CHIMIA

CHIMIA 2023, Volume 77
ISSN 0009-4293
www.chimia.ch
Supplementa to Issue 7-8/2023

SCS
Swiss Chemical
Society

SCS Fall Meeting 2023
Lecture, Short Talk and Poster Abstracts

Session of Computational Chemistry

August 24-25, 2023
University of Bern, vonRoll Areal
https://fm23.scg.ch

Swiss Chemical Society
Haus der Akademien
Postfach
3001 Bern
Switzerland
info@scg.ch
www.scg.ch

Capturing tunnelling, wavepacket splitting and single vibronic level excitation effects on vibronic spectra with Hagedorn wavepackets

Z. Zhang[1], J. Vaníček[1]*

[1]Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne

Vibrationally resolved electronic spectroscopy provides important information on the structure and dynamics of polyatomic molecules. On-the-fly ab initio implementation [1,2,3] of the thawed Gaussian approximation (TGA) [4] has been successful in describing such spectra within the Condon approximation but cannot capture non-Condon effects, tunneling or wavepacket splitting. While the extended TGA [5,6] captures first-order non-Condon effects, Hagedorn's wavepackets [7,8], i.e. superposition of states, each of which is the Gaussian multiplied by a cleverly chosen polynomial, enable the propagation of non-Condon wavepackets of arbitrary shapes. I will demonstrate that a variational implementation [9] of the Hagedorn wavepackets can capture tunneling and wavepacket splitting, as well as their effects on spectra. To demonstrate that the method is not restricted to small model systems, I will combine our recently developed algorithm for an algebraic evaluation of the overlap of Hagedorn wavepackets with on-the-fly ab initio semiclassical dynamics to compute single vibronic level fluorescence spectra of polyatomic molecules.

[1] M. Wehrle, M. Šulc, J. Vaníček, J. Chem. Phys., **140**, 244114 (2014)
[2] M. Wehrle, S. Oberli, J. Vaníček, J. Phys. Chem. A, **119**, 5685-5690 (2015)
[3] T. Begušić, E. Tapavicza, J. Vaníček, J. Chem. Theory Comput., **18**, 3065-3074 (2022)
[4] E. Heller, J. Chem. Phys, **62**, 1544-1555 (1975)
[5] A. Patoz, T. Begušić, J. Vaníček, J. Phys. Chem. Lett., **9**, 2367-2372 (2018)
[6] A. Prlj, T. Begušić, Z. Zhang, G. Fish, M. Wehrle, T. Zimmermann, S. Choi, J. Roulet, J. Moser, J. Vaníček, J. Chem. Theory Comput., **16**, 2617-2626 (2020)
[7] G. Hagedorn, Annals of Physics, **269**, 77-104 (1998)
[8] C. Lasser, C. Lubich, Acta Numerica, **29**, 229-401 (2020)
[9] E. Faou, V. Gradinaru, C. Lubich, SIAM J. Sci. Comput., **31**, 3027-3041 (2009)

A local diabatisation method for two-state adiabatic conical intersections

E. Vandaele[1], M. Mališ[1], S. Luber[1]*

[1]Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

The topography around a conical intersection (CI) regulates the non-adiabatic transitions, thereby governing the deactivation pathway, products and lifetime of non-radiative decay processes.[1-3] Hence, a detailed depiction of a CI provides valuable information on a molecule's photodeactivation. A new methodology to characterise CIs between two adiabatic electronic states will be presented, from which the non-adiabatic coupling vectors (NACs) can be calculated in a wave function-free, energy-based approach.[4] The branching space coordinates are identified only based on the state's Hessians and gradients at the CI geometry. NAC terms are constructed after converting the adiabatic potential energy surface (PES) to a diabatic representation by fitting the PES around the CI in the branching space, as shown in Figure 1.

To prove the universality of the developed methodology, the minimum-energy CI (MECI) between the first (S1) and second (S2) singlet excited states of formamide is investigated at the state-averaged complete active space self-consistent field (SA-CASSCF) and extended multi-state complete active space second-order perturbation theory (XMS-CASPT2) levels of theory. In addition, the asymmetrical MECI between the ground state (S0) and S1 of cyclopropanone is evaluated using SA-CASSCF, as well as (ME)CIs between the S1 and S2 states of benzene using SA-CASSCF and time-dependent density functional theory (TDDFT). Finally, a CI between the S1 and S2 excited states of thiophene was analysed using TDDFT. The evolution of the states' characters, and projection of the NACs and gradients on the branching plane computed ab-initio and composed from the developed methodology are analogue for all electronic structure methods and software programmes used.

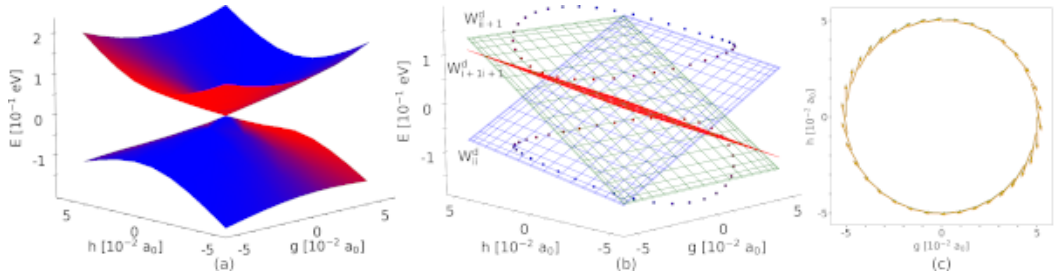


Figure 1: SA-CASSCF(8,7) PES around the S0-S1 asymmetrical MECI of cyclopropanone along the branching space coordinates (a) and diabatic approximation of this PES with circumferences of the adiabatic energy cones (b) coloured according to the state characters. Projection of the NAC vectors calculated using SA-CASSCF(8,7) (green) and composed from our methodology (orange) on the branching space coordinates in the vicinity of cyclopropanone's S0-S1 asymmetrical MECI (c).

[1] Wolfgang Domcke, David R. Yarkony, *Annual review of physical chemistry*, **2012,** 63, 325-352.
[2] Lluis Blancafort, *ChemPhysChem*, **2014**, 15(15), 3166-3181.
[3] Spiridoula Matsika, *Chemical Reviews*, **2021**, 121(15), 9407-9449.
[4] Eva Vandaele, Momir Malis, Sandra Luber, *Manuscript submitted for publication*.

A DMRG-based Framework for Large-scale Quantum Many-body Calculations

N. Glaser[1], A. Baiardi[1], M. Reiher[1]*

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

Quantum chemical algorithms based on tensor factorizations are continuously expanding the scope of wave function-based molecular simulation methods. By leveraging very compact many-body wave function parametrizations, tensor network-based methods such as the density matrix renormalization group (DMRG) [1] can tame the computational cost of full configuration interaction (CI)-type calculations for strongly correlated molecular systems. While the DMRG algorithm is routinely applied to ground-state electronic structure problems, we present a versatile framework enabling its application to a broader range of many-body quantum problems [2,3]. We demonstrate the capabilities of our DMRG-based framework by introducing the *n*-mode vibrational DMRG method [4], which allows for an accurate treatment of the correlated nuclear problem for a reliable characterization of strongly anharmonic molecules described by complex potential energy surfaces. By expressing the vibrational CI wave function as a matrix product state, *n*-mode vDMRG can target systems with up to 30 fully coupled vibrational modes.

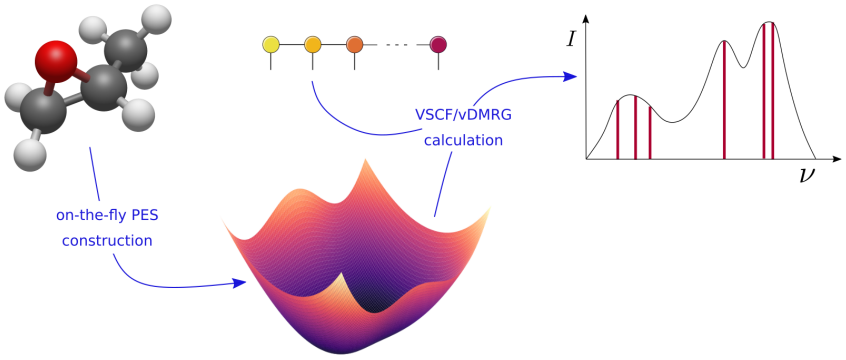


Figure 1: Illustration of the computational workflow of a *n*-mode vDMRG calculation

We further extend our framework to excited-states algorithms [5] enabling the large-scale calculation of both
low- and high-energy excitations. The combination of powerful tensor-based wave function representations
and efficient eigenstate targeting algorithms paves the route towards the accurate characterization of a
large variety of high-dimensional quantum systems.

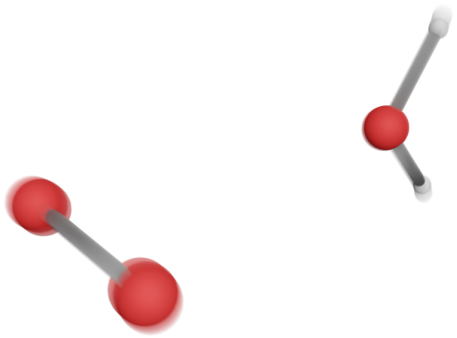
[1] S. R. White, *Phys. Rev. Lett.*, **1992**, 69, 2863–2866.
[2] A. Baiardi, C. J. Stein, V. Barone, M. Reiher, *J. Chem. Theory Comput.*, **2017,** 13, 3764–3777.
[3] N. Glaser, A. Baiardi, M. Reiher. In *Vibrational Dynamics of Molecules,* World Scientific, **2022,** 80-144.
[4] N. Glaser, A. Baiardi, M. Reiher, *manuscript in preparation*
[5] A. Baiardi, A. K. Kelemen, M. Reiher, *J. Chem. Theory Comput.*, **2022,** 18, 415–430.

Heavy-atom tunneling in singlet oxygen deactivation in water

I. M. Ansari[1], E. R. Heller[1], G. Trenins[1], J. O. Richardson[1]*

[1]Department of Chemistry and Applied Biosciences

The singlet state of the oxygen molecule is an important reactive precursor[1] that has applications in biochemistry, synthetic, atmospheric and environmental chemistry. The lifetime of singlet oxygen in solution is principally determined by the rate of its nonradiative decay to the triplet ground state, which is a spin-crossover process (also known as intersystem crossing) in the Marcus inverted regime. The rate constant for the decay can in principle be calculated using Fermi's Golden Rule (FGR), but as this requires solutions to the exact nuclear Schrödinger equation, it is computationally infeasible. Previous attempts at approximating the rate have not accounted for nuclear quantum effects such as tunnelling and zero-point energy[2]. Golden-rule instanton theory is a semiclassical approximation to FGR that captures multidimensional tunnelling and zero-point energy effects by locating the optimal tunnelling path, called the instanton[3]. However, the flux autocorrelation function for this spin-crossover reaction exhibits a singularity and golden-rule instanton theory can therefore not be applied. We present extensions to golden-rule instanton theory that tackle such cases and apply it to calculate the decay rate constant of singlet oxygen interacting with a water molecule using on-the-fly multireference ab initio calculations[4]. The new instanton methods predict a reaction mechanism with significant heavy-atom tunnelling, which corresponds to a 27 order of magnitude tunnelling factor and a large kinetic isotope effect of 20 at room temperature, consistent with experimental data.



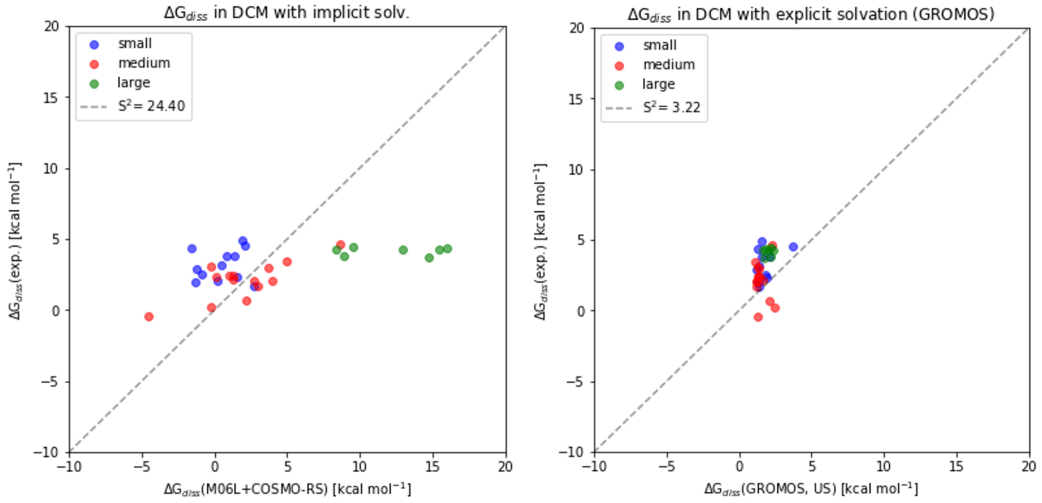
[1] Maria C. DeRosa, Robert J. Crutchley, *Coordination Chemistry Reviews*, **2002**, 233–234, 351-371.
[2] Frederik Thorning, Frank Jensen, Peter R. Ogilby,The Journal of Physical Chemistry B, **2020,** 124 (11), 2245-2254.
[3] Imaad M. Ansari, Eric R. Heller, George Trenins, Jeremy O. Richardson, *Phil. Trans. R. Soc.*, **2022,** A 380: 20200378.
[4] Imaad M. Ansari, Eric R. Heller, George Trenins, Jeremy O. Richardson, in preparation.

Solvation Free Energies of Ion Dissociations in Dichloromethane: En Route to Accurate Computations

A. Savoy[1], E. Paenurk [2], R. Pollice [3], P. H. Hünenberger[4], P. Chen[1]*

[1]Laboratorium für Organische Chemie, ETH Zurich, Vladimir-Prelog-Weg 2, CH-8093 Zurich, Switzerland, [2]Department of Chemistry Yale University, 225 Prospect Street, New Haven, CT 06520-8107, United States, [3]Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, Netherlands, [4] Laboratorium für Physikalische Chemie, ETH Zurich, Vladimir-Prelog-Weg 2, CH-8093 Zurich, Switzerland

An extensive dataset of proton bound pyridine dimers was previously synthesized. An experimental and computational study of their dissociation, in the gas phase and in condensed phase, was previously reported [1][2]. Comparison of gas phase and condensed phase data gave considerable insights into solvation effects and the compensation of London dispersion interactions in condensed phase. State of the art DFT calculations and implicit solvent models failed spectacularly to reproduce free energies of dissociation in solvent, especially for larger quinoline dimers. We report a complementary computational study on the dissociation of proton-bound pyridine dimers in the gas phase and in dichloromethane (DCM). In an effort to determine the prerequisites to reproduce experimental trends and magnitudes of free energy of dissociation (ΔG*diss*) in solvent, different contributions were investigated: The configurational and translational entropic contribution were estimated using xTB. Explicit solvation with classical mechanics was used as the most relevant method to treat not only the solute but also the solvent configurational entropy. This approach is a crude but fast computational method capable of predicting ΔG*diss* in DCM for this test system within a reasonable error margin. Despite the encouraging results in DCM, there are some limitations to the prediction of the free energy of solvation (ΔG*solv*) which were unveiled and discussed extensively.



[1] R. Pollice, M. Bot, I. J. Kobylianskii, I. Shenderovich, P. Chen, *J. Am. Chem. Soc.*, **2017**, 139, 37, 13126–13140 [2] R. Pollice, F. Fleckenstein, I. Shenderovich, P. Chen, *Angew. Chem. Int. Ed.*, **2019**, 58, 40, 14281-14288

Computational chemistry in the life sciences industries

A. Feldman-Salit[1]

[1]Arcondis Group, Hegenheimermattweg 167b, 4123 Allschwil - anna.feldman-salit@arcondis.com

Computational chemistry is not only a tool to crack a scientific drug-design riddle but also an essential contributor to productive manufacturing for life sciences industries.

In our talk we want to address the advantages of computational chemistry field for drug or chemical production processes as the ultimate future vision of digitalized and automated manufacturing era. We will also walk you through known challenges and corresponding technical or administrative solutions, which we witnessed during our 20-year-old collaborative work with various pharmaceutical, chemical, and biotechnological companies.

It will be shown, how e.g., good project management or technical advisory and support may facilitate more efficient use of the computational predictive tool to reduce workload in the wet labs, to accelerate generated data analysis and transfer, thereby improving a production performance, also increasing patient safety and compliance of final products.

We will demonstrate how essential and valuable a contribution of computational predictions in the life sciences industry is, and not only due to its scientific mystery.

Grand canonical ensemble approaches for modeling electrochemical problems at constant electrode potentials in CP2K

Z. Chai[1], S. Luber[1]

[1]Department of Chemistry, University of Zurich, Zurich 8057, Switzerland

In many electrochemical experiments, the number of electrons of the electrode immersed in the electrolyte is variable, and the number of adsorbed substances on the surface of the electrode can also vary. However, treating electrochemical solid-liquid interfaces with the typical canonical DFT tends to be a challenge. This can be addressed by using grand canonical approaches. We present the implementation of two grand canonical approaches that go beyond the existing canonical ensemble paradigm in the open-source computational chemistry software CP2K. The first approach includes a number of recent developments: (a) grand canonical self-consistent field (GC-SCF) method[1] allowing the electron number of the system to fluctuate naturally and accordingly with the experimental electrode potential, (b) planar counter charge (PCC)[2,3] salt model completely screening the net charge of the electrode model, (c) the solvent-aware interfaces[4] between solute and solvent in continuum solvation for overcoming the unphysical isolated cavities or pockets of the dielectric function. In contrast with the previous studies, in our implementation, the work function (absolute electrode potential) is the constrained quantity during an SCF optimization instead of the Fermi energy. We derived the analytical expressions of the potential and force compatible with the Quickstep framework of the CP2K software package. The second approach (referred to as the two-surface method and the numerical litmus method)[5–7] is used to calculate the absolute electrode potential corresponding to an equilibrium electrochemical half-reaction (M^((n+m)+)+ne^-→M^(m+)) which involves DFT-MD and explicit modeling of the solvent molecules. The systematic tests have verified that the implementation of both two methods in CP2K is reliable. This opens the way for forefront electrochemical calculations in CP2K for a broad range of systems.

(1) Sundararaman, R. et al. **2017**, *146* (11), 114104.
(2) Nattino, F. et al. **2018**, *150* (4), 041722.
(3) Lozovoi, A. Y.; Alavi, A. **2003**, *68* (24), 245416.
(4) Andreussi, O. et al. **2019**, *15* (3), 1996–2009.
(5) Tavernelli, I. et al. **2002**, *88* (21), 213002.
(6) Tateyama, Y. et al. **2005**, *122* (23), 234505.
(7) Blumberger, J. et al. **2004**, *126* (12), 3928–3938.

Transition Metal Binding Site Predictions for Drug Design Applications

A. Levy[1], S. L. Dürr[1], U. Röthlisberger[1]*

[1]Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Transition metal-based drugs represent promising candidates for cancer treatment. In the field of computational chemistry, hybrid quantum mechanics/molecular mechanics (QM/MM) is often the method of choice for this type of system, allowing an accurate description of metal species and the formation or breaking of covalent bonds. However, QM/MM is computationally expensive, limiting its applications to drug design, where the ability to study multiple drug candidates binding to diverse sites is crucial.

In this context, we present Metal3D [1], an accurate metal site predictor recently developed in our group showing promising applications: it employs 3D convolutional neural networks and, despite being initially designed for predicting the location of zinc ion binding sites, it shows accurate predictions for other transition metal ions and, more interestingly for drug design applications, also for transition metal-based drugs. We assessed Metal3D's ability to predict binding sites for different transition metal-based compounds complexed with nucleosome core particles. Remarkably, Metal3D accurately identifies the experimentally located sites in the majority of cases.

Metal3D's ability to accurately identify binding sites for transition metal-based drugs in biological systems makes it a promising tool for drug design applications. By efficiently identifying putative binding sites, Metal3D can be a valid asset for drug design, providing a limited number of likely starting points for QM/MM simulations.

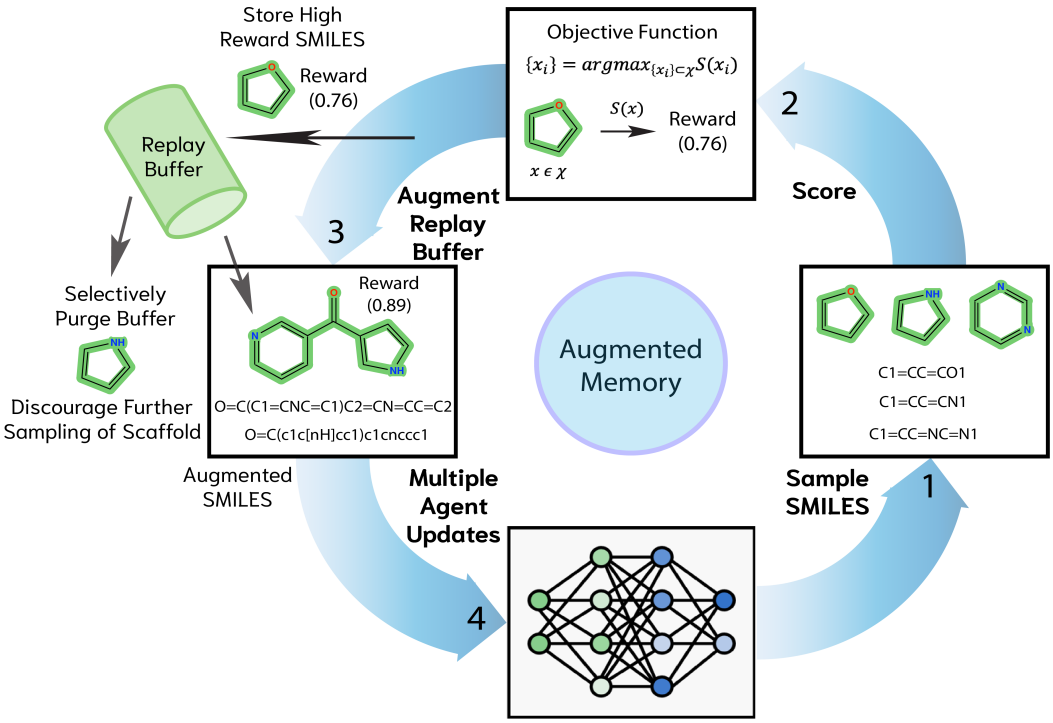
[1] Simon L. Dürr, Andrea Levy, and Ursula Rothlisberger, *Nature Communications*, **2023**, 14(1), 2713.

Augmented Memory: Capitalizing on Experience Replay to Accelerate De Novo Molecular Design

J. Guo[1,2], P. Schwaller[1,2]*

[1]Laboratory of Artificial Chemical Intelligence (LIAC), Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [2]National Centre of Competence in Research (NCCR) Catalysis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Sample efficiency is a fundamental challenge in de novo molecular design. Ideally, molecular generative models should learn to satisfy desired objectives under minimal oracle evaluations (computational prediction or wet-lab experiment). This problem becomes more apparent when using oracles that can provide increased predictive accuracy but impose a significant cost. Molecular generative models have shown remarkable sample efficiency when coupled with reinforcement learning, as demonstrated in the Practical Molecular Optimization (PMO) benchmark. Here, we propose a novel algorithm called Augmented Memory that combines data augmentation with experience replay. We show that scores obtained from oracle calls can be reused to update the model multiple times. We compare Augmented Memory to previously proposed algorithms and show significantly enhanced sample efficiency in an exploitation task and a drug discovery case study requiring both exploration and exploitation. Our method achieves a new state-of-the-art in the PMO benchmark which enforces a computational budget, and outperforms the previous best performing method on 19/23 tasks.



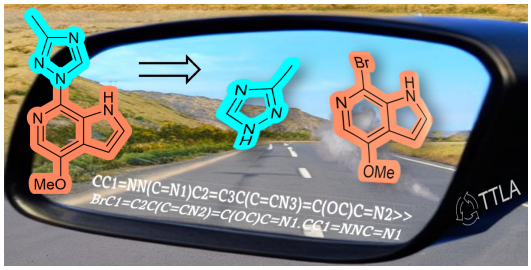
Jeff Guo, Philippe Schwaller, *ChemRxiv*, **2023**. doi: 10.26434/chemrxiv-2023-qmqmq-v3 D O I: 10.26434/chemrxiv-2023-qmqmq-v3

Disconnection-Aware Triple Transformer Loop with a Route-Penalty Score for Multistep Retrosynthesis

D. Kreutter[1], J. L. Reymond[1]*

[1]Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

Computer-aided synthesis planning (CASP) plays a crucial role in automating retrosynthetic analyses of unseen molecules by learning organic reactivity from literature. To address the challenges of (1) proposing realistic disconnections while maintaining reaction novelty and diversity, and (2) exploring efficient short synthetic sequences, we present an innovative open-source CASP tool.



Our approach uses a triple transformer loop (TTL) that separately predicts starting materials (T1), reagents (T2), and products (T3). It explores multiple disconnections sites through a combination of exhaustive, template-based, and transformer-based tagging procedures prior to T1, allowing an extensive chemical space exploration.

Furthermore, we integrate the single-step TTL into a multistep tree search algorithm (TTLA) that prioritizes sequences based on a route penalty score (RPScore). The RPScore considers factors such as the number of steps, confidence scores, and the simplicity of intermediates along the route. This scoring scheme enables TTLA to prioritize shorter synthetic routes to readily available commercial starting materials during the tree search exploration. The effectiveness of our approach is demonstrated by showcasing retrosynthetic analyses of recently approved drugs.

Overall, our open-source multistep retrosynthesis tool provides a broader chemical space exploration in synthesis planning and can predict short synthetic routes for drug-like molecules. Moreover, separating the prediction of starting material and reagents might be adapted to more complex reaction types.[1]

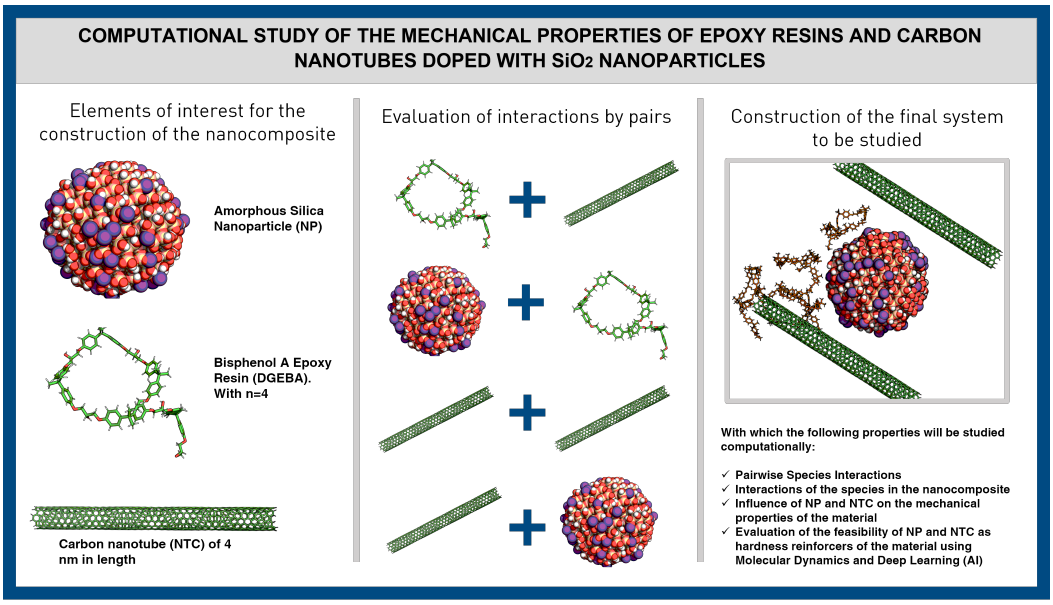
[1] D. Kreutter, J.-L. Reymond, *ChemRxiv*, **2023**, DOI: 10.26434/chemrxiv-2022-8khth-v2

Computational Study of the Mechanical Properties of Epoxy Resins and Carbon Nanotubes Doped with SiO_2 Nanoparticles

A. J. Bastos R.[1], J. G. Parra[2], R. Hernández[3], J. Castillo[1], V. Mujica[4]

[1]Central University of Venezuela, Chemistry School, Computational Chemistry department, Venezuela., [2]University of Carabobo, Physical Chemistry, Venezuela., [3] Industrial Research and Development Center (CIDESI), Strategic and technology postgraduate management, Mexico., [4] Arizona State University, Center for Biological Physics, United States.

Composite materials have been widely used due to their high mechanical properties and great resistance, generally proven in the aeronautical industry. However, it is sought to place reinforcements of other types of particles to be useful in applications of interest such as medicine, specifically for the design of materials for a foot prosthesis. It is for this reason, that in this work a study of the mechanical properties of polymeric nanocomposites was carried out from computational techniques of semiclassical mechanics and neural networks. The GROMACS 2019.2 software and the CHARMM36 force field were used to perform molecular dynamics simulations in polymer nanocomposite systems with different polymer chain lengths and number of carbon nanotubes. In addition, a Neural Network was used to predict the mechanical properties of the nanocomposites. The most relevant results of this study include the determination of the relative energies between the studied species, which allowed us to understand at the molecular level the effectiveness of the nanocomposite assembly as a fundamental reason for the improvement that the mechanical properties could have. In addition, it was possible to obtain inferences regarding the proportions of the three studied polymer sizes suggested by the neural network, which allowed predicting the mechanical properties of the nanocomposites with an accuracy of less than 10%. Finally, it was possible to study the percentage change in the Poissons ratio according to the distortions suffered by the lengths parallel to the direction of application of the external pressure, which made it possible to compare the responses of the different systems as the size of the pipe increased. polymer or the number of Carbon Nanotubes was increased. In conclusion, this study demonstrates the utility of molecular dynamics simulations and neural networks to predict the mechanical properties of polymer nanocomposites and improve their use in industrial or scientific applications.



Molecular Hypergraph Neural Network

J. Chen[1]

[1]Laboratory of Artificial Chemical Intelligence (LIAC) & National Centre of Competence in Research (NCCR) Catalysis, Institute of Chemical Sciences and Engineering, EPFL

Graph neural networks (GNNs) have exhibited promising performance on many chemistry-related tasks. However, ordinary graphs only model the pairwise connectivities in molecules, can’t adequately describe higher-order connectivities like multi-center bonds and conjugated bonds. To address this challenge, we introduce molecular hypergraphs and propose Molecular Hypergraph Neural Network (MHNN) to predict the optoelectronic properties of organic semiconductors. MHNN outperforms all baseline models on the most tasks of organic photovoltaic (OPV) dataset. Without any 3D geometric information, MHNN also surpasses the baseline models using atom positions. Moreover, MHNN achieves better performance than pretrained GNNs under limited training data, demonstrating its excellent data efficiency.

A machine learning-based QSAR approach to predict biological removal of organic micropollutants during wastewater treatment

J. A. Cordero[1], K. Fenner[1,2]*

[1]Department of Environmental Chemistry, Eawag: Swiss Federal Institute of Aquatic Science and Technology, [2]Department of Chemistry, University of Zürich

Accurate prediction of removal of micropollutants during wastewater treatment is urgently needed for improved chemical risk assessment and for the design of greener chemical substances and processes. One of the major challenges hindering development of accurate models is the lack of large homogenous databases. However, experimental data on removals of hundreds of micropollutants for several wastewater treatment plants (WWTPs) have been recently collected, thus providing new opportunities for modeling.

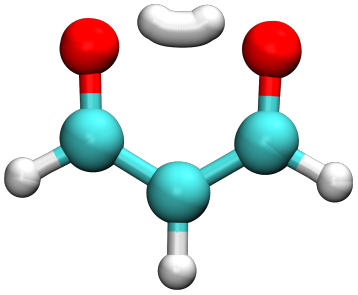
In this study, a first attempt to predict removals using molecular descriptors (i.e., Padel descriptors, MACCS fingerprints and enviPath biotransformation rules) and machine learning algorithms (i.e., Random Forests, Support Vector Machines and Neural Networks) is presented. Two independent datasets from two major sampling campaigns that collected removal data from WWTPs in Switzerland, Australia and Sweden were used. The first dataset (hereafter referred as AUS) includes data for 293 chemical compounds in 15 WWTPs and the second dataset (hereafter referred as AMAR) includes 384 compounds for 8 WWTPs. The AUS dataset is characterized by compounds with small breakthroughs ranging mostly between 0 and 0.2, that is, compounds that are largely removed. Differently, the AMAR dataset contains more compounds with breakthrough in the range of 0.2-0.4, but only few compounds with intermediate and large breakthroughs, challenging model training. Upon log transformation of breakthrough, it was possible to produce models that explain, at least partially, breakthrough in terms of the presence or absence of molecular substructures. The best performance (R^2_{test}= 0.1-0.4) was achieved using a random forest regressor and MACCS fingerprints as features. All models show large differences in performance with different random train-test splits, which is attributed to the large chemical space covered by the test set and the limited number of training examples. Therefore, we suggest the following strategies for further development: 1) applying transfer learning techniques using lab-generated data but relying on WWTP data to train the final model; 2) targeting more recalcitrant compounds through exhaustive suspect screening. Our models are available at: https://c4science.ch/source/pepper/repository/

Perturbatively corrected ring-polymer instanton theory for accurate tunneling splittings

J. Dušek[1], J. E. Lawrence[1], J. O. Richardson[1]*

[1]Department of Chemistry and Applied Biosciences

We introduce an approach for calculating perturbative corrections to the ring-polymer instanton approximation to tunneling splittings (RPI+PC), by computing higher-order terms in the asymptotic expansion in $\hbar$. The resulting method goes beyond standard instanton theory by using information on the third and fourth derivatives of the potential along the tunneling path to include additional anharmonic effects. This leads to significant improvements both in systems with low barriers and in systems with anharmonic modes. We demonstrate the applicability of RPI+PC to molecular systems by computing the tunneling splitting in full-dimensional malonaldehyde and a deuterated derivative. Comparing to both experiment and recent quantum-mechanical benchmark results, we find that our perturbative correction reduces the error from −11% to 2% for hydrogen transfer and performs even better for the deuterated case. This makes our approach more accurate than previous calculations using diffusion Monte Carlo and path-integral molecular dynamics, while being more computationally efficient.



Joseph E. Lawrence, Jindrich Dusek, Jeremy O. Richardson, arXiv (to be published in JChemPhys), 2023, arXiv:2304.10963

Insights into the nature of host-guest interactions in emergent framework materials

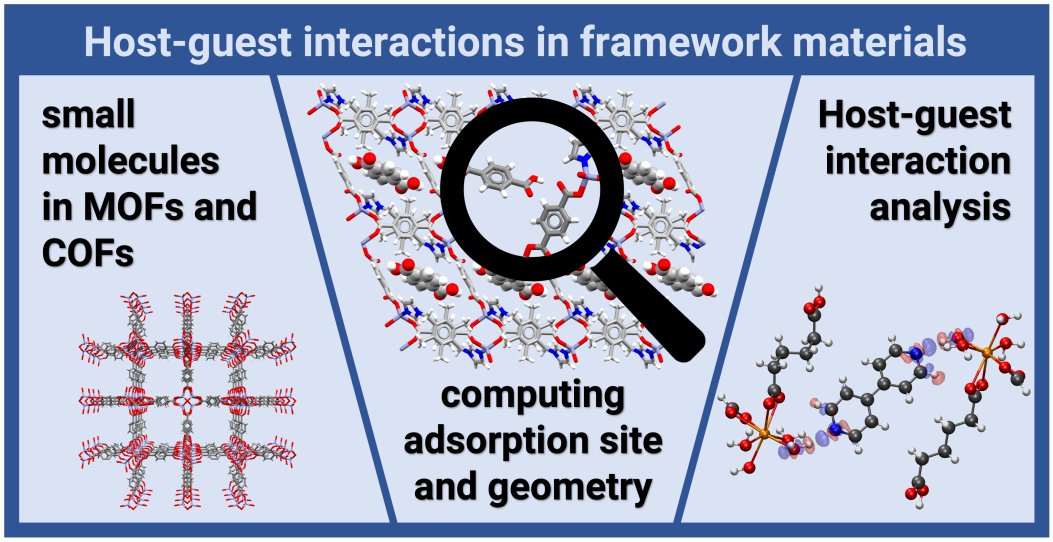
M. Ernst[1], G. Gryn'ova[2,3]*

[1]Department of Chemistry, University of Zurich, [2]Heidelberg Institute for Theoretical Studies (HITS gGmbH), [3]Interdisciplinary Center for Scientific Computing, Heidelberg University

A key feature of metal-organic frameworks (MOFs) and covalent organic frameworks (COFs) is their ability to capture, transport, and release guest molecules. The nature, quality, and quantity of the associated absorption depend on pore size and volume, surface area, chemical environment, and in particular on the host-guest intermolecular interactions.

We use *in silico* tools to study and characterize these interactions: molecular docking to identify adsorption sites, periodic and finite DFT simulations to compute interaction energies, and in-depth analyses of the non-covalent interactions between host frameworks and guests to characterize them. Based on several different examples and application cases, we show how to identify, quantify, and describe these host-guest interactions and how to relate our results to experimental data.

Eventually, these insights pave the way towards a set of design guidelines to tune the host-guest interactions for future targeted applications.

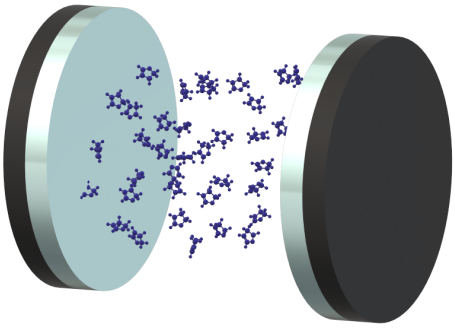


Mirrors and reaction rates: how 'quantum' is vibrational polariton chemistry?

M. R. Fiechter[1], J. E. Runeson[2], J. E. Lawrence[1], J. O. Richardson[1]*

[1]Department of Chemistry and Applied Biosciences, ETH Zürich, [2]Department of Chemistry, University of Oxford, Physical and Theoretical Chemistry Laboratory

A recent series of experiments [1-4] has shown that certain chemical reactions can be sped up or slowed down just by placing the reaction mixture between a pair of mirrors (in a so-called 'optical cavity'). The origins of this surprising phenomenon are still poorly understood. Simple transition state theory maintains that an optical cavity cannot affect reaction rates. More advanced theories do predict a cavity effect, but fail to replicate the experimental finding that rates only change if the cavity is tuned ‘on resonance’, i.e. if one of the standing light waves inside the cavity closely matches a molecular vibrational frequency. It is only recently that this resonant effect emerged for the first time in a computational study [5], which involved a full quantum-dynamical simulation of a model system. To better understand the underlying mechanism, we studied the same model system using an approximate method (ring-polymer molecular dynamics, or RPMD) that captures quantum statistical effects, such as zero-point energy and tunneling, but treats dynamics classically. In general this approach accurately predicts rate constants for reactions in solution, while remaining computationally tractable for atomistic simulations. Whether or not RPMD can reproduce the resonance effects will help us understand how an optical cavity alters chemical reactivity. This has important practical implications for future work in the field, since many affordable but approximate methods would be disqualified if the cavity effect turned out to be fully quantum dynamical in nature.



[1] Anoop Thomas, Jino George, Atef Shalabney, Marian Dryzhakov, Sreejith J. Varma, Joseph Moran, ..., Thomas W. Ebbesen, *Angewandte Chemie International Edition,* **2016**, 55, 11462-11466.
[2] Anoop Thomas, Lucas Lethuillier-Karl, Kalaivanan Nagarajan, Robrecht M.A. Vergauwe, Jino George, Thibault Chervy, ..., Thomas W. Ebbesen, *Science,* **2019**, 363, 615-619.
[3] Kenji Hirai, Rie Takeda, James A. Hutchison, Hiroshi Uji-i, *Angewandte Chemie International Edition,* **2020**, 59, 5332-5335.
[4] Abhijit Sau, Kalaivanan Nagarajan, Bianca Patrahau, Lucas Lethuillier-Karl, Robrecht M.A. Vergauwe, Anoop Thomas, Joseph Moran, Cyriaque Genet, Thomas W. Ebbesen. *Angewandte Chemie International Edition,* **2021**, 60, 5712-5717.
[5] Lachlan P. Lindoy, Arkajit Mandal, David R. Reichman*, Nature Communications,* **2023**, 14, 2733.

Redox-Based Defect Detection in Packed DNA: Insights from Hybrid Quantum Mechanical/Molecular Mechanics Molecular Dynamics and Feature Selection Studies

S. K. Johnson[1], M. Kılıç[1], P. Diamantis[1], O. Toth[1], U. Rothlisberger[1]*

[1]Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL)

A hybrid quantum mechanics/molecular mechanics molecular dynamics (QM/MM MD) study explored the effect of oxidative damage on DNA redox properties in both unraveled and packed (nucleosome core particle, NCP) DNA systems under biological conditions. From the QM/MM MD studies, we applied a correlation-based feature selection machine learning algorithm to the packed DNA simulation data revealing the electrostatic and structural features which most directly influence the redox properties of a packed DNA system. Furthermore, our work to date shows that DNA lesions caused by oxidative damage yield subtle structural effects, but notable redox changes in guanine-rich regions of both unraveled and packed DNA. In addition to providing quantum-level accuracy of redox properties in DNA affected by oxidation, the findings support the proposed charge transfer (CT) mechanism for DNA damage recognition in guanine-rich DNA sites, regardless of DNA structure [1]. We aim to be able to identify potential oxidative damage “hotspots” in any nucleosome core particle in the future utilizing the data generated from ongoing QM/MM MD studies and the following feature selection analysis.

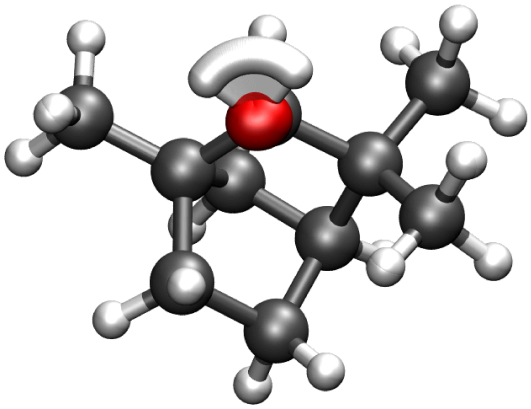
[1] Amie K Boal, Jacqueline K Barton, *Bioconjugate Chemistry*, **2005**, 16, 312–321.

Tunnelling in Complex Molecular Systems: Bridging Theory and Experiment

G. Laude[1], J. O. Richardson[1]*

[1]Department of Chemistry and Applied Biosciences, ETH Zurich

Quantum tunnelling is an ubiquitous phenomenon, relevant to many fields such as astrochemistry and biochemistry. It can for instance, result in larger-than-expected reaction rates, or a splitting of energy levels. However, unfavourable scaling of computational times with respect to system size hamper many attempts at determining tunnelling effects. With instanton theory, [1,2] one will only have to locate an optimal tunnelling pathway (called the instanton) to determine reaction rates and tunnelling splittings. The instanton also provides us greater insight on the tunnelling process by giving us a visual picture, as well as identifying the contribution each atom makes to the tunnelling process. In combination with machine-learning approaches, [3,4] one can choose to employ a higher level of electronic structure theory at a reduced computational cost, thereby increasing the accuracy of the calculations even further. We have now extended instanton theory such that we are able to evaluate tunnelling splittings in asymmetric systems, [5–7] and in vibrationally-excited states. [8] We apply our method to a wide array of complex molecular systems such as malonaldehyde, tropolone, α−fenchol and the vinyl radical, for which our results compare favourably with experiments.



[1] J. O. Richardson, *Int. Rev. Phys. Chem.* **2018**, 37, 171.
[2] J. O. Richardson, *J. Chem. Phys.* **2018**, 148, 200901.
[3] G. Laude, D. Calderini, D. P. Tew and J. O. Richardson, *Faraday Discuss*. **2018**, 212, 237.
[4] G. Laude, D. Calderini, R. Welsch and J. O. Richardson, *Phys. Chem. Chem. Phys.* **2020**, 22, 16843.
[5] E. Jahr*, G. Laude* and J. O. Richardson, *J. Chem. Phys.* **2020**, 153, 094101.
[6] M. Fiechter*, G. Laude* and J. O. Richardson, **2023**, *in preparation*.
[7] A. Nandi, G. Laude, S. S. Khire, N. D. Gurav, C. Qu, R. Conte, Q. Yu, S. Li, P. L. Houston, S. R. Gadre, J. O. Richardson, F. A. Evangelista and J. M. Bowman, *J. Am. Chem. Soc.* **2023**, 145, 9655.
[8] G. Laude and J. O. Richardson, **2023**, *in preparation*.

Explicit treatment of the time-dependent electromagnetic excitation in the nonadiabatic quantum dynamics in the adiabatic basis

Y. Lee[1], J. Vaníček[2]*

[1]Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), [2]Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL)

The shape of the excitation pulse has an effect on the subsequent nonadiabatic dynamics when the location of the initial wavepacket is close to the coupling region between different potential energy surfaces. Most computational studies make the sudden approximation, where the initial wavepacket after the (de-)excitation has the same shape as the ground state of the original surface. In this study, the excitation by the electromagnetic pulse is explicitly included in the Hamiltonian in the adiabatic representation. Since the nonadiabatic vector coupling makes the usual explicit split-operator algorithm unavailable, implicit integrators such as Crank-Nicolson method must be used instead [1]. We discuss and numerically show the conservation of the geometric properties by the implicit integrators and their symmetric composition to achieve an arbitrary even-order accuracy in the timestep [1] with a time-dependent potential. Furthermore, we study the effect of the exact shape of the excitation pulse on nonadiabatic dynamics to differentiate the regime where the excitation affects the subsequent dynamics and the sudden approximation is no longer valid.

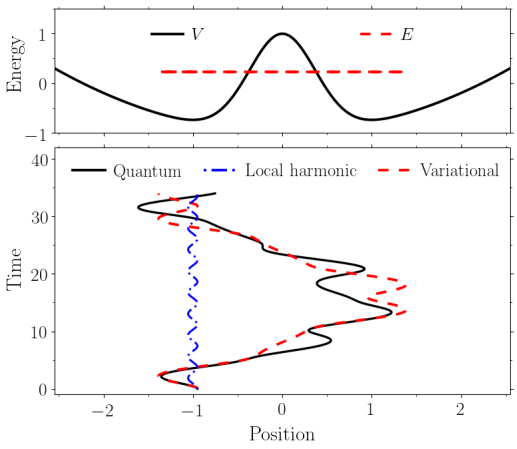
[1] Seonghoon Choi, Jiří Vaníček, *The Journal of Chemical Physics*, **2019**, 150, 204112.

Efficient high-order symplectic integrators for the variational Gaussian wavepacket dynamics

R. Moghaddasi Fereidani[1], J. Vanicek[1]

[1]Laboratory of Theoretical Physical Chemistry, Chemistry and Chemical Engineering

Among the single-trajectory Gaussian-based methods for solving the time-dependent Schrödinger equation, the variational Gaussian wavepacket dynamics [1,2] is the most accurate one. In contrast to Heller's original thawed Gaussian approximation [3], it is symplectic, conserves energy exactly, and partially takes into account tunneling. However, the variational method is also much more expensive. To improve its efficiency, we symmetrically compose the second-order symplectic integrator of Faou and Lubich [4] and obtain geometric integrators that can achieve an arbitrary even order of convergence in the time step. We demonstrate that the high-order integrators can drastically speed up convergence compared to the second-order algorithm (we show an example, where this speedup is by a factor of 100 if a moderate accuracy of 10^{-6} is required for the wavefunction) and, in contrast to the popular fourth-order Runge-Kutta method, are time-reversible and conserve the norm and the symplectic structure exactly, regardless of the time step. To show that the method is not restricted to low-dimensional systems, we perform most of the analysis on a non-separable twenty-dimensional model of coupled Morse oscillators. We also show that the variational method can include tunneling and, in calculation of vibronic spectra, improves accuracy over the non-variational thawed Gaussian approximation.



[1] E. J. Heller, Time dependent variational approach to semiclassical dynamics, J. Chem. Phys. **64**, 63 (1976).
[2] R. D. Coalson and M. Karplus, Multidimensional variational Gaussian wave packet dynamics with application to photodissociation spectroscopy, J. Chem. Phys. **93**, 3919 (1990).
[3] E. J. Heller, Time-dependent approach to semiclassical dynamics, J. Chem. Phys. **62**, 1544 (1975).
[4] E. Faou and C. Lubich, A Poisson integrator for Gaussian wavepacket dynamics, Comput. Visual. Sci. **9**, 45 (2006).

Autonomous Active Space Calculations through AutoCAS

M. Mörchen[1], M. Reiher[1]*

[1]Department of Chemistry and Applied Biosciences

In order to describe strongly correlated systems correctly, the choice of an active space imposes one of the greatest problems in multi-configurational quantum chemistry.
In the AutoCAS algorithm, concepts from quantum information theory are exploited in order to automatically and consistently select orbitals for an active space.
We present our Python-based AutoCAS [1, 2, 3, 4] module, which can be employed in existing workflows to streamline multi-configurational calculations in a black-box manner.
Due to the black-box-like selection of active spaces, post-active space methods like Tailored Coupled Cluster [5, 6, 7] or second order perturbation theory [8] can be routinely applied to recover dynamic correlation.
Furthermore, in the AutoRXN workflow [9], a workflow for the exploration of chemical reaction networks, we automatically validated CCSD(T) energies with potential multi-reference character through the AutoCAS module.

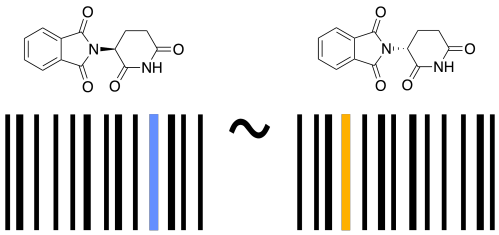
[1] Christopher J. Stein, Markus Reiher, *Journal of Computational Chemistry*, **2019**, 40, 2216-2226.
[2] Christopher J. Stein, Markus Reiher, *Journal of Chemical Theory and Computation*, **2016**, 12, 1760-1771.
[3] Christopher J. Stein, Vera von Burg, Markus Reiher, *Journal of Chemical Theory and Computation*, **2016**, 12, 3764-3773.
[4] Christopher J. Stein, Markus Reiher, *Chimia*, **2017**, 71, 170-176.
[5] Tomoko Kinoshita, Osamu Hino, Rodney J. Bartlett, *Journal of Chemical Physics*, **2005**, 123, 074106.
[6] Libor Veis, Andrej Antalík, Jiří Brabec, Frank Neese, Örs Legeza, Jiří Pittner, *Journal of Physical Chemistry Letters*, **2016**, 7, 4072–4078.
[7] Maximilian Mörchen, Leon Freitag, Markus Reiher, *Journal of Chemical Physics*, **2020**, 153, 244113.
[8] Leon Freitag, Stefan Knecht, Celestino Angeli, Markus Reiher, *Journal of Computational Chemistry*, **2017**, 13, 451-459.
[9] Jan P. Unsleber, Hongbin Liu, Leopold Talirz, Thomas Weymuth, Maximilian Mörchen, Adam Grofe, Dave Wecker, Christopher J. Stein, Ajay Panyala, Bo Peng, Karol Kowalski, Matthias Troyer, Markus Reiher, *Journal of Chemical Physics*, **2023**, 158, 084803.

Encoding Stereochemistry in Molecular Fingerprints

M. Orsi[1], D. Probst[2], J. Reymond[3]*

[1]University of Bern, Department of Chemistry, Biochemistry and Pharmacy, [2]École Polytechnique Federale de Lausanne, [3]University of Bern

Molecular fingerprints are important tools enabling fast similarity comparisons between molecules[1]. These fingerprints encode different molecular properties, such as pharmacophores[2], atom distances[3,4] and substructures[4–6], and are particularly useful in the context of virtual screening, where one carries out millions of comparisons to detect potential hits. However, a serious limitation of the most well-known fingerprints is their inability to account for chirality in a chemically meaningful manner. Specifically, the challenge lies in encoding chirality to distinguish enantiomers and diastereomers as highly similar yet distinct entities. In this study, we introduce MHFP6* and MAP4*, enhanced versions of our MHFP6[6] and MAP4[4] fingerprints, which correctly encode CIP chirality in a chemically meaningful manner. Both fingerprints demonstrate the capability to correctly differentiate isomers, thereby overcoming the previously mentioned limitations.



[1] Peter Willet, *Drug Discov. Today* **2006**, *11* (23–24), 1046–1053.

[2] Alice Capecchi, Mahendra Awale, Daniel Probst, Jean-Louis Reymond, *Mol. Inform.* **2019**, *38* (5), 1900016.

[3] Raymond E. Carhart, Dennis H. Smith, Rengachari Venkataraghavan,*J. Chem. Inf. Comput. Sci.* **1985**, *25* (2), 64–73.

[4] Alice Capecchi, Daniel Probst, Jean-Louis Reymond, *J. Cheminformatics* **2020**, *12* (1), 43.

[5] David Rogers, Mathew Hahn, *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754.

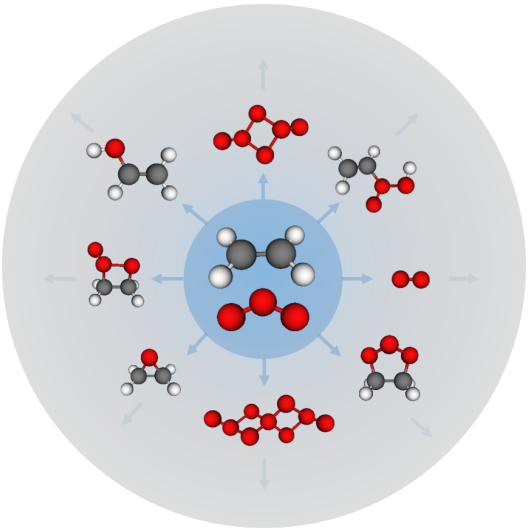
[6] Daniel Probst, Jean-Louis Reymond, *J. Cheminformatics* **2018**, *10* (1), 66.

Automated Reaction Network Exploration of Ozonation Processes in Water Treatment

E. Petrus[1,2], J. P. Unsleber[2], T. Weymuth[2], M. Reiher*[2], U. von Gunten[1,3]*, T. B. Hofstetter[1]*

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf 8600, Switzerland, [2]Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland, [3]School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), 1015, Lausanne, Switzerland

Ozone has been one of the flagship disinfectants in drinking water and wastewater treatment throughout the past decades, as well as for potable water reuse.[1] However, in many of its applications, the oxidant is scavenged in reactions with dissolved organic matter (DOM). These reactions lead to the formation of potentially toxic oxidation byproducts. Experimental identification of the byproducts is not only very laborious but also inherently incomplete and, thus, alternative strategies have been developed. In this context, quantum chemical (QC) computations have been applied to construct quantitative structure activity relations (QSAR) to ultimately predict rate constants for ozone consumption and byproduct formation routes.[2] Despite the advances of QC in this field, however, there are currently no examples focusing on determining the chemical reaction network, mainly due to the challenging task of finding numerous transition states with electronically complex structures.



Herein, we have addressed this challenge using the automated exploration tool named Chemoton, recently developed in the Reiher group.[3] We have generated a complex reaction network with species that represent the structural moieties of DOM (such as differently substituted olefins) that react with ozone. These explorations will enable us to carry out kinetic simulations which are directly comparable to the experimentally-determined rate constants and product yields.[4] Moreover, a systematic analysis of the network could unveil whether the preferred pathway reproduces the reported Criegee mechanism. Overall, we envision that this work can pave the way for future QC explorations of byproduct formation in disinfection/oxidation processes in aqueous solutions.

[1] Urs von Gunten *Environ. Sci. Technol*. **2018**, 52, 5062−5075. [2] Peter R. Tentscher, Minju Lee, Urs von Gunten *Acc. Chem. Res.* **2019**, 52, 3, 605–614. [3] Jan P. Unsleber, Stephanie A. Grimmel, Markus Reiher *J. Chem. Theory. Comput.* **2022**, 18, 5393-5409. [4] Dowideit, P., von Sonntag, C., **1998**. *Environ. Sci. Technol.* 1998, 32, 1112-1119.

Data-Driven Discovery of Electrocatalysts for the CO2 Reduction Reaction: Getting Into the Right Shape

B. Ranković[1,3], L. E. Zaza[2,3], R. Buonsanti[2,3]*, P. Schwaller[1,3]*

[1]EPFL SB-ISIC LIAC, [2]EPFL SB ISIC LNCE, [3]NCCR Catalysis

In the face of ongoing climate change, harnessing carbon dioxide (CO2) and converting it to useful chemicals via electrocatalytic processes stands out as a promising strategy [1, 2]. However, a major challenge is the lack of efficient and selective catalysts. Copper, despite its suitable binding energy for generating a range of products, grapples with selectivity issues in its bulk form.

Our study tackles this challenge by utilizing machine learning and data-driven computational tools to accelerate catalyst discovery. Integrating Electronic Lab Notebooks (ELN) [3], we have devised a process that makes data sharing across different chemistry laboratories machine-learning readable. This collaborative approach enables us to catalogue and analyze important synthesis and catalysis parameters. Notably, we enhance the dataset covering our wet-lab experiments with curated literature data, ensuring a comprehensive dataset for models to learn from.

Our models demonstrate promising results, accurately predicting nanocrystal shapes from reaction vectors derived from reagents and reaction conditions, features significantly associated with electrocatalytic properties [4]. We can classify multiple and intermediate shapes using multi-output classification models. Prediction probabilities provide an overview of different shapes present in the reaction output, underscoring our ability to decipher the rules guiding the formation of specific shapes from reaction vectors.

This uncovered correlation between reaction vectors and resulting nanoparticle shapes validates our research direction towards Bayesian optimization strategies [5]. Bayesian search facilitates exploration of the reaction space for specific shapes known to enhance selectivity. Importantly, it is not limited to known conditions but can potentially guide reactions towards direct selectivity optimization. An intriguing byproduct of this approach could be the emergence of nanoparticle shapes that were not initially targeted but have proven beneficial for catalytic selectivity, emphasizing the adaptive potential of our methodology.

In summary, our work presents a progressive approach towards the data-driven discovery of electrocatalysts, with a focus on CO2 reduction. It promises significant strides not only in enhancing the efficiency and selectivity of catalysts but also in transforming the traditionally slow process of lab data analysis into a more dynamic, automated one, paving the way for faster and more accurate catalyst discovery.

[1] Jianfeng Huang, Raffaella Buonsanti, *Chemistry of Materials*, **2018**, 31.1: 13-25.
[2] Giulia Mangione, Jianfeng Huang, Raffaella Buonsanti, Clémence Corminboeuf, *The Journal of Physical Chemistry Letters*, **2019**, 10.15: 4259-4265.
[3] Raffaella Buonsanti, *Chemistry of Materials*, **2023**, 35.3: 805-806.
[4] Rachel M. Anderson, David F. Yancey, Liang Zhang, Samuel T. Chill, Graeme Henkelman, Richard M. Crooks, *Accounts of chemical research*, **2015**, 48.5: 1351-1357.
[5] Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Martinez Alvarado, Jacob M. Janey, Ryan P Adams, Abigail G. Doyle, *Nature*, **2021**, 590.7844: 89-96.

ChORISO: a highly curated organic reaction SMILES dataset

V. Sabanza Gil[1], A. M. Bran[1], M. Franke[1], P. Schwaller[1,2]*, J. Luterbacher[1,2]*

[1]LIAC, EPFL EPFL Lausanne, [2]NCCR Catalysis

Artificial Intelligence (AI) and machine learning (ML) have been successfully applied to core problems in organic chemistry, like reaction outcome prediction and synthesis planning (1). Many of the underlying ML models are commonly trained using data in the form of Simplified Molecular Input Line Entry System notation2 (SMILES). Despite the popularity and use of some public available datasets like the US Patent and Trademark Office3 (USPTO), high quality data are scarce and difficult to extract manually. Therefore, new publicly available curated data could improve and leverage the existing ML models for chemical reaction tasks.

This work presents ChORISO (Chemical Organic ReactIon SMILES Omnibus), a new benchmark containing highly curated organic chemistry reaction SMILES. We have completed, processed and cleaned 3.2 M reaction SMILES extracted from chemical literature4, obtaining a final dataset with around 700k examples. We have analysed the most relevant features and compared them to the standard USPTO dataset. In addition, we have used our new dataset to train and benchmark different reaction prediction models (molecular transformer and graph-2-SMILES). This work offers a new high-quality organic reaction SMILES dataset which can contribute to the development and assessment of new chemical reaction prediction models.

References:
1. Schwaller et al.; WIREs Comput Mol Sci. 12 (2022) e1604.
2. Weininger D. J Chem Inf Comput Sci. 28 (1988) 31-6
3. Lowe, D.; Extraction of Chemical Structures and Reactions from the Literature. Ph.D. Thesis, University of Cambridge, 2012
4. S. Jiang et al. IEEE Access, 9 (2021) 85071-85083.

Frozen density embedding of CASSCF wavefunctions in CP2K

L. Schreder[1], S. Luber[1]*

[1]University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

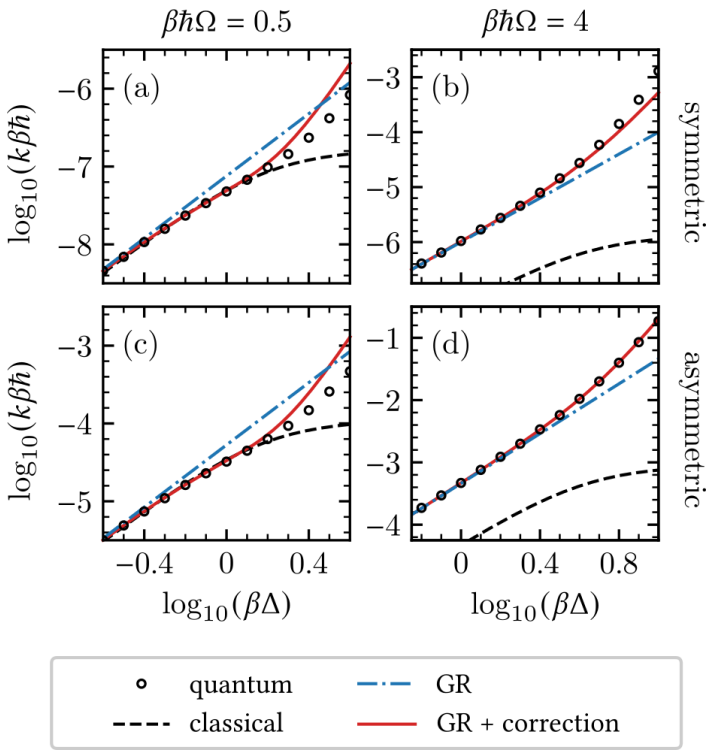
Most chemical processes happen at a local scale where only a subset of the molecular orbitals (MOs) is excited, and only a subset of covalent bonds may be rearranged. To capture such excited states or reactions, the efficient Density Functional Theory (DFT) is often inadequate, and the use of correlated wavefunction (CW) methods is required for accurate results. Due to their unfavorable computational scaling compared to DFT, this usually limits their use to small systems. Thus, studying a local excitation or chemical reaction in an extensive, relatively static environment is a balancing act between the accuracy of CW methods and the system size DFT can afford. A mixed-resolution approach backed by embedding theory is therefore preferable. Based on the self-consistent frozen density embedding algorithm, we developed a method of embedding complete active space self-consistent field (CASSCF) simulations run in the OpenMolcas code in a DFT environment calculated in CP2K.

Nonadiabatic reactions and tunnelling: beyond the golden-rule approximation

G. Trenins[1], J. O. Richardson[1]*

[1]Department of Chemistry and Applied Biosciences, ETH Zürich

Fermi's golden rule (GR) [1] is commonly used to calculate the rates of nonadiabatic reactions. Treated exactly, it is only tractable for simple models, but several approximate formulations have been developed, whose accuracy and computational efficiency make them a valuable tool in the study of chemical systems. GR instanton theory is one such formulation that accurately accounts for zero-point energy effects and multidimensional nuclear tunnelling in polyatomic systems [2-4]. However, not all nonadiabatic reactions are adequately described by Fermi's golden rule, and it has long been known that it can be improved with the help of perturbation theory. Much like the golden rule itself, exact perturbative corrections are too expensive for practical applications, and so we have developed a corresponding practical instanton formulation [5]. Our findings reveal how nonadiabatic effects can either suppress or enhance reaction rates, giving new mechanistic insight into the interplay of electronic transitions and nuclear tunnelling. We discuss how our new theory can be used to easily test and improve the accuracy of previous instanton calculations. We also present a novel procedure that uses our theory to find the optimal diabatic nuclear Hamiltonian for golden-rule calculations, expanding the scope of reactions that can be tackled with instanton techniques.



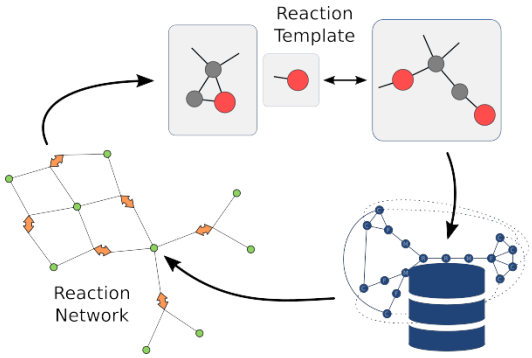
[1] Paul A. M. Dirac, *Proc. R. Soc. Lond. A,* **1927,** 114, 243-265.
[2] Jeremy O. Richardson, Rainer Bauer, and Michael Thoss, *J. Chem. Phys.,* **2015,** 143, 134115.
[3] Eric R. Heller and Jeremy O. Richardson, *J. Chem. Phys.,* **2020,** 152, 034106.
[4] Eric R. Heller and Jeremy O. Richardson, *J. Am. Chem. Soc.,* **2021,** 143, 49, 20952.
[5] George Trenins and Jeremy O. Richardson, *J. Chem. Phys.,* **2022,** 156, 174115.

Efficiently Charting Chemical Reaction Space with First-Principles Methods

J. P. Unsleber[1], M. Reiher[1]*

[1]Departement Chemie und Angewandte Biowissenschaften, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

Much algorithm development has been devoted to the computer-based automated exploration of chemical reaction networks in recent years.[1] These reaction network-based methods aim to establish a sufficiently complete reaction network to allow for reliable predictions, detailed understanding, and, eventually, the design of chemical processes. A comprehensive reaction network consists of all thermodynamically and kinetically relevant intermediates (under reaction conditions) and all reaction paths that connect them. Due to conformational diversity, a reaction network generated for relatively simple chemical processes can be vast if it is complete. This poses a set of challenges for methodology and software.[2] Here, we will present the current developments of our mechanism exploration automaton, called Chemoton[3], in its latest release version[4,5] and new developments to accelerate explorations[6].



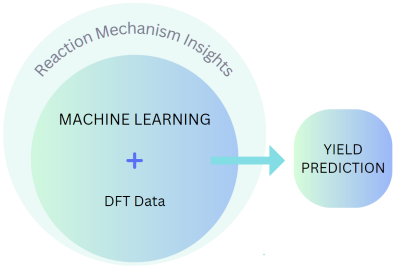
[1] G. N. Simm, A. C. Vaucher, and M. Reiher, *J. Phys. Chem. A,* **2018**, 123, 385
[2] J. P. Unsleber and M. Reiher, *Annu. Rev. Phys. Chem.,* **2020**, 71, 121
[3] G. N. Simm and M. Reiher, *J. Chem. Theory Comput.,* **2017**, 13, 6108
[4] J. P. Unsleber, S. A. Grimmel and M. Reiher, *J. Chem. Theory Comput.*, **2022**, 18, 5393
[5] Website: https://scine.ethz.ch, Source Code: https://github.com/qcscine
[6] J. P. Unsleber, *J. Chem. Inf. Model.*, **2023** in press, DOI: 10.1021/acs.jcim.3c00102

Machine Learning-Driven Yield Prediction in Organometallic Cross-Coupling Reactions

V. Vijay[1], C. Rajalakshmi[1], A. Vijayakumar[1], V. I. Thomas[1]*

[1]CMS college Kottayam, Kerala, India , Department of Chemistry Affiliated to Mahatma Gandhi University

We present a comprehensive study aimed at developing a machine learning-based yield prediction model for copper and zinc catalyzed Cadiot-Chodkiewicz reactions. Leveraging regression algorithms, our approach focuses on the integration of different electronic and chelating effect descriptors obtained from Density Functional Theory (DFT) optimized metal-ligand complexes.



To train and test our model, we utilize a dataset comprising both DFT obtained data and experimental results. Our prediction model exhibits remarkable performance, as indicated by a high correlation coefficient and low mean absolute error. Notably, we identify the HOMOLUMO gap and Ligating atom distance as the most influential factors governing yield in the studied catalysts.This research contributes to the understanding of a general mechanism for first-row high valent transition metal catalyzed reactions, enabling the development of more effective machine learning models for reaction optimization.

Prediction of Chemical Reaction Yields for C-O Cross-Coupling Reaction Using Machine Learning Technique

A. Vijayakumar[1], C. Rajalakshmi[1], V. Vijay[1], V. I. Thomas[1]*

[1]C M S College Kottayam

Prediction of Reaction Yield and catalytic activity using machine learning is an emerging field that has recently gained attention. We designed a Multiple Linear Regression Algorithm (MLRA) model using experimental yield for C-O cross-coupling reactions. We focused on 10 descriptors, including Homo-Lumo gap, bite angle, Mullicken charges, and dipole moment, which are important in considering electronic and steric effects. Various combinations of these descriptors were used for fitting analysis, including two, three, and four descriptors. Notably, we observed strong correlations with the experimental yield, particularly when considering the combination of Homo-Lumo gap and Mullicken charge in both cobalt and zinc complexes. Here, we have found that Homo-Lumo gap gives more contribution in determining yield for both cobalt and zinc complexes. Through our analysis, we developed an MLRA model for both cobalt and zinc complexes. This model enables the prediction of unknown yields, which plays a crucial role in evaluating and optimizing reaction conditions. The findings of this study provide valuable insights into yield prediction and have broad implications for various applications in the field of chemistry.