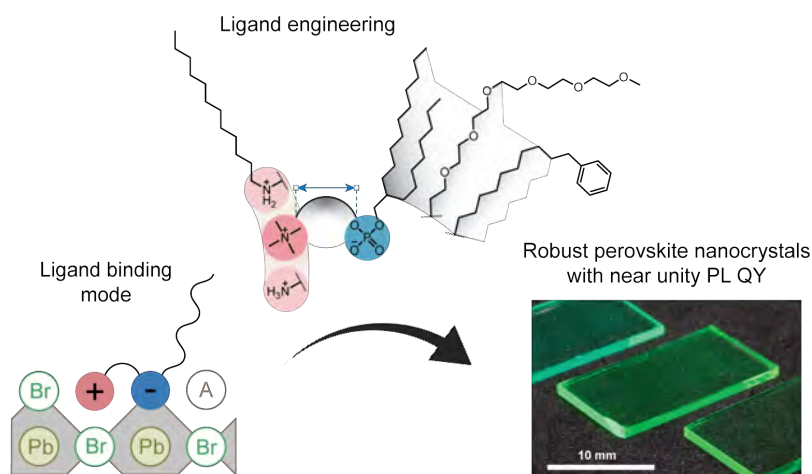
[1] Anastasia Athanasakoglou, Kathrin Fenner, *Environmental Science & Technology*, **2022**, 56, 313-324.

## Computational Design of Capping Ligands for Lead Halide Perovskite Nanocrystals

A. Stelmakh<sup>1,2</sup>, V. Morad<sup>1,2</sup>, M. Svyrydenko<sup>1,2</sup>, L. Feld<sup>1,2</sup>, M. Aebli<sup>1,2</sup>, A. Baumketner<sup>3\*</sup>, M. V. Kovalenko<sup>1,2\*</sup>

<sup>1</sup>Laboratory of Inorganic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir Prelog Weg 1, CH-8093 Zürich, Switzerland, <sup>2</sup>Laboratory for Thin Films and Photovoltaics, Empa – Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, CH-8600 Dübendorf, Switzerland, <sup>3</sup>Institute for Condensed Matter Physics, NAS of Ukraine, 1 Svientsitsky, Lviv 79011, Ukraine

Versatile surface functionalization of highly ionic surfaces, ubiquitous among inorganic nanomaterials, remains a formidable challenge in view of inherently non-covalent surface bonding. Colloidal lead halide perovskite nanocrystals (NCs), which are of interest for classical and quantum light generation,[1,2] are one of the examples. Despite some recent empirical progress in surface chemistry of lead halide perovskite NCs, the general strategy towards their robust surface functionalization still remains a challenge.[3] In this study we present the first structural investigation of perovskite NC surfaces capped with zwitterionic phospholipid molecules. In line with molecular dynamics simulations and solid-state NMR, zwitterionic phospholipid ligands bind to the surface of NCs with both head-groups by displacing the native ions of the perovskite. The ligand head-group affinity to the surface is therefore primarily governed by a geometric fitness of its cationic and anionic moieties into the crystal lattice. As a result, stable and colloiddally robust nanocrystals of inherently soft and chemically labile lead halide perovskites – FAPbX<sub>3</sub> and MAPbX<sub>3</sub> (X – Br, I) – can be obtained with a lattice-matched phosphoethanolamine head-group. Stable surface passivation enables excellent optical performance of the NCs. As an example, alkylphospholipid-capped FAPbBr<sub>3</sub> NCs display stable emission with a near-unity photoluminescence quantum yield in a broad concentration range, as well as in thick films. Ligand tail engineering, on the other hand, allows diverse surface functionalization of the NCs, broadening the scope of their potential applications.



[1] Mingming Liu, Qun Wan, Huamiao Wang, Francesco Carulli, Xiaochuan Sun, Weilin Zheng, Long Kong, Qi Zhang, Congyang Zhang, Qinggang Zhang, Sergio Brovelli, Liang Li, *Nature Photonics*, **2021**, 15, 379-385.

[2] Gabriele Rainò, Michael A. Becker, Maryna I. Bodnarchuk, Rainer F. Mahrt, Maksym V. Kovalenko, Thilo Stöferle, *Nature*, **2018**, 563, 671-675.

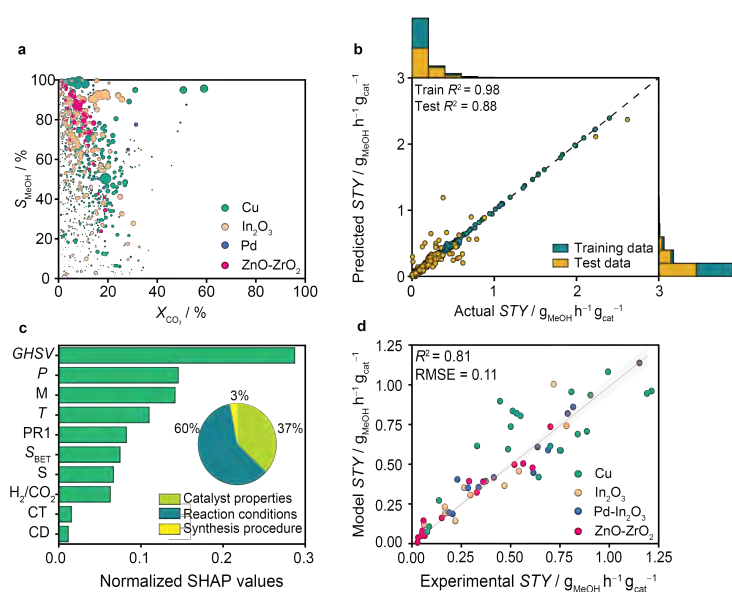
[3] Quinten A. Akkerman, Gabriele Rainò, Maksym V. Kovalenko, Liberato Manna, *Nature Materials*, **2018**, 17, 394-405.

## A generalized machine learning framework to predict the space-time yield of thermocatalytic CO<sub>2</sub> hydrogenation to methanol

M. Suvarna<sup>1</sup>, T. Pinheiro Araújo<sup>1</sup>, J. Pérez-Ramírez<sup>1\*</sup>

<sup>1</sup>ETH Zurich

Thermocatalytic CO<sub>2</sub> hydrogenation to methanol is an attractive decarbonization technology to combat climate change while producing a valuable platform chemical and energy carrier. However, predicting the performance of the diverse catalytic systems for this process remains a challenge, as it is governed by several intricate and non-linear interactions between the catalyst composition, its structure as well as reaction mechanism, properties, conditions, and kinetics.<sup>[1]</sup> Herein, we present an explainable machine learning (ML) framework to predict catalyst performance, purely from experimental descriptors.<sup>[2]</sup> A database of 1425 datapoints comprising of Cu-, Pd-, In<sub>2</sub>O<sub>3</sub>-, and ZnO-ZrO<sub>2</sub>-based catalysts is compiled from literature and subjected to data mining (**Fig. 1a**). The methanol space-time yield (STY) is then predicted based on a set of 12 descriptors including catalyst compositions, synthesis procedure and reaction conditions using accurate ensemble-tree models (**Fig. 1b**). The greater significance of space velocity, pressure, and metal content on the model prediction is revealed *via* causal analytics (**Fig. 1c**), which accounts for 60% of the model variance. Eventually the insights of the explainable ML model are experimentally validated, with a root mean squared error of 0.11 g<sub>MeOH</sub> h<sup>-1</sup> g<sub>cat</sub><sup>-1</sup> between the experimental and model predicted methanol STY (**Fig. 1d**). The framework is purely data-driven and cross-deployable to other catalytic processes.



**Fig. 1a** Data mining representing selectivity towards methanol as a function of CO<sub>2</sub> conversion for all the catalysts compiled from literature. **b** Model prediction by extreme gradient boost algorithm on the training and test set. **c** Feature importance analysis represented in the form of horizontal bar plots. *GHSV* (gas hourly space velocity, cm<sup>3</sup> h<sup>-1</sup> g<sub>cat</sub><sup>-1</sup>), *P* (pressure, MPa), *M* (metal content, wt.%), *T* (temperature, K), *PR1* (promoter 1, wt.%), *S<sub>BET</sub>* (catalyst surface area, m<sup>2</sup> g<sup>-1</sup>), *S* (support, wt.%), feed H<sub>2</sub>/CO<sub>2</sub> ratio (H<sub>2</sub>/CO<sub>2</sub>, -), *CT* (calcination temperature, K) and *CD* (calcination time, h). **d** Experimental validation of the model based for the four different catalyst families on 60 experimental runs

[1] J. Zhong, X. Yang, Z. Wu, B. Liang, Y. Huang, T. Zhang, *Chem. Soc. Rev.* **2020**, *49*, 1385.

[2] M. Suvarna, T.P. Araújo, J. Pérez-Ramírez, *Appl. Catal. B Environ.* **2022**, in press.

## Chemoton: Automated Exploration of Reaction Networks

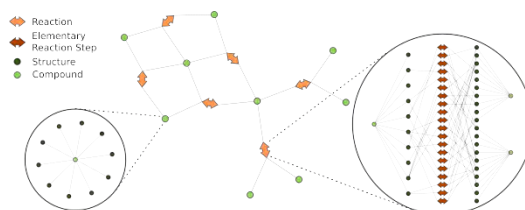
J. P. Unsleber<sup>1</sup>, S. A. Grimmel<sup>1</sup>, M. Reiher<sup>1\*</sup>

<sup>1</sup>Laboratorium für Physikalische Chemie, ETH Zürich

In recent years, much algorithm development has been devoted to the computer-based automated exploration of chemical reaction networks.<sup>[1]</sup> In parts, this has been driven by its role in the automation and acceleration of Design-Make-Test-Analyze (DMTA) cycles for new materials and chemical processes. All of these reaction network based methods aim at establishing a sufficiently complete reaction network in order to allow for reliable predictions, detailed understanding, and eventually, the design of chemical processes.

A complete reaction network consists of all thermodynamically and kinetically relevant intermediates (under reaction conditions) and all reaction paths that connect them. Already due to conformational diversity, a reaction network generated for relatively simple chemical processes can be vast if it shall be complete. This poses a set of challenges for methodology and software, which we have analyzed and discussed recently.<sup>[2]</sup>

Here, we will present the current developments of our mechanism exploration automaton, called Chemoton<sup>[3]</sup>, in its latest release version 2.0<sup>[4,5]</sup> and beyond.



1. G. N. Simm, A. C. Vaucher, and M. Reiher, *J. Phys. Chem. A*, **2018**, 123, 385.
2. J. P. Unsleber and M. Reiher, *Annu. Rev. Phys. Chem.*, **2020**, 71, 121.
3. G. N. Simm and M. Reiher, *J. Chem. Theory Comput.*, **2017**, 13, 6108.
4. J. P. Unsleber, S. A. Grimmel and M. Reiher, *arXiv*, **2022**, DOI: 10.48550/arXiv.2202.13011.
5. Website: <https://scine.ethz.ch>, Source Code: <https://github.com/qcscine>.

**The Transferability Limits of Static Benchmarks**

T. Weymuth<sup>1</sup>, M. Reiher<sup>1\*</sup>

<sup>1</sup>Laboratory of Physical Chemistry, ETH Zurich

All practical quantum chemical methods introduce certain approximations, which lead to a systematic error. For computational chemistry applications, quantifying this error for the specific case under consideration is very important. The traditional way of establishing the reliability of a method is what we call "static" benchmarking, i.e., one compares the results for a predefined, fixed set of molecules to some reference data. However, as we will show, the transferability and reliability of such static benchmark results remains always questionable to a certain degree. We advocate to instead adopt a "dynamic", i.e., rolling and system-focused approach for efficiently quantifying quantum chemical uncertainties.

[1] Thomas Weymuth, Markus Reiher, arXiv:2204.06659 [physics.comp-ph].

## Capturing tunnelling and wavepacket splitting effects on electronic spectra with Hagedorn wavepackets

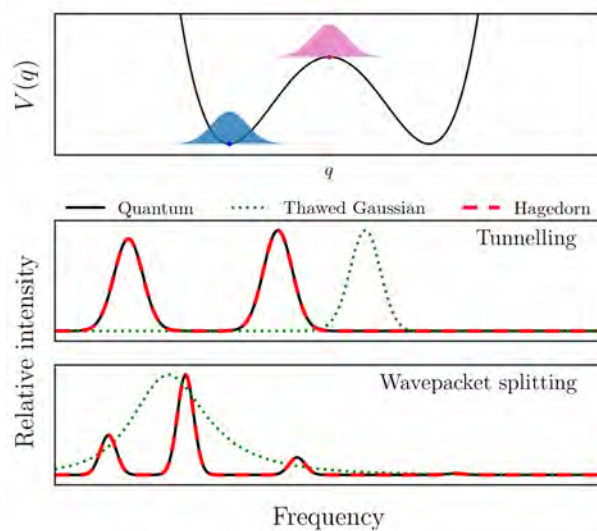
Z. Zhang<sup>1</sup>, J. Vaníček<sup>1\*</sup>

<sup>1</sup>Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Semiclassical methods such as the thawed Gaussian approximation (TGA) have been used to efficiently simulate vibrationally resolved electronic spectroscopy in order to better understand experimental results and guide further investigations into molecular processes. In TGA, a single Gaussian wavepacket (GWP) is propagated using classical-like equations under the local harmonic approximation (LHA), which keeps the wavepacket in the Gaussian form.

To explore dynamics where a single GWP does not suffice, we can construct a class of semiclassical wavepackets introduced by Hagedorn from a complete orthonormal basis in the form of polynomials times a particular GWP. The time evolution of the GWP parameters associated with such wavepackets retains the simple propagation form as in TGA, while a variational method is applied to the coefficients of basis functions to account for effects beyond LHA. As a result, the Hagedorn wavepackets (HWP) can be applied to much higher-dimensional systems than grid-based quantum approaches.

On several examples of double-well and anharmonic potentials, we demonstrate the convergence of the HWP approach to exact quantum results with a sufficient number of basis functions and its ability to treat spectroscopy problems involving quantum tunnelling and wavepacket splitting (see figure) that commonly arise in chemical dynamics but are beyond the reach of a single thawed GWP.



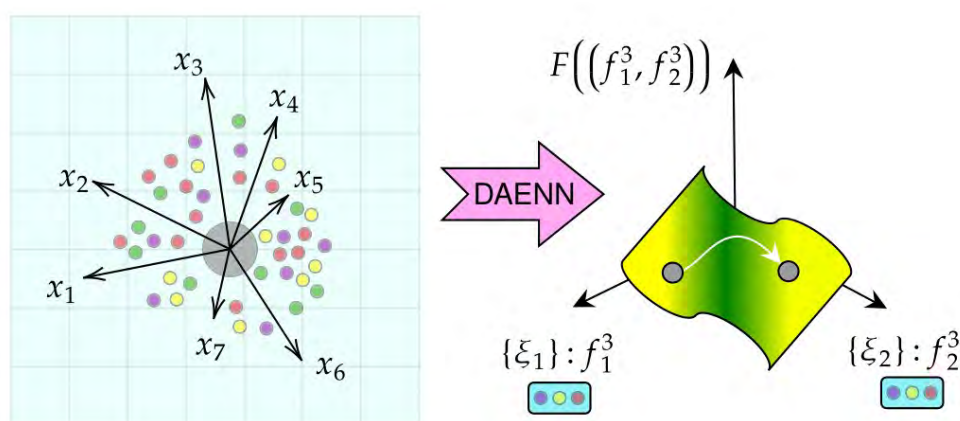
- [1] E. Heller, *Journal of Chemical Physics*, **1975**, 62, 1544-1555
- [2] M. Wehrle, S. Oberli, J. Vaníček, *Journal of Physical Chemistry A*, **2015**, 119, 5685-5690
- [3] G. Hagedorn, *Annals of Physics*, **1998**, 269, 77-104
- [4] E. Faou, V. Gradinaru, C. Lubich, *SIAM Journal on Scientific Computing*, **2009**, 31, 3027-3041
- [5] V. Gradinaru, G. A. Hagedorn, A. Joye, *Journal of Chemical Physics*, **2010**, 132, 184108
- [6] T. Begušić, M. Cordova, J. Vaníček, *Journal of Chemical Physics*, **2019**, 150, 154117

## Generalizing Collective Variables using Deep Learning for Free Energy Surface Calculation

R. Ketkaew<sup>1</sup>, F. Creazzo<sup>1</sup>, S. Luber<sup>1\*</sup>

<sup>1</sup>University of Zurich

Enhanced sampling simulation techniques have been widely used for a broad range of chemical reactions and biological systems. The ability of those techniques such as metadynamics essentially depends on the quality of a set of chosen atomic coordinates (collective variables (CVs)) that has to be defined before performing the simulations. In this work, we present an efficient *Deep AutoEncoder Neural Network* (DAENN) algorithm for effortlessly computing general-purpose CVs that can be used with enhanced sampling methods to reconstruct free energy surfaces of chemical reactions. Carefully manifested by primitive organic reactions such as the Diels-Alder reaction, DAENN-generated CVs are capable of identifying transitions between rare-event metastable states. The results show that the adopted CVs give accurate computed free energy that is in line with those obtained by previously developed CVs and experimental results. In addition, the proposed method is implemented in *Deep learning for Collective Variables* (DeepCV), a computer code that provides an efficient and customizable implementation of the DAENN, which can communicate with other molecular dynamics software, namely CP2K and PLUMED.



[1] W. Chen, A. R. Tan, A. L. Ferguson. *J. Chem. Phys.* **2018**, 149, 072312

[2] H. Sidky, W. Chen, A. L. Ferguson. *Mol. Phys.* **2020**, 118, 5, e1737742

[3] R. Ketkaew, F. Creazzo, S. Luber. *J. Phys. Chem. Lett.* **2022**, 13, 1797-1805



**Oxygen Spin-crossover Reaction with Instanton Theory**I. M. Ansari<sup>1</sup>, J. Richardson<sup>1\*</sup><sup>1</sup>ETH Zurich, D-CHAB, Vladimir-Prelog-Weg 1-5/10, 8093 Zurich, Switzerland

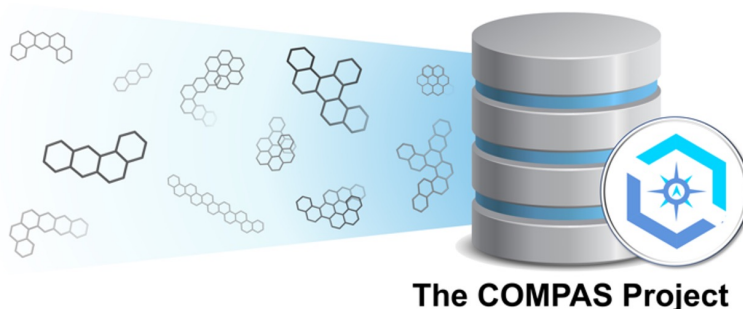
The first excited state of the oxygen molecule is a chemically-reactive singlet that plays a key role in various processes of biological and chemical significance including cellular damage, cancer treatment, polymer synthesis and water disinfection. As an excited-state reactive intermediate, its reactivity is significantly affected by its lifetime, which can vary based on the solvent environment and temperature. The non-radiative spin-crossover deactivation of O<sub>2</sub> from the singlet <sup>1</sup>Δ<sub>g</sub> to the <sup>3</sup>Σ<sub>g</sub><sup>-</sup> ground state is therefore crucial to understanding, both qualitatively and quantitatively, its reactivity. This is a reaction that takes place in the Marcus inverted regime and is therefore more likely to exhibit tunnelling, which we treat in an ab-initio manner using golden-rule instanton theory. However, due to the presence of a singularity in the flux-flux correlation function, it cannot be directly applied to this system. We present a modified version of golden-rule instanton theory that takes into account this singularity and use it to calculate the rate of the deactivation process in water, by considering a 1:1 complex. Additional calculations on isotope-substituted species reveal insights into the mechanism of deactivation.

## The COMPAS Project - a COMputational Database of Polycyclic Aromatic Systems. Phase 1: cata-condensed Polybenzenoid Hydrocarbons

A. Wahab<sup>1</sup>, L. Pfuderer<sup>1</sup>, E. Paenurk<sup>1</sup>, R. Gershoni-Poranne<sup>1,2</sup>, P. Chen<sup>1\*</sup>

<sup>1</sup>ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 1-5/10, 8093 Zürich, <sup>2</sup>Schulich Faculty of Chemistry, Technion-Israel Institute of Technology, Haifa 32000

Chemical databases are an essential tool for data-driven investigation of structure–property relationships and for the design of novel functional compounds. Polycyclic aromatic systems are one of most prevalent types of compounds in nature and are important in many fields in chemistry. To the best of our knowledge, no well-curated and decently sized database of those molecules exist to this day. For this purpose, we started the COMPAS project: a COMputational database of Polycyclic Aromatic Systems. We present here<sup>1</sup> the first phase of this project, where we focused on *cata*-condensed polybenzenoid hydrocarbons (PBHs). We developed two data sets: (1) COMPAS-1x containing ~34k molecules with up to 11 rings, and (2) COMPAS-1D containing ~9k molecules with up to 10 rings. Both datasets contain the optimized ground-state structures and a selection of molecular properties, calculated at the GFN2-xTB and the B3LYP-D3BJ/def2-SVP levels, respectively. Herein, we describe the workflow of data generation and curation, and detail the information available within the data sets. Using those data sets, we compare the two types of computations as well as study the structure–property relationships of the calculated species. The data and insights gained from them can drive rational design of novel functional polycyclic aromatic molecules with applications in, e.g., organic electronics, and can provide a basis for additional data-driven machine- and deep-learning studies in chemistry.



[1] Wahab, A.; Pfuderer, L.; Paenurk, E.; Gershoni-Poranne, R. The COMPAS Project: A Computational Database of Polycyclic Aromatic Systems. Phase 1: Cata-Condensed Polybenzenoid Hydrocarbons. *J. Chem. Inf. Model.* 2022. <https://doi.org/10.1021/acs.jcim.2c00503>.