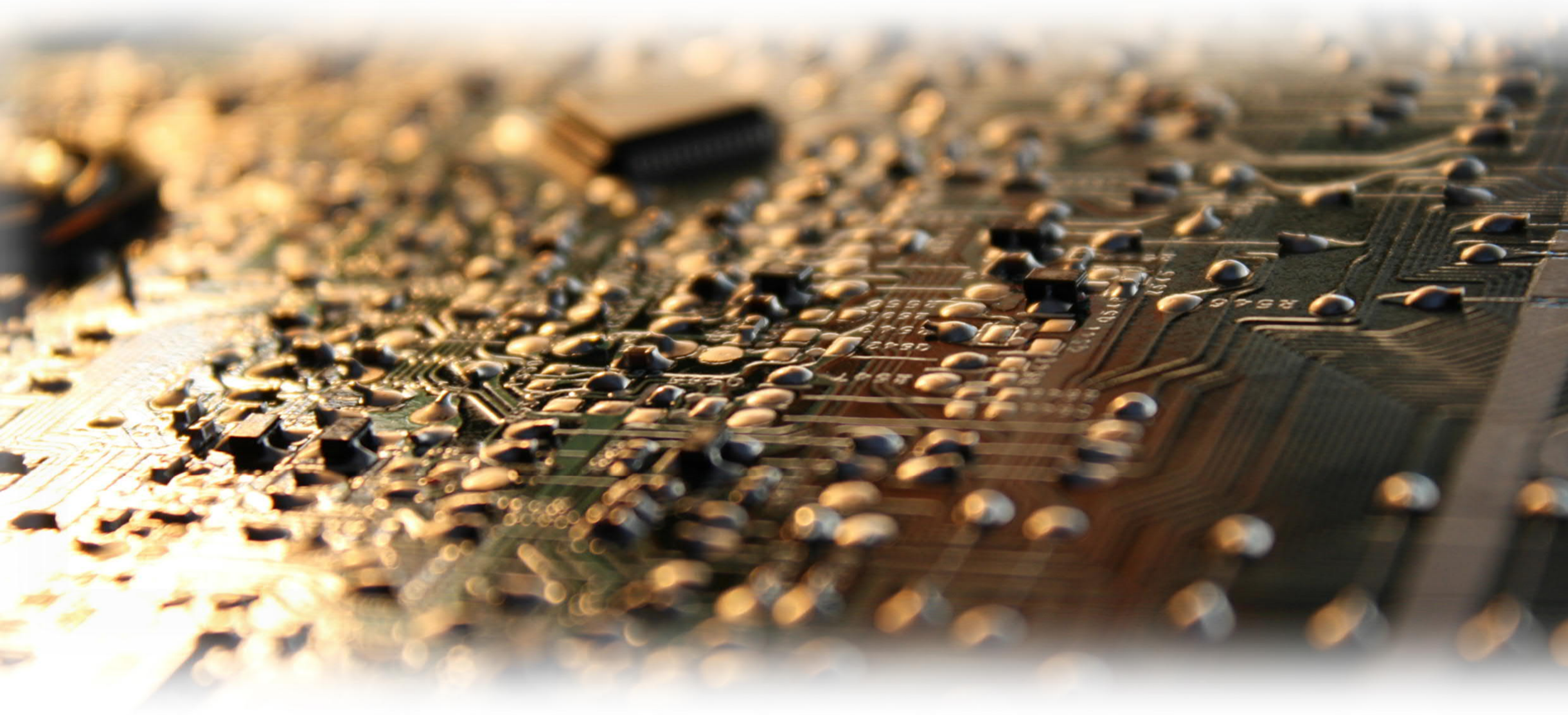
The 34th Winter School in Theoretical Chemistry

Machine Learning

University of Helsinki, December 10th—13th, 2018



Book of Abstracts

Organisers

Dage Sundholm

Mikael P. Johansson

http://www.chem.helsinki.fi/ws2018.html

Programme

Monday 10 December

9:00	Welcome	
9:15	Behler 1	High-Dimensional Neural Networks: Concepts and Applications, Part 1
10:00	Behler 2	High-Dimensional Neural Networks: Concepts and Applications, Part 2
10:45	Break	
11:00	Behler 3	High-Dimensional Neural Networks: Concepts and Applications, Part 3
11:45	Lunch	
13:00	Sillanpää	HPC-Europa3: Travel and collaborate with EC funding
13:10	Sluydts	Accelerating materials screening with machine learning
13:30	Popelier 1	Next generation force field design: state of the art and challenges
14:15	Popelier 2	Background to FFLUX: Quantum Chemical Topology and Gaussian Processes
15:00	Break	
15:15	Deringer 1	Machine-learning potentials for materials chemistry: fundamentals
16:00	Deringer 2	Machine-learning potentials for materials chemistry: applications to amorphous materials
16:45	Break	
17:00	Poster Session	

Tuesday 11 December

9:15	Goodpaster 1	Kernel Ridge Regression, Gaussian Processes, and Neural Networks in Quantum Chemistry
10:00	Ceriotti 1	Atom-density based representations for machine learning
10:45	Break	
11:00	Ceriotti 2	Not only potentials: learning vectors and tensors
11:45	Lunch	
13:00	Westermayr	Machine learning for excited-state molecular dynamics
13:20	Settels & Palmer	High-throughput forecasting of molecular properties in solution
13:45	Isayev 1	Predicting properties of inorganic materials with machine learning
14:30	Isayev 2	Neural Networks learning Quantum Chemistry
15:15	Break	
15:30	Popelier 3	Results and Future Work
16:15	Todorovic 1	Global atomistic structure search with Bayesian Optimization

Wednesday 12 December

9:15	Lilienfeld 1	Quantum Machine Learning in Chemical Space: Part 1
10:00	Lilienfeld 2	Quantum Machine Learning in Chemical Space: Part 2
10:45	Break	
11:00	Goodpaster 2	Learning Electron Correlation: Part 1
11:45	Goodpaster 3	Learning Electron Correlation: Part 2
12:30	Lunch and free afternoon	
13:00	(Annual meeting of the Computational Chemistry Section of the Finnish Chemical Societies)	

Thursday 13 December

9:15	Todorovic 2	Predicting molecular orbital energies with Kernel Ridge Regression
10:00	Todorovic 3	Deep-learning molecular spectra with Neural Networks
10:45	Break	
11:00	Deringer 3	Data-driven learning and prediction of inorganic crystal structures
11:45	Isayev 3	Deep Learning and Generative Models for Inverse Molecular Design
12:30	Closing and departure	

Understanding X-ray spectroscopy of carbonaceous materials by combining Density Functional Theory with Machine Learning

Anja Aarva,[1,*] Volker L. Deringer,[2,3] Tomi Laurila,[1] and Miguel A. Caro[1,4]

[1] *Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland*
[2] *Engineering Laboratory, University of Cambridge, Cambridge CB2 1PZ, United Kingdom*
[3] *Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom*
[4] *Department of Applied Physics, Aalto University, 02150 Espoo, Finland*

Carbonaceus materials, amorphous carbon (a-C) in particular, where carbon can be bonded in various ways (sp^1, sp^2 and sp^3) and form several different kinds of ring structures [1], are known to have many desirable properties to be used in biological applications. These materials are biocompatible and resistant to bacterial adhesion. They are widely used as protective coatings and biomedical coatings [2], but they are also potential materials for electrochemical detection of biomolecules *in vivo* [3].

Carbon-based materials often contain also elements other than carbon, such as hydrogen and oxygen. These elements can have a profound effect on the electrochemistry, electrical properties, and even on mechanical properties of the materials. Oxygen and hydrogen can be present in elementary form or as part of different functional groups on the surface. Although these carbon-based materials have been characterized in detail and their chemistry has been widely studied [3–5], because of the highly convoluted nature of interpretation of the experimental data, profound understanding of the relationship between structural characteristics and properties of these materials has not been achieved. Better understanding of the structure and especially the composition of the surface would be extremely valuable in development of carbon based applications.

X-ray absorption spectroscopy (XAS) and X-ray photoelectron spectroscopy (XPS), are very powerful methods for analyzing the surface characteristics and for studying functional groups present on the surface [6, 7]. However, since the materials in question are so complex, interpretation of the experimental data is challenging. Literature search about amorphous or other less organized forms of carbon characterization utilizing X-ray spectroscopy quickly reveals that interpretations of the spectra are often controversial. Hence, a simulation tool that can recreate the spectrum from first principles would aid the interpretation of the spectra remarkably.

In this work we utilize machine learning (ML) based structure construction [8]. In order to rationalize the complexity of the amorphous substance and to aid in deconvolution of the spectra, the structures are classified by employing clustering. The clustering is based on k-medoids [9, 10] where the dissimilarity matrix is obtained from Smooth Overlap of Atomic Positions (SOAP) [11] kernels. Although the main focus is on a-C, we also include graphene and diamond simulations for reference. Density functional theory based (DFT) simulations are employed to create the XAS spectra of these materials. In addition, the surfaces are functionalized with hydrogen and oxygen containing groups. We also calculate the Δ-Kohn-Sham (ΔKS) energy shifts which can be used to interpret XPS spectra [12]. Finally, our ambition is to be able to fit experimental spectra with our simulated spectra to provide a sophisticated interpretation tool for the community to be used in quantitative interpretation of spectroscopy of carbonaceous materials and to provide atomic level chemical information about the surfaces. This information is necessary, though not sufficient alone, when technologically important aims such as catalysis and biosensing are explored.

[1] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, and G. Csányi, Phys. Rev. Lett. **120**, 166101 (2018).
[2] J. Robertson, Mat. Sci. Eng. R **37**, 129 (2002).
[3] T. Laurila, S. Sainio, and M. A. Caro, Prog. Mater. Sci. **88**, 499 (2017).
[4] E. Peltola, J. J. Heikkinen, K. Sovanto, S. Sainio, A. Aarva, S. Franssila, V. Jokinen, and T. Laurila, J. Mater. Chem. B **5**, 9033 (2017).
[5] S. Sainio, D. Nordlund, M. A. Caro, R. Gandhiraman, J. Koehne, N. Wester, J. Koskinen, M. Meyyappan, and T. Laurila, J. Phys. Chem. C **120**, 8298 (2016).
[6] J. Stöhr, NEXAFS spectroscopy, Vol. 25 (Springer Science & Business Media, 2013).
[7] C. S. Fadley, J. Electron Spectrosc. **178**, 2 (2010).
[8] V. L. Deringer, M. A. Caro, R. Jana, A. Aarva, S. R. Elliott, T. Laurila, G. Csányi, and L. Pastewka, Chem. Mater. **30**, 7438 (2018).
[9] M. A. Caro, A. Aarva, V. L. Deringer, G. Csányi, and T. Laurila, Chem. Mater. **30**, 7446 (2018).
[10] C. Bauckhage, Numpy/scipy Recipes for Data Science: k-Medoids Clustering, Tech. Rep. (University of Bonn, 2015).
[11] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).
[12] L. Triguero, L. G. M. Pettersson, and H. Ågren, Phys. Rev. B **58**, 8097 (1998).

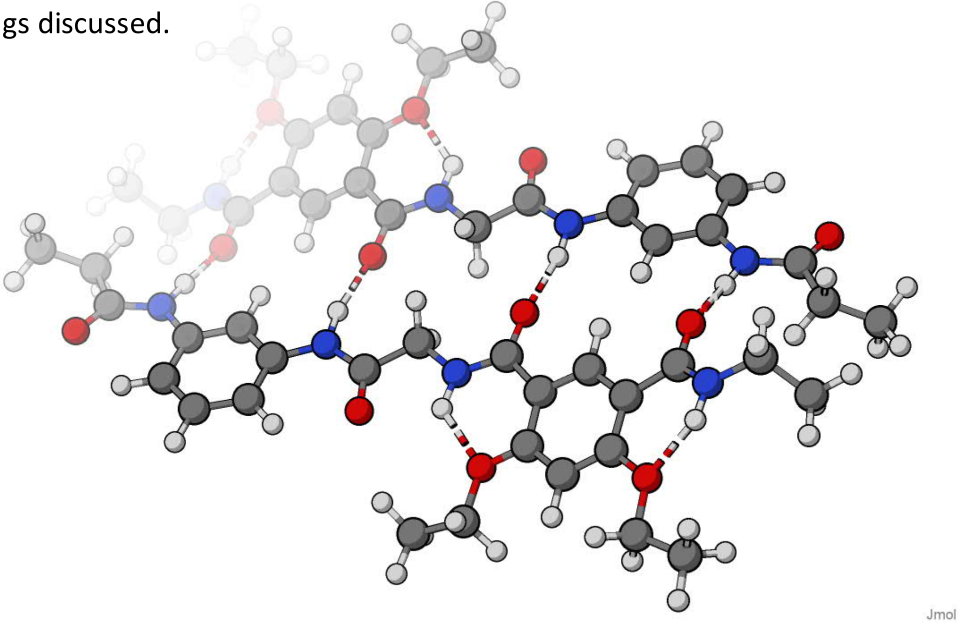
*anja.aarva@aalto.fi

Dissecting the interactions in quadruply H-bonded complexes

Usman Ahmed[1] and Mikael P. Johansson[1,2]
[1]Department of Chemistry, University of Helsinki, Finland
[2]Helsinki Institute of Sustainability Science (HELSUS), Finland
usman.ahmed@helsinki.fi

In recent years, self-assembly of molecules has drawn the interest of chemists to study the phenomenon of aggregation and the control of the aggregates through non-covalent interactions. Self-assembly of these systems is traditionally thought to arise mainly from hydrogen-bonding interactions.

Here, we studied various molecules that undergo spontaneous self-assembly using complementary hydrogen bonding. Complementary H-bonding can occur, if the molecule has hydrogen donor (D) and acceptor (A) groups suitably ordered. With four hydrogen-bonding groups, two main families exist: DDAA and DADA molecules. The interactions are investigated using Symmetry-Adapted Perturbation Theory (SAPT) recently extended for functional group analysis [1], and the findings discussed.



[1] R. M. Parrish, T. M. Parker, and C.D. Sherrill*, J. Chem. Theory Comput*. **2014**, 10, 4417.

Speeding up Constant pH Molecular Dynamics

N. Aho[1], G. Groenhof[1], B. Hess[2]

[1]University of Jyväskylä, P.O. Box 35, FI-40014 University of Jyväskylä, Finland
[2]KTH Royal Institute of Technology, Stockholm, Sweden
e-mail: noora.s.aho@jyu.fi

pH is one of the key parameters affecting the function and dynamics of proteins and other biomolecules. The effect of pH is due to changes in the protonation states of titratable groups, which alter the electrostatic properties of the biomolecules. Still, being able to dynamically change the protonation states and altering the pH is not a standard option in most of the classical computer simulation software. Usually in simulations, protonation states for titratable groups are initially chosen at the given pH, by selecting the statistically most probable protonation. This prevents the examination of pH-dependent phenomena, which is why an efficient constant pH molecular dynamics (MD) method in explicit solvent is highly desired.

The previous version of constant pH MD was released for GROMACS 3.3.3. There, the dynamics of the λ-particles was calculated by linearly interpolating between the Hamiltonians of all possible protonation states [1]. In the version under development, the λ-dynamics relies on the interpolation of the charges of the atoms in titratable groups, instead of interpolating the Hamiltonians. With this approach, only the calculation of the electrostatic potential for the system with interpolated charges is required. Thus, for N titratable sites only one PME call is needed, instead of the previously required 2N PME calls, resulting in a huge improvement in speed compared to the previous version.

Implementation of charge interpolation approach is still ongoing and needs to be followed by efficient parallelization and plenty of testing. The final goal is to implement the new version of constant pH MD as a permanently updated method in GROMACS, to have it available in all versions in the future.

References

[1] S. Donnini, F. Tegeler, G. Groenhof, H. Grubmüller, *Journal of Chemical Theory and Computation.,* **7**, 1962-1978 (2011).

Novel multiscale framework for modelling of liquid battery electrolytes

Rasmus Andersson [a], Fabian Årén[a], Alejandro A. Franco[b,c,d,e], Patrik Johansson [a,d]
[a] *Dept. of Physics, Chalmers University of Technology, 412 96, Gothenburg, Sweden*
[b] *Laboratoire de Réactivité et de Chimie des Solides (LRCS), Université de Picardie Jules Verne & CNRS, UMR 7314 – 33 rue Saint Leu, 80039 Amiens, France*
[c] *Réseau sur le Stockage Electrochimique de l'Energie (RS2E), CNRS FR3459, 80039 Amiens, France*
[d] *Alistore-ERI European Research Institute*
[e] *Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005 Paris, France*
E-mail: rasmus.andersson@chalmers.se

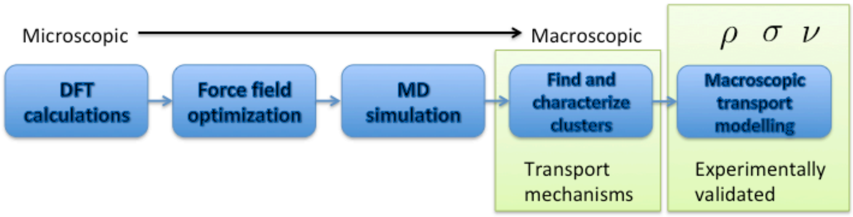
Transport properties of liquid battery electrolytes are conventionally modelled using variants of the Nernst-Planck equation, typically resolving single cationic and anionic species, respectively, hence assuming no ion-pairing, *i.e.* full cation solvation and also vehicular transport in stable solvation shells. Such modelling can straightforwardly use input obtained by electrochemical methods.

There are, however, several emerging classes of electrolytes that render these models inadequate, mainly due to a wider variation in electrolyte composition and charge carriers. As one example, for Li-S battery electrolytes the transport and solvation dynamics of polysulfides are crucial for the cell operation [1]. Another example is highly concentrated electrolytes, where too few solvent molecules are available to fully solvate the cation, leading to complex structure and dynamics by many partially solvated ionic aggregates [2].

To address the transport within such electrolytes, we here propose a novel multiscale method for modelling liquid electrolytes, starting from first principles. The method is based on three consecutive steps building from the microscopic to the mesoscopic: i) A system-specific classical force field is optimized using machine learning algorithms, overcoming problems of conventional force fields [3], while avoiding expensive polarization modelling and with reference data provided by density functional theory calculations, ii) molecular dynamics simulations for nanoseconds to microseconds and systems of 10^4-10^5 atoms provide statistics for, iii) trajectory analysis in a hierarchical manner by identifying clusters of ions and molecules and their structure, transport, and solvation and population dynamics.

The method predicts the structures to be included in continuum models, whose outputs can be directly validated against macroscopic quantities such as ionic conductivity, viscosity and density. Through its multiscale nature, this model yields not only macroscopic predictions, but also insights into how the macroscopic properties emerge from smaller scales.

References:
[1] J. Scheers, S. Fantini, P. Johansson, J. Power Sources. 255 (2014) 204-218
[2] D.M. Seo, O. Borodin, S-D. Han, Q. Ly, P.D. Boyle, W.A. Henderson, J. Electrochem. Soc.. 159 (2012) A553-A565
[3] D. Bedrov, O. Borodin, Z. Li, G.D. Smith, J. Phys. Chem. B 114 (2010) 4984-4997



SIMPLIFIED CHARGE DENSITY MODELS FOR ACCURATE ESTIMATION OF ELECTROSTATIC PROPERTIES

Sławomir A. BOJAROWSKI, Prashant KUMAR and Paulina M. DOMINIAK

Biological and Chemical Research Centre, Department of Chemistry, University of Warsaw,. Żwirki i Wigury 101, 02-089, Warsaw, Poland
sbojarowski@chem.uw.edu.pl

Fast and accurate estimation of electrostatic properties is still a challenging task. The electrostatic properties give a lot of valuable information of studied systems. The molecular electrostatic potential (MEP) demonstrates the negative and positive regions of molecules. It gives the information of the potential sites for nucleophilic or electrophilic attacks and can be a base for quantitative structure-activity relationship models. The electrostatic interaction energy (E_{es}) is commonly believed to be the most meaningful component of total interaction energy E_{int}, in the case of biomacromolecules, since they are highly polar systems. The E_{es} correlates the most with the E_{int}. This feature makes the accurate estimated E_{es}, apart from being important contribution itself, the best base for a scoring function in the docking procedure. However, the most popular models for electrostatic properties estimation, based on the simple Coulomb law with point charges (E_{mtp}), are fast, but they do not account for the penetration energy (E_{pen}). The latter occurs when the interacting molecules are so close that interaction energies cannot be estimated by just E_{mtp}. At such distances: $E_{es}=E_{mtp}+E_{pen}$. The E_{pen} can be up to 50% of total electrostatic interaction energy for polar molecules at the equilibrium distance.[1] Finally the model for fast and accurate E_{es} estimation, with the penetration effects, can be a basis for the electrostatic term in force fields. Furthermore, it can be the basis for new generation of *non-bonded* model in force fields.

Our concept is taken from the theoretical crystallography. Firstly, we applied aspherical pseudoatom databank (UBDB) which can easily reproduce the averaged electron density and calculate the electrostatic properties.[2] The UBDB relies on a Hansen-Coppens formalism of electron density. The multipole expansions is truncated at the hexadecapole level. With UBDB one can obtain a fully anisotropic model of charge density. It was already applied in biomacromolecular chemistry, particularly in such issues as interactions between influenza neuraminidase with inhibitors, syntenin PDZ2 domain interactions with short peptides, protein kinases interacting with sunitinib and many more.

However, to meet all the expectations mentioned in the first paragraph, we introduce an aug-PROmol model. It relies on a promolecule model of an electron density, originally augmented with monopoles fitted to electrostatic potential (RESP). It enables estimation of the exact electrostatic interaction energy including penetration effects. Its simplicity allows to omit computational-costly integration procedure. The integration procedure between reproduced, simplified charge densities from aug-PROmol, can be replaced by an analytical function.[3] The form of this function follows how E_{es} is computed from aug-PROmol model by the use of exact integration, but it is much simpler and faster in terms of computation time. Our recent investigations show that the RESP charges can be successfully replaced by the point charges from other, easy attainable, sources: from database or from semiempirical.[4] The electrostatic interaction energies estimated by aug-PROmol with different sources of point charges relies mostly on the E_{mtp}, and thus on the quality of point charges. Penetration energies computed with aug-PROmol model seems to be universal and not sensitive to applied source of point charges. For the benchmark molecular dimers, from S66x8 dataset, the mean absolute errors of estimated E_{es} are equal ~ 1.0 kcal mol^{-1} at the equilibrium distance, when referred to quantum mechanics at the DFT-SAPT/aug-cc-pVTZ level of theory. Error at this level is very satisfactory and is well known as a level of chemical accuracy. The molecular electrostatic potentials from the aug-PROmol model were calculated for small organic molecules from S66x8 dataset and further for dataset of HIV-1 protease inhibitors.[5] The aug-PROmol model with semi-empirical point charges give partially anisotropic potentials, far different from these obtained with just point charges. Obtained potentials show a great compatibility with the reference from quantum mechanic at the B3LYP/6-31G** level of theory.

Aug-PROmol model meets the expectations given in the introduction. It estimates the E_{es} with reasonable accuracy, it reproduces the molecular potential close to the reference and, by its simplicity, it can be easily applied as a base of scoring function in docking procedure or hereafter, it can be applied to force fields as an electrostatic term.

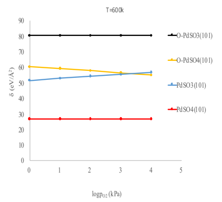
1. Bojarowski, S. A.; Kumar, P.; Dominiak, P. M., Acta Cryst., 2017, B73, 598-609.
2. Kumar, P.; Bojarowski, S. A.; Dominiak, P. M., et al. J. Chem. Theory Comput. 2014, 10, 1652-1664.
3. Bojarowski, S. A.; Kumar, P.; Dominiak, P. M., ChemPhysChem, 2016, 17, 2455-2460.
4. Bojarowski, S. A.; Kumar, P.; Wandtke, C. M.; Dittrich, B.; and Dominiak, P. M.; J. Chem. Theory Comput DOI: 10.1021/acs.jctc.8b00781
5. Kumar, P.; Gruza, B.; Bojarowski, S. A.; and Dominiak, P. M.; *Extension of transferable aspherical pseudoatom databank towards comparison of molecular electrostatic potentials for structure-activity studies*, submission stage

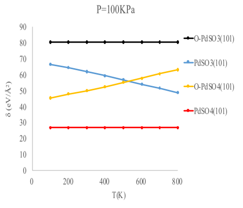
I would like to acknowledge the National Science Centre for PRELUDIUM-UMO/2015/19/N/ST4/00010 grant.

The stability of palladium sulphate and the adsorption on it: A first principle atomistic approach based on DFT

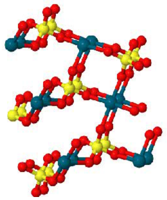
Tajudeen I Busari, Tapani Pakkanen, Mikko Linnolahti and Janne Hirvi
Department of Chemistry, University of Eastern Finland

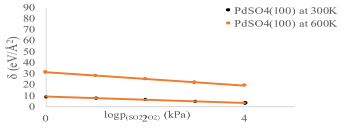
Abstract

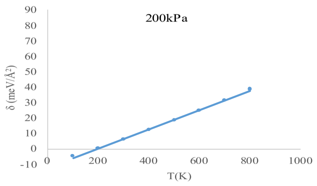




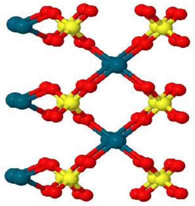
PdSO$_4$(101)







PdSO$_4$(100)



Catalytic combustion of methane is very important in natural gas engines. Natural engine is one of the fastest growing energy source in the world. In this system, complete methane removal is very essential because of its potential threat to the environment as a greenhouse gas. The catalyst of choice, PdO however is poisoned to $PdSO_4$ due to the presence of SO_2 and H_2O in the engine. In this regard, DFT calculations from the first principle atomistic approach were used to investigate the stability of $PdSO_4$ surfaces with different gas terminations and the adsorption of H_2O and CO on the stoichiometric $PdSO_4$(101) surface. Adsorption was only performed on this surface because there is no favourable adsorption site on the $PdSO_4$(100) counterpart. Stoichiometric surfaces, $PdSO_4$(101), where surface oxygen atom resides on the four coordinated sulphur and O-$PdSO_3$(101), where the oxygen resides on the four coordinated palladium had surface energies of 27meV/Å^2 and 81meV/Å^2 respectively. These values are independent of the temperature and pressure conditions. The surface energies of non-stoichiometric palladium sulphate surfaces, O-$PdSO_4$(101)-oxygen excess and $PdSO_3$(101)-oxygen deficient, were dependent on Gibbs energy of the oxygen as a function of temperature and pressure. Oxygen excess surface was more stable at high pressure and low temperature while oxygen deficient surface was more stable at low pressure and high temperature. $PdSO_4$(100) with excess of SO_2 and O_2 at the surface, had its surface energy dependent also on temperature and pressure. As temperature increases, its surface energy increases, whereas it decreases with increasing pressure. The computed adsorption energies for CO and H_2O on the three-fold coordinated surface palladium of the stoichiometric $PdSO_4$(101) surface showed that former was strongly adsorbed compared to latter. In conclusions, for the temperature and pressure conditions considered the $PdSO_4$(100) and $PdSO_4$(101) surfaces appeared to be the most stable while O-$PdSO_3$(101), the least stable. The $PdSO_4$(101) surface will however be a more useful catalyst because of the availability of a vacant site for adsorption as compared to the $PdSO_4$(100) where there is no vacant site.

An accurate and generally applicable atomic-charge based dispersion correction for density functional theory (DFT-D4)

Eike Caldeweyher[1], **Sebastian Ehlert**[1], **Andreas Hansen**[1], **Hagen Neugebauer**[1],
Sebastian Spicher[1], **Christoph Bannwarth**[2], **Stefan Grimme**[1]

[1] *Mulliken Center for Theoretical Chemistry, Beringstr. 4, D-53115 Bonn, Germany*
[2] *Department of Chemistry, Stanford University, Stanford, CA 94305, United States of America*

The so-called DFT-D4 model is presented for the accurate computation of London dispersion interactions in density functional theory (DFT) approximations. In this successor to the DFT-D3 model, the atomic coordination-dependent dipole polarizabilities are scaled based on classical atomic partial charges, which are obtained by solving a system of linear equations for which efficient analytical charge derivatives – with respect to nuclear positions – are developed for the first time in the present work. For this purpose, a new charge-dependent parameter-economic function is designed. A numerical Casimir-Polder integration of these atom-in-molecule dynamic polarizabilities then yields charge- and geometry-dependent dipole-dipole dispersion coefficients. Similar to the D3 model, the dynamic polarizabilities are pre-computed at the TD-PBE38/daug-def2-QZVP level and all elements up to radon ($Z = 86$) are included within an extended reference set. The two-body dispersion energy expression has the usual sum-over-atom-pairs form and includes dipole-dipole, as well as dipole-quadrupole interactions. For a benchmark set of 1225 molecular dipole-dipole dispersion coefficients, the D4 model achieves an unprecedented accuracy with a mean relative deviation of 3.9% compared to 4.7% for D3. In addition to the two-body dispersion energy, many-body effects are described up to third order by an Axilrod-Teller-Muto term. A common many-body dispersion description was extensively tested and assigned beneficial for larger system sizes in terms of an energy correction. Becke-Johnson-type damping parameters are determined for more than 60 common density functionals. For various standard energy benchmark sets DFT-D4 slightly but rather consistently outperforms DFT-D3. Especially for metal containing systems, the introduced charge dependence of the dispersion coefficients improves thermochemical properties. We suggest the new DFT-D4 model as a more sophisticated model in place of DFT-D3 for DFT calculations.

[1] Caldeweyher, E. and Bannwarth, C. and Grimme S. *J. Phys. Chem.* **2017**, *147*, 034112.

[2] Caldeweyher, E. and Ehlert, S. and Hansen, A. and Neugebauer, H. and Spicher, S. and Bannwarth, C. and Grimme S. *ChemRxiv*, in progress.

F-RESP: a fast and accurate polarizable force field

F. Cappelluti[1*], D. Ottaviani[2], A. Di Luca[3], D. Di Girolamo[4], L. Guidoni[1]

1 Università degli Studi dell'Aquila - L'Aquila
2 Cineca – Casalecchio di Reno
3 Technical University of Munich - Munich
4 Università degli Studi di Roma "La Sapienza" - Roma
** francesco.cappelluti@graduate.univaq.it*

ABSTRACT

F-RESP (Fluctuating Restrained ElectroStatic Potential) is a new polarizable model for classical molecular dynamics. In this model, atomic charges depend on the electric fields exerted along the bonds of the molecule they belong to by the other particles in the system.

The main advantage and innovation of this technique is that electric fields along bonds are calculated using fixed "generating" charges, which are obtained from RESP[1] ones during the parametrization procedure along with coefficients for charge variation function. In this way, the slow self-consistent process that would have been necessary using real charges is avoided.

The calculation of fictitious electric field (whose name is given by the use of the fictitious generating charges) can be performed using Ewald summation[2] technique or Fennell's modification[3] of Wolf potential[4].

Both models have been implemented in LAMMPS[5] and parallelized; for the second one, corrections to the virial are calculated, making possible to perform NPT simulations.

Good results have been obtained with simulations on water, while the application of the technique on more complex molecules was more tricky, mainly for the choice of parameters. This prompted us to engage in the improvement and generalization of the parametrization process, maybe through the use of a Machine Learning approach.

REFERENCES

1. Bayly, Christopher I., et al. "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model." *The Journal of Physical Chemistry* 97.40 (1993): 10269-10280.
2. Ewald, Paul Peter. "Die Berechnung optischer und elektrostatischer Gitterpotentiale." *Annalen der physik* 369.3 (1921): 253-287.
3. Fennell, Christopher J., and J. Daniel Gezelter. "Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics." *The Journal of chemical physics* 124.23 (2006): 234104.
4. Wolf, D., et al. "Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r− 1 summation." *The Journal of chemical physics* 110.17 (1999): 8254-8282.
5. http://lammps.sandia.gov

Poster contribution winter school Helsinki 2018

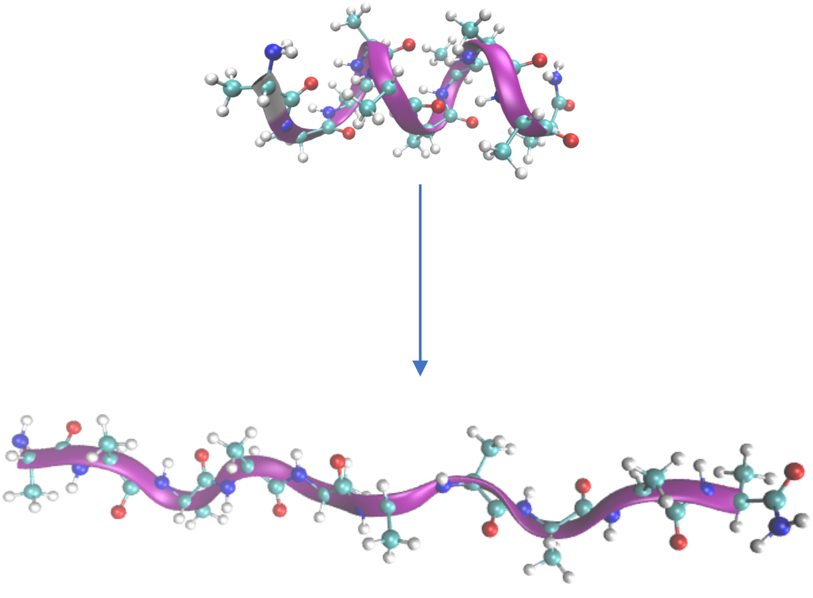
Title: Massively Parallel implementation of the Steered Molecular Dynamic in Tinker-HP for polarizable and non-polarizable force fields

Nom: Frédéric CELERSE

Supervisors: Jean-Philip PIQUEMAL (LCT) and Etienne DERAT (IPCM)

Sum:

Steered Molecular Dynamics (SMD) is a powerful methodology to accelerate rare events occurring in a Molecular Dynamics (MD) by applying an external force to a set of chosen atoms [1]. While SMD provides non-equilibrium trajectories, it is of a main interest to be able to reconstruct equilibrium properties such as Potential of Mean Force (PMF) [2,3]. To be able to use multiple distributed polarizable force fields (such as AMOEBA), the methodology has been implemented in the Tinker-HP software [4]. The SMD implementation has been firstly compared to NAMD, considered as a reference [1]. We further tested it with different type of force fields (polarizable and non-polarizable) with use of systems of different size. Polarization effect has been isolated and characterized by a decrease of the free energy barrier of the PMF compared to non-polarizable force fields in most of the cases.



[1] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... & Schulten, K. (**2005**). *J. Comput. Chem.*, *26*(16), 1781-1802.
[2] Park, S., Khalili-Araghi, F., Tajkhorshid, E., & Schulten, K. (**2003**). *J. Chem. Phys.*, *119*(6), 3559-3566.
[3] Park, S., & Schulten, K. (**2004**). *J. Chem. Phys.*, *120*(13), 5946-5961.
[4] Lagardère, L., Jolly, L. H., Lipparini, F., Aviat, F., Stamm, B., Jing, Z. F., ... & Gresh, N. (**2018**). *Chem. Sci.*, *9*(4), 956-972.

PTC-Simu: MACLAREN project 2018-2019

Ziyad Chaker[1], Thibault Charpentier[1], Mathieu Salanne[2], Jean-Marc Delaye[3]

[1]*French Alternative Energies and Atomic Energy Commission - CEA*
[2]*Sorbonne University and Maison De La Simulation (MDLS - CEA)*
[3]*French Alternative Energies and Atomic Energy Commission - CEA*

Abstract

Nowadays, due to the intensive use of nuclear energy in our modern society, the treatment of radioactive wastes through vitrification becomes a thematic of central interest. In particular, the question of the efficiency, viability and durability resulting from such sequestration methods refers directly to our ability to understand and predict the structure and properties of these nuclear glasses. We focus on the NMR investigation of these complex glasses to establish clearly the structure to property relationships characterizing them. Many simulation methods from MD, DFT to aiMD are used to investigate in this sense. Nevertheless, severe limitations exist such as the accuracy of the potentials used (MD) or the speed and feasibility of the calculations (aiMD). We propose to apply the state-of-the-art machine learning (ML) methods [1, 2] to: **(i)** accelerate the procedures already used for these calculations (construct ML force fields), **(ii)** use ML to provide a systematic approach for analyzing experimental NMR data (disentangle the atomic structure to NMR signals relationships) and **(iii)** go beyond simple scalar predictions and predict tensor DFT observables such as NMR, polarizability or stress tensors. The MACLAREN project subscribes in a global perspective aiming at the amelioration of modeling tools hitherto available for nuclear glasses studies. We plan to incorporate ML optimized schemes and the outputs from many spectroscopic approaches (NMR, RAMAN, structure factor measurements) into a reverse Monte-Carlo routine (fpNMR [3]) currently being developed. The aim is to have enough constraints to reach unprecedented accuracy in the establishment of the atomic structures of real samples.

References

[1] A. P. Bartók and G. Csányi, "Gaussian approximation potentials: A brief tutorial introduction," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1051–1057, 2015.

[2] A. S. Christensen, F. A. Faber, and O. A. von Lilienfeld, "Operators in machine learning: Response properties in chemical space," *arXiv preprint arXiv:1807.08811*, 2018.

[3] T. Charpentier, P. Kroll, and F. Mauri, "First-principles nuclear magnetic resonance structural analysis of vitreous silica," *The Journal of Physical Chemistry C*, vol. 113, no. 18, pp. 7917–7929, 2009.

Chiral Resolution of the Active Pharmaceutical Intermediates on the Phenylethylamine Stationary Phase.

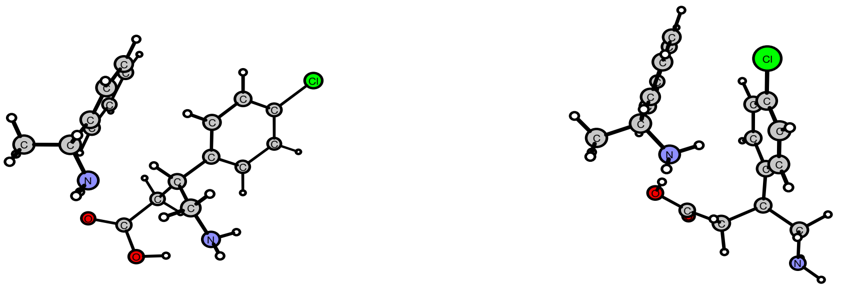
Michał Chojecki[1], Dorota Rutkowska-Żbik[2], Tatiana Korona[1]

[1] *Faculty of Chemistry, University of Warsaw, Warsaw, Poland*
[2] *Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Kraków, Poland*

Many bioactive compounds exist in two forms which only differ with the way they rotate light. These forms have identical chemical and physical properties but interact differently with other chiral molecules, including many important biochemicals. Therefore, when a new candidate for an active pharmaceutical intermediate (API) molecule is identified, a synthetic route that produces solely the desired enantiomer must be developed.

To this end, we aim to study the effectiveness of modelling of the interaction energy differences between the pair of the same enantiomers of the selected stationary phases, such as: phenylethanamine, glucose, and proline with selected APIs molecules such as: baclofen or ibuprofen. We utilise several popular theoretical methods, which allow for the energy partitioning, like symmetry-adapted perturbation theory (SAPT) [1] and functional-group SAPT (F-SAPT) [2], as well as a supermolecular MP2. The F-SAPT approach turns out to be especially useful for this analysis since it enables for an identification of a specific spot of the molecule responsible for the increased or decreased attraction or repulsion.

Figure 1: Optimised complexes of the baclofen enantiomers with phenylethanamine.



(a) R-baclofen $\cdots$ (1R)-1-phenylethanamine

(b) S-baclofen $\cdots$ (1R)-1-phenylethanamine

The support from the National Science Centre of Poland through grant 2015/19/B/ST4/01812 is gratefully acknowledged. This research was supported in part by PL-Grid Infrastructure.

[1] B. Jeziorski, R. Moszynski, K. Szalewicz, Chem. Rev. **94** (1994) 1887.

[2] R.M. Parrish, T.M. Parker, C.D. Sherill, J. Chem. Theory Comput. **10** (2014) 4417.

Comparing different machine learning force fields: a case study of aluminium

Maarten Cools-Ceuppens[1], Joni Dambre[2], and Toon Verstraelen[1]

[1] CENTER FOR MOLECULAR MODELING, Ghent University, Technologiepark-Zwijnaarde 903, 9052 Zwijnaarde, Belgium
[2] IDLab, ELIS, Ghent University, Technologiepark-Zwijnaarde 15, 9052 Zwijnaarde, Belgium

Machine learning force fields have become a popular and reliable method to sample the potential energy surface of molecular systems. One does not have to come up with a physical expression for the energy unlike conventional force fields, which makes them very flexible and a strong competitor to these traditional force fields (e.g. the embedded atom model). In the past few months and years, the amount of different machine learning force fields has increased rapidly, all excelling in other areas and on various datasets.

In this poster, three machine learning force fields are compared on an aluminium dataset: a linear model, SOAP [1] (kernel ridge regression with Gaussian approximation potentials) and SchNet [2] (a deep neural network making use of message passing). The full dataset contains a variety of MD trajectories, simulating bulk structures, surfaces, vacancies, grain boundaries, surfaces ... and was partially taken from literature [3] and extended with GPAW calculations. Before training, some MD trajectories were excluded to serve as an external validation set. The other structures are shuffled and split into the training and test set (90% - 10%). Evidently, the latter two sets consist of representative systems for the external validation set.

We show that more complex machine learning force fields have a greater predictive power when interpolating (have a lower error on the test set). Furthermore, they benefit more from a larger training set. On the contrary, they seem to suffer when extrapolating or making predictions of unseen structures (a larger error on the external validation set). Even the use of early stopping by having access to the unseen structures does not decrease the SchNet error below the value of the linear model. Moreover, the linear model can predict the lattice constants and bulk moduli of the equilibrium fcc, bcc and hcp crystal structures of aluminium at the same accuracy as the well-tested SOAP force field.

References

[1] Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, May 2013.

[2] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. Schnet – a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.

[3] V. Botu, R. Batra, J. Chapman, and R. Ramprasad. Machine learning force fields: Construction, validation, and outlook. *The Journal of Physical Chemistry C*, 121(1):511–522, 2017.

Towards efficient conformational sampling of short peptides

Martin Culka*[1], Daniel Herman*[1] and Lubomír Rulíšek[1]

martin.culka@uochb.cas.cz, daniel.herman@uochb.cas.cz

[1] Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic
* equal contribution, both presenting authors

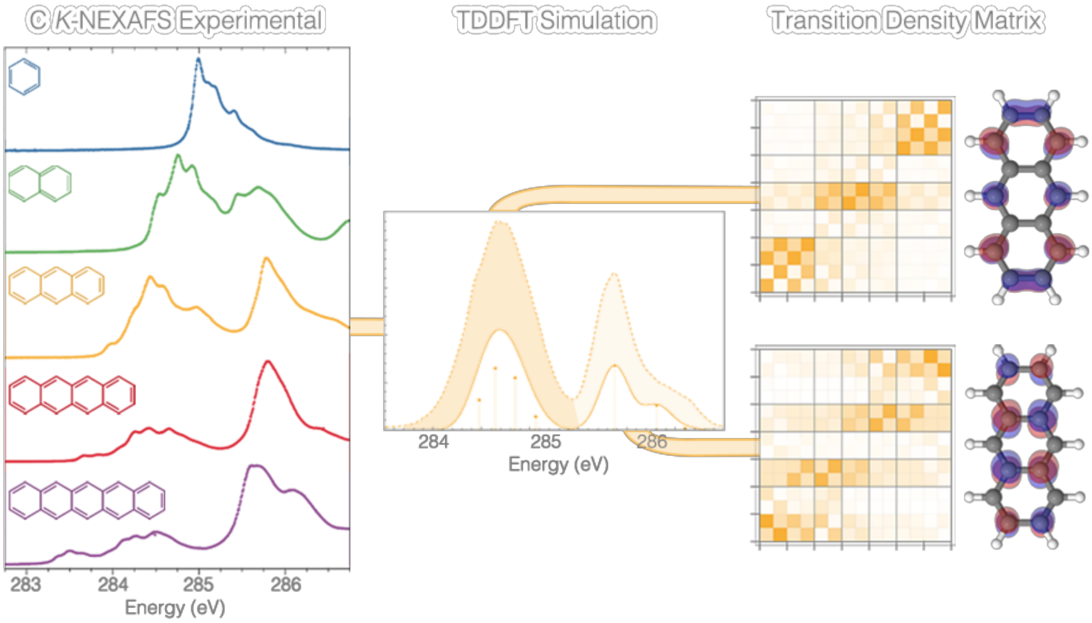
Efficient conformation sampling is an important issue in various areas of computational chemistry. While small molecules have limited well-defined number of conformations (often only one), more complex molecules and especially macromolecules show huge conformational variability, especially in solution. Proteins, biomacromolecules with complex folding patterns, represent a characteristic example of a chemical substance that only based on its conformation has different properties. Protein folding is a longstanding problem in computational biochemistry, since it is impossible to fully sample the conformational space of a protein using physical models such as empirical force fields, let alone more accurate quantum-chemical methods. We have recently (paper under review) identified sequences of tripeptide fragments with strong secondary structure preference in the protein structural database. We have shown that this preference is correlated with the characteristics of energetically low-lying conformers of the corresponding tripeptides in solution. We used metadynamics-based large-scale conformational sampling in combination with DFT-based geometry optimization and free energy calculation. We have also shown that empirical forcefields cannot provide as reliable energy ranking as the DFT. It was, however, necessary to optimize tens of thousands of conformers for each tripeptides – an enormous computational task. Therefor, we were able to evaluate only six most promising tripeptide sequences out of 8000. Furthermore, we obtained conformers with quite large energy span (up to 20 kcal/mol). In the study presented here we attempt to use machine learning approaches (such as neural networks) as a preselection tool. The conformers and free energies of the six tripeptides have been used as a training and validation set. We have then used the trained model to predict conformer energies. The low-energetic fraction has then been evaluated using our DFT-based approach in order to validate the preselection.

Electronic Structure of the Carbon *K*-Edge States in Polyacenes Molecules

Carlos E. V. de Moura[1], M. Barbatti[1], M. L. M. Rocco[2], A. B. Rocha[2],
M. Häming[3], A. Schöll[3], and E. Umbach[3]

1 *Aix Marseille Univ, CNRS, ICR (Marseille, France)*
2 *Universidade Federal do Rio de Janeiro, Instituto de Química (Rio de Janeiro, Brazil)*
3 *Universität Würzburg, Experimentelle Physik II (Würzburg, Germany)*

We present an experimental and theoretical study of high-resolution NEXAFS spectra of well-ordered films of benzene and a series of polyacenes molecules, namely naphthalene, anthracene, tetracene, and pentacene. The spectral complexity increases with the molecular size: NEXAFS features decrease in intensity and moves to lower photon energy (redshift), as the size of the aromatic system grows. Moreover, a second group of transitions arises. Vibrational fine structures coupled to the $C\,1s \rightarrow \pi^*$ transitions are apparent for all investigated molecules.



To improve the interpretation of the experimental measurements, theoretical chemistry models were applied to obtain the absolute band envelopes. The Nuclear Ensemble method was employed based on TDDFT electronic structure. TDDFT was performed with the Restricted Excitation Window approach with the ωB97X-D functional to the description of the electronic structure of the core excited states. We also analyzed the Transition Densities Matrices for the whole series of molecules. This approach improves the comprehension of the unique electronic nature of the spectroscopic features observed in the first two spectral bands.

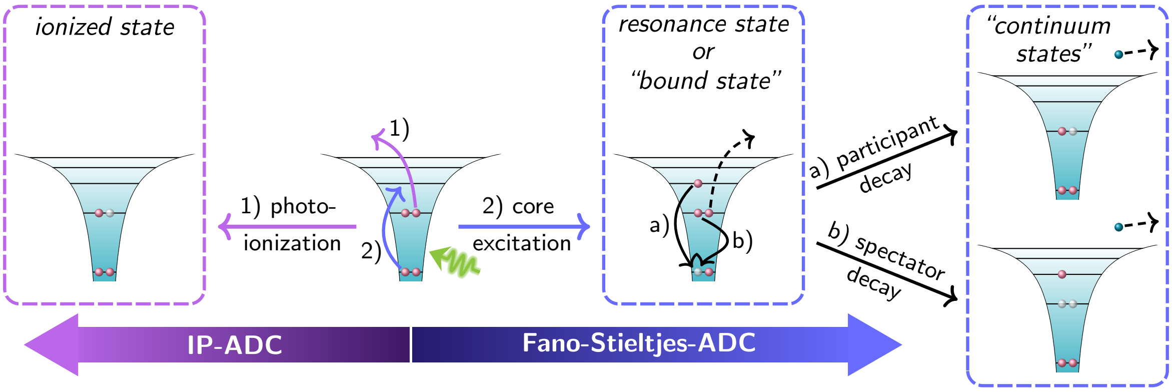
Accepted to publication:
High-Resolution NEXAFS Study of Condensed Polyacenes
The Journal of Physical Chemistry C, (2018) DOI:10.1021/acs.jpcc.8b08945

Resonance Lifetimes in Large Systems

ADRIAN L. DEMPWOLFF, ANDREAS DREUW

Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany.
email: adrian.dempwolff@iwr.uni-heidelberg.de

The theoretical modeling of ionization processes and simulation of corresponding photoelectron spectra is a demanding task for quantum chemistry. With the Algebraic Diagrammatic Construction approximation for the particle propagator (IP-ADC), a successful approach to this problem has been developed.[1] Here, we show an efficient implementation of IP-ADC up to third order in `adcman`,[2] a suite of ADC methods for electronically excited and ionized states, which is available through the Q-CHEM quantum-chemical program package.[3]
A second part of the presentation focuses on resonant autoionization. As illustrated in the right part of the figure below, such processes occur if an electronic system is excited above its ionization threshold. Due to the unbound nature of the involved resonance states, their theoretical description is challenging. An implementation of the Fano-ADC-Stieltjes method,[4] which combines Fano-Stieltjes theory[5] with the Algebraic Diagrammatic Construction scheme for the polarization propagator (ADC),[6] is presented.



Different ionization processes: Photoionization (left), autoionization/resonant Auger (right).

References

[1] J. Schirmer, A. B. Trofimov, G. Stelter, *J. Chem. Phys.* **1998**, *109*, 4734.

[2] M. Wormit, D. R. Rehn, P. H. P. Harbach, J. Wenzel, C. M. Krauter, E. Epifanovsky, A. Dreuw, *Mol. Phys.* **2014**, *112*, 774.

[3] Y. Shao *et al.*, *Mol. Phys.* **2014**, *113*, 184.

[4] K. Gokhberg, V. Averbukh, L. S. Cederbaum, *J. Chem. Phys.* **2007**, *126*, 154107.

[5] A. U. Hazi, *J. Phys. B* **1978**, *11*, L259.

[6] A. Dreuw, M. Wormit, *WIREs Comput. Mol. Sci.* **2015**, *5*, 82.

CH fragments in strong magnetic fields

Maria Dimitrova,[1,2*] Dage Sundholm,[1,2] Trygve Helgaker,[1,3] Trond Saue,[1,4]
Stella Stopkowicz,[1,5] Andrew Teale,[1,6] Wim Klopper,[1,7] Peter Schwerdtfeger[1,8]

[1]*CAS, Norwegian Academy of Sciences and Letters*
[2]*University of Helsinki, Finland*
[3]*University of Oslo, Norway*
[4]*Université Toulouse III – Paul Sabatier, France*
[5]*Johannes Gutenberg – Universität Mainz, Germany*
[6]*Karlsruhe Institute of Technology, Germany*
[7]*University of Nottingham, UK*
[8]*Massey University, Auckland, New Zealand*
* *maria.dimitrova@helsinki.fi*

The electrons in a molecule interact with an external magnetic field, affecting the electronic structure, energy and geometry. Atomic orbitals with high angular momenta lower their energy and can become the ground state. The Schrödinger equation involves an additional term, where the angular momentum operator appears, thus requiring that atomic and molecular orbitals are complex functions. We are investigating the electronic states and geometries of methane, and small its building blocks: H, C, H_2, H_3, H_4, CH, CH_2, and CH_3. The properties of the molecules are investigated in magnetic fields up to one atomic unit B_0 = 235 000 T. We compare the energies, geometries and the orbital symmetry of the electron configuration of the molecules. The atoms and molecules are studied in magnetic fields up to B_0 using the LONDON program[1] at the unrestricted Hartree-Fock level using the uncontracted aug-cc-pVTZ basis set and London orbitals to ensure gauge-origin independence. The multiplicity and stability of molecules change with the increase in the magnetic field strength. The CH fragment exhibits paramagnetic bonding when $B > 0.4B_0$ and exists as a bound sextet. CH_4 exists in a singlet or triplet state until $B \approx 0.1B_0$. The molecules become flat and orient perpendicularly with respect to the magnetic field vector.

References

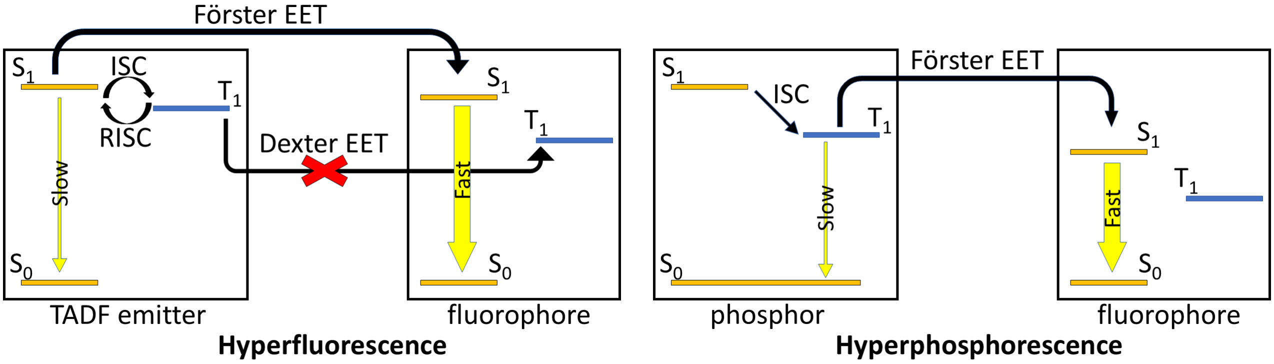
[1] Tellgren, Soncini, Helgaker. *J. Chem. Phys.* **129** (2008) 154114.

Excitonic Energy Transfer: An DFT/MRCI monomer transition density approach

Fabian Dinkelbach, Martin Kleinschmidt and Christel M. Marian

Institute of Theoretical and Computational Chemistry, Heinrich-Heine University Düsseldorf, Germany

Modern organic light emitting diodes employ phosphorescent or thermally activated delayed fluorescent (TADF) emitters. These emitters tend to have radiative lifetimes of microseconds, which is enough time for non-radiative deactivation processes to take place and to reduce the quantum yield. To shorten the radiative lifetime, energy transfer processes like Förster- excitation energy transfer (EET) can be used to move population to a strongly fluorescent emitter. In case of the TADF emitter, the quantum yield is increased by hyperfluorescence, where Förster EET moves singlet population towards the fluorophore. In case of a phosphorescent emitter, hyperphosphorescence takes place and population is moved from the excited triplet towards the fluorophore's excited singlet state. For these processes to be efficient Dexter EET should be avoided.



We use the monomer transition density approach[1,2] (MTD) which calculates the excitonic coupling for Förster and Dexter energy transfer employing the transition densities of monomers and the two-electron integrals of the dimer.

$$V_{Foerster} = \sum_{ijkl} [(\phi_i\phi_j|\phi_k\phi_l) - \frac{1}{2}(\phi_i\phi_l|\phi_k\phi_j)] \cdot D^{SS}_{D,ij} D^{SS}_{A,kl} \tag{1}$$

$$V_{Dexter} = -\sum_{ijkl} (\phi_i\phi_l|\phi_k\phi_j)[D^{ST_+}_{D,ij} D^{T_+S}_{A,kl} + D^{ST_-}_{D,ij} D^{T_-S}_{A,kl} + \frac{1}{2} D^{ST_0}_{D,ij} D^{T_0S}_{A,kl}] \tag{2}$$

The transition density matrices are obtained with the DFT/MRCI method[3] which proved to be a fast, accurate and reliable method. For close-distance interactions, like Dexter EET, the MTD-approach suffers from the non-orthogonality of the monomer bases, which has to be corrected[4,5].

The method can be used with DFT/MRSOCI[6] coupled states to investigate systems where relativistic effects play an important role.

References

1. R. Fink, J. Pfister, A. Schneider, H. Zhao, B. Engels, Chem. Phys., 343, 353-361 (2008)
2. R. Fink, J. Pfister, H. Zhao, B. Engels, Chem. Phys., 346, 275-285 (2008)
3. C. Marian, A. Heil, M. Kleinschmidt, WIREs Comput Mol Sci. e1394 (2018)
4. K. Fujimoto, J. Chem. Phys. 137, 034101 (2012)
5. H. Nagae, T. Kakitani, T. Katoh, M. Mimuro, J. Chem. Phys., 98, 8012 (1993)
6. M. Kleinschmidt, J. Tatchen, C. Marian, J. Chem. Phys., 124, 124101 (2006)

Novel proton transfer mechanism in the HIV-1 RNase H phosphodiester cleavage

Simon L. Dürr[1,2], Olga Bohuszewicz[1], Reynier Suardiaz[1], Pablo G. Jambrina[1], Christine Peter[2], Yihan Shao[3], Edina Rosta[1]

[1]Department of Chemistry, King's College London, London, SE1 1DB, United Kingdom [2]Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany [3] Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, United States

Contact email: edina.rosta@kcl.ac.uk

RNase H is a prototypical example for two metal ion catalysis in enzymes. An RNase H activity is present in the HIV-1 reverse transcriptase but also in many other nucleases such as *Homo sapiens* (Hs) or *Escherichia coli* (Ec) RNase H. The mechanism of the reaction has already been extensively studied based on the *Bacillus halodurans* (Bh) RNase H crystal structure[1–3] most recently using time-resolved X-Ray crystallography. However, kinetic and mutation experiments with HIV-1, Hs and Ec RNase H implicate a catalytic histidine in the reaction not present in Bh RNase H and the protonation of the leaving group also remains poorly understood.[4]

We use quantum mechanics/molecular mechanics (QM/MM) calculations combining Hamiltonian replica exchange with a finite-temperature string method to study the cleavage of the ribonucleic acid (RNA) backbone of a DNA/RNA hybrid catalyzed by the HIV-1 RNase H with focus on the proton transfer pathway and the role of the histidine. The reported pathway is consistent with kinetic data obtained with mutant HIV-1, Hs and Ec RNase H, the calculated pK_a values of the DEDD residues and crystallographic studies. The overall reaction barrier of ~16 kcal mol^{-1}, encountered in the first step, matches the slow experimental rate of ~1-100 min^{-1}.

Using Molecular dynamics (MD) calculations we are able to sample the recently identified[3] binding site for a third transient divalent metal ion in the vicinity of the scissile phosphate in the product complex. Based on data from our simulations, a similar observation of a third metal ion facilitating product release in an *Aquifex aeolicus* RNase III crystal structure[5] and the *in crystallo* reaction we are able to show that the third ion and the histidine are key to product release as had been hypothesized[4].

(1) Rosta, E.; Nowotny, M.; Yang, W. *J. Am.* ... **2011**, 8934–8941.

(2) Nowotny, M.; Yang, W. *EMBO J.* **2006**, *25* (9), 1924–1933.

(3) Samara, N. L.; Yang, W. *Nat. Struct. Mol. Biol.* **2018**, *25* (8), 715–721.

(4) Alla, N. R.; Nicholson, A. W. *FEBS J.* **2012**, *279* (24), 4492–4500.

(5) Gan, J.; Shaw, G.; Tropea, J. E.; Waugh, D. S.; Court, D. L.; Ji, X. *Mol. Microbiol.* **2008**, *67* (1), 143–154.

ANALYSIS OF THE TAILORED COUPLED-CLUSTER METHOD IN QUANTUM CHEMISTRY

FABIAN M. FAULSTICH [*], MIHÁLY MÁTÉ [†], ANDRE LAESTADIUS[*], MIHÁLY ANDRÁS CSIRIK [†], LIBOR VEIS [‡], ANDREJ ANTALIK [‡], JIŘÍ BRABEC [‡]REINHOLD SCHNEIDER[§], JIŘÍ PITTNER [‡], SIMEN KVAAL[*], AND ÖRS LEGEZA[†]

We analyze mathematical properties of tailored coupled-cluster (TCC) methods as finite dimensional non-linear Galerkin schemes. This multi-reference formalism combines the single-reference coupled-cluster (CC) approach with a full configuration interaction solution covering the static correlation [3]. This keeps the computational costs low as the full configuration interaction solution is calculated for a subsystem and guarantees the high accuracy of the CC method for the dynamical correlation.

Based on Zarantonello's Theorem, we prove that TCC schemes attain locally unique solutions fulfilling a quasi-optimal error bound. To that end, we formally characterize the TCC function and show local strongly monotonicity and Lipschitz continuity. We introduce a CAS-EXT-gap assumption for multi-reference problems replacing the HOMO-LUMO gap which is unreasonable for statically correlated systems. From this new assumption, results from previous analyses [6, 4, 5] are adaptable to the TCC formalism presented here [1]. Further we perform a first error analysis revealing the mathematical complexity of the TCC-methods. Due to the basis-splitting nature of the TCC formalism, the error decomposes into several parts. Using the Aubin-Nitsche-duality method we derive a quadratic (Newton type) error bound. However, as TCC-methods do not converge to the Full-CI solution a methodological error enters this bound. This can be qualitatively explained by the basis-splitting, although the quantitative connection corresponds to a long known and unsolved problem in quantum information theory.

We present numerical investigations on the robustness with respect to the bond dimensions of the single orbital entropy and the mutual information, which are quantities that are used to choose the complete active space. Furthermore, we extend the mathematical analysis with a numerical study on the complete active space dependence of the error [2].

REFERENCES

[1] F. M. Faulstich, A. Laestadius, S. Kvaal, Ö. Legeza, and R. Schneider, *Analysis of the coupled-cluster method tailored by tensor-network states in quantum chemistry*, arXiv preprint arXiv:1802.05699, (2018).

[2] F. M. Faulstich, M. Máté, A. Laestadius, M. A. Csirik, L. Veis, A. Antalik, J. Brabec, R. Schneider, J. Pittner, S. Kvaal, et al., *Numerical and theoretical aspects of the dmrg-tcc method exemplified by the nitrogen dimer*, arXiv preprint arXiv:1809.07732, (2018).

[3] T. Kinoshita, O. Hino, and R. J. Bartlett, *Coupled-cluster method tailored by configuration interaction*, The Journal of chemical physics, 123 (2005), p. 074106, https://doi.org/10.1063/1.2000251.

[4] T. Rohwedder, *The continuous coupled cluster formulation for the electronic schrödinger equation*, ESAIM: Mathematical Modelling and Numerical Analysis, 47 (2013), pp. 421–447, https://doi.org/10.1051/m2an/2012035.

[5] T. Rohwedder and R. Schneider, *Error estimates for the coupled cluster method*, ESAIM: Mathematical Modelling and Numerical Analysis, https://doi.org/10.1051/m2an/2013075.

[6] R. Schneider, *Analysis of the projected coupled cluster method in electronic structure calculation*, Numerische Mathematik, 113 (2009), pp. 433–471, https://doi.org/10.1007/s00211-009-0237-3.

---

[*] Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway(f.m.faulstich@kjemi.uio.no).

[†] Strongly Correlated Systems "Lendület" Research Group, Wigner Research Center for Physics, H-1525, P.O. Box 49, Budapest, Hungary.

[‡] J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, v.v.i., Dolejškova 3, 18223 Prague 8, Czech Republic

[§] Modeling, Simulation and Optimization in Science, Department of Mathematics, Technische Universität Berlin, Sekretariat MA 5-3, Straße des 17. Juni 136, 10623 Berlin,Germany

Machine learning of correlated dihedral potentials for atomistic molecular force fields

Pascal Friederich [1,2*], *Manuel Konrad* [1], *Timo Strunk* [3], *Wolfgang Wenzel* [1]

[1] Institute of Nanotechnology, Karlsruhe Institute of Technology, Germany
[2] Department of Chemistry, University of Toronto, Canada
[3] Nanomatch GmbH, Germany

* pascal.friederich@kit.edu & pascal.friederich@utoronto.ca

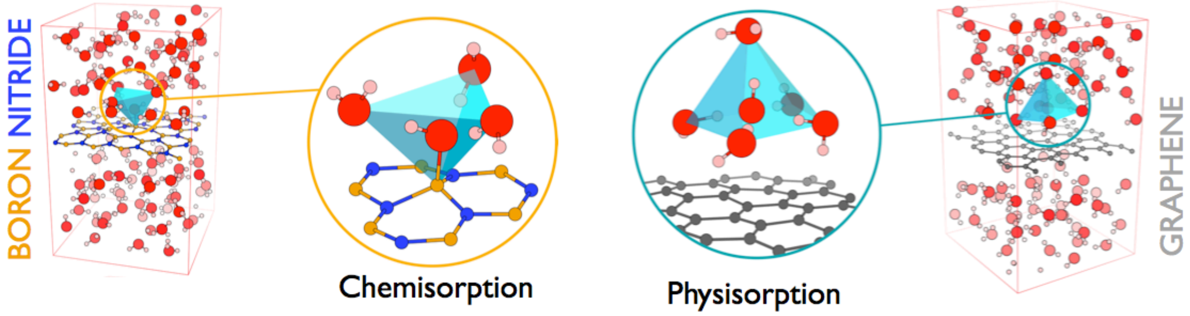
Computer simulation increasingly complements experimental efforts to describe nanoscale structure formation [1,2]. Molecular mechanics simulations and related computational methods fundamentally rely on the accuracy of classical atomistic force fields for the evaluation of inter- and intramolecular energies [3,4]. One indispensable component of such force fields, in particular for large organic molecules, is the accuracy of molecule-specific dihedral potentials which are the key determinants of molecular flexibility. We show in this work that non-local correlations of dihedral potentials play a decisive role in the description of the total molecular energy - an effect which is neglected in most state-of-the-art dihedral force fields. We furthermore present an efficient machine learning approach to compute intramolecular conformational energies [5]. We demonstrate with the example of α-NPD, a molecule frequently used in organic electronics, that this approach outperforms traditional force fields by decreasing the mean absolute deviations by one order of magnitude to values smaller than 0.37 kcal/mol (16.0 meV) per dihedral angle.

[1] Friederich *et al.*, *Chemistry of Materials*, 2017, 29 (21), pp 9528-9535
[2] Friederich *et al.*, *ACS Appl. Mater. Interfaces*, 2018, 10 (2), pp 1881-1887
[3] Friederich *et al.*, *Advanced Functional Materials*, 2017, 26, pp 5757-5763
[4] Friederich *et al.*, *Advanced Materials*, 2017, 29 (43), pp 1703505
[5] Friederich *et al.*, *Scientific Reports*, 2018, 8, pp 2559

How Graphene and Hexagonal Boron Nitride Get Electrified in Water?

Benoît Grosjean[1], Marie-Laure Bocquet[1], and Rodolphe Vuilleumier[1]
[1]*PASTEUR, Département de chimie,*
École normale supérieure, PSL University,
Sorbonne Université, CNRS, 75005 Paris, France

The recent emergence of nanofluidics has highlighted the exceptional properties of graphene and its boron-nitride counterpart (hBN) as confining materials for water and ion transport[1-4]. Surprinsingly ionic transport experiments have unveiled a large electrification of the water-BN surfaces, with a contrasting response for its water-carbon homologue[1]. This charging was conjectured to originate in the differential hydroxide adsorption on the 2D materials, but the challenge of simulating this elusive anion has precluded a proper explanation up to now. Here we report free energy calculations based on *ab initio* molecular dynamics simulations of a hydroxide in water near graphene and hBN layers. Our results[5] disclose that both surfaces electrify by hydroxide adsorption via different mechanisms. OH- shows strong chemisorption on hBN, but only weak physisorption on graphene. Interestingly OH- is shown to keep a fast lateral interfacial mobility while physisorbed. Taking into account the resulting large ionic surface mobility, an analytical transport model allows to reproduce quantitatively the experimental data. Our results offer new foundations for the chemical reactivity of carbon and BN materials in water and suggest new perspectives for advanced membrane technologies for water purification and energy harvesting[6].



References

[1] Siria, A. *et al*. Giant osmotic energy conversion measured in a single transmembrane boron nitride nanotube. *Nature* 494, 455–458 (2013).
[2] Secchi, E., Niguès, A., Jubin, L., Siria, A. & Bocquet, L. Scaling behavior for ionic transport and its fluctuations in individual carbon nanotubes. *Phys. Rev. Lett.* 116, 154501 (2016).
[3] Radha, B. *et al*. Molecular transport through capillaries made with atomic-scale precision. *Nature* 538, 222–225 (2016).
[4] Majumder, M., Siria, A. & Bocquet, L. Flows in one-dimensional and two-dimensional carbon nanochannels: Fast and curious. *MRS Bull*. 42, 278–282 (2017).
[5] Grosjean, B., Bocquet, M.-L. & Vuilleumier, R. *Submitted*, 2018.
[6] Siria, A., Bocquet, M.-L. & Bocquet, L. New avenues for the large-scale harvesting of blue energy. *Nat. Rev. Chem.* 1, 0091 (2017).

Application of *GW*-BSE theory to transition-metal complexes

Xin Gui[1], **Wim Klopper**[1]

[1] *Institute of Physical Chemistry, Karlsruhe Institute of Technology, Fritz-Haber-Weg 2, 76131 Karlsruhe, Germany*

The Bethe-Salpeter equation (BSE) [1] was initially derived in nuclear physics and later imported to the field of computational solid-state physics. Recently, it has been widespread for calculating molecular excitation energies and provided optimistic results in a variety of systems.[2-6] The BSE formalism is very similar to the time-dependent density functional theory (TDDFT) in the language of linear response, and thus has the same scaling with respect to system size as TDDFT. However, BSE can be applied for description of excitations that are particularly problematic for TDDFT, e.g. excitations with charge-transfer character. The BSE approach has recently been implemented in our TURBOMOLE program using a resolution-of-the-identity (RI) approximation for all two-electron integrals that are required to solve the equation.[7] Since BSE normally requires quasiparticle energies from preceding *GW* calculations as input, *GW* and BSE have often emerged together, which is known as the *GW*-BSE formalism. The performance of the BSE approach for the computation of singlet and triplet excitation energies of small molecules has been assessed with respect to the quasiparticle energies used in the BSE calculations.[8] Here we present several examples of application of *GW*-BSE to transition-metal complexes.

[1] E. E. Salpeter, H. A. Bethe, *Phys. Rev.* **84**, 1232 (1951).

[2] D. Jacquemin, I. Duchemin, X. Blase, *J. Chem. Theory Comput.* **11**, 3290 (2015).

[3] D. Jacquemin, I. Duchemin, X. Blase, *J. Chem. Theory Comput.* **11**, 5340 (2015).

[4] D. Jacquemin, I. Duchemin, X. Blase, *Mol. Phys.* **114**, 957 (2016).

[5] X. Blase, P. Boulanger, F. Bruneval, M. Fernandez-Serra, I. Duchemin, *J. Chem. Phys.* **144**, 034109 (2016).

[6] D. Jacquemin, I. Duchemin, X. Blase, *J. Phys. Chem. Lett.* **8**, 1524 (2017).

[7] K. Krause, W. Klopper, *J. Comput. Chem.* **38**, 383 (2017).

[8] X. Gui, C. Holzer, W. Klopper, *J. Chem. Theory Comput.* **14**, 2127 (2018).

Large scale atomistic simulations on respiratory complex I

Outi Haapanen[1], Amina Djurabekova[1], Vivek Sharma[1,2]
[1]Department of Physics, University of Helsinki, P.O. Box 64, 00014 Helsinki, Finland
[2]Institute of Biotechnology, University of Helsinki, FI-00014 Helsinki, Finland
e-mail: outi.haapanen@helsinki.fi

Energy production is the one of the most important processes in our biosphere. Cells use the energy in the form of ATP, which is produced via oxidative phosphorylation (OXPHOS) in the inner mitochondrial membrane by respiratory complexes, to catalyse biochemical reactions.

Respiratory complex I is the first member of the respiratory chain and one of the largest known protein assemblies (500 kDa – 1 MDa) with two proteinaceous arms, one spanning along the inner mitochondrial membrane with 70 transmembrane helices and the other embedded in the mitochondrial matrix.

Complex I takes part in OXPHOS by pumping protons across the membrane contributing to the proton electrochemical gradient required by ATP synthase for ATP generation. Even though the structure of complex I was solved already in 2013, the functional mechanism remains unknown. The functional cycle consists of two separate, however coupled, reactions: reduction of quinone to quinol, and pumping of four protons across the membrane. The coupled reactions are separated by a distance of ca. 20 nm and temporally by reaction speed of one order of magnitude. The supposedly unique coupling mechanism is one the main questions in the field of bioenergetics.

The major research question in our study is to enlighten the unknown coupling mechanism with computational methods. Using multi-scale simulation approaches, we are studying the dynamics of quinone and in both bacterial and mammalian variants of the enzyme. Our results provide details for the recent mechanistic proposals [1] and new exciting viewpoints for experimental research. The significance of the research on complex I lies within the diseases caused by the mutations and other dysfunctions of complex I, such as Alzheimer's and Parkinson's disease.

[1] Outi Haapanen and Vivek Sharma, BBA Bioenergetics, **1859,** 510-523 (2018).

The Novel Materials Discovery Laboratory - Towards Data-driven Materials Science

L. Himanen[1], A. S. Foster[1,2,3], P. Rinke[1], R. M. Nieminen[1]
[1]Aalto University, P.O. Box 11100, 00076 Aalto, Espoo, Finland
[2]Graduate School Materials Science in Mainz, Staudinger Weg 9, 55128, Germany
[3]WPI Nano Life Science Institute Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan
e-mail: *lauri.himanen@aalto.f*

NOMAD, the Novel Materials Discovery Laboratory, is a European Centre of Excellence (CoE) which was established in fall 2015. It comprises of eight complementary research groups in computational materials science along with four high performance computing (HPC) centers. NOMAD creates, collects, stores, and cleanses atomistic computational materials science data, computed with a broad set of materials-science codes available today. These underlying tasks then allow the project to develop tools for mining this data to find structure, correlations, and novel information that could not be discovered from studying smaller data sets. In a larger context the goal of the project is to use the large volume of data and innovative tools to enable researchers in basic science and engineering to advance materials science, identify new physical phenomena, and help industry to improve existing and develop novel products and technologies.

In this contribution we discuss the shift towards data-driven science and present the services that NOMAD provides for researchers working with large amounts of atomistics simulations: free raw data storage (recommended by the journal Nature), data mining and exploration tools and big-data analytics tools for in-depth data analysis and machine learning.

NOMAD

NOVEL MATERIALS DISCOVERY

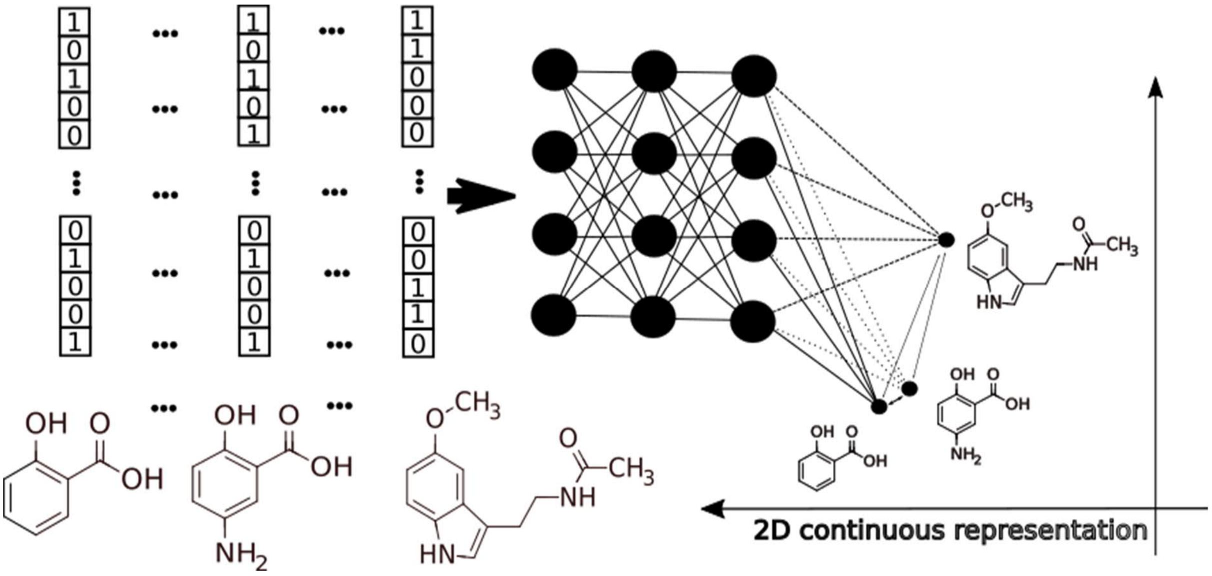
References

[1] https://nomad-coe.eu/

Chemical Space exploration guided by deep neural networks

Dmitry S. Karlov[a], Sergey Sosnin[a], Igor V. Tetko[b,c], Maxim V. Fedorov[a]
[a] *Center for Computational and Data-intensive Science and Engineering, Skolkovo Institute of Science and Technology, Nobelya Ulitsa 3 Moscow, 121205 Russia.*
[b] *Helmholtz Zentrum München - Research Center for Environmental Health (GmbH), Institute of Structural Biology, Ingolstädter Landstraße 1, D-85764 Neuherberg Germany.*
[c] *BIGCHEM GmbH, Ingolstädter Landstraße 1, G. 60w, D-85764 Neuherberg, Germany*

Parametric t-SNE approach based on deep feed-forward neural network was applied to the chemical space visualization problem. It is able to retain more information than certain dimensionality reduction techniques used for this purpose (Principal component analysis (PCA), Multidimensional scaling (MDS)). The applicability of this method for some chemical space navigation tasks (activity cliffs and activity landscapes identification) is discussed.



Approximation of water dimer potential energy surface with moment tensor polynomials.

Andre Lomaka, **Toomas Tamm**

Tallinn University of Technology, Department of Chemistry and Biotechnology, Tallinn, Estonia

Moment tensor polynomial basis that satisfies both rotational and permutational invariance restrictions, has been recently proposed in the literature as a promising novel family of descriptors for systematic approximation of Born-Oppenheimer surfaces. Its extension to multicomponent systems [1] containing more than one atom type requires nonlinear optimization of model parameters that is harder to implement than linear optimization and can suffer from multiple minima problem. We provide an alternative relaxed formulation of the original potential energy function for multicomponent systems where parameter optimization can be expressed as a linear sparse approximation problem. Since the number of descriptors in the reformulated problem grows combinatorically with the number of atom types giving rise to possibly millions of descriptors with high multi-collinearity, there is a need for ultra-large dimensional feature selection. In order to reduce the number of descriptors to a manageable size, we applied high-dimensional ordinary least-squares projection (HOLP) [2] for initial screening of variables. Further reduction of descriptor space was achieved through recursive feature elimination [3] based on standardized regression coefficients. We demonstrate the utility of the proposed method for the approximation of potential energy surfaces by taking water dimer, a 12-dimensional system with 2 chemical elements, as a test example. A moment tensor potential model with up to four-body terms and tensor rank limited by 2 (matrices) has been developed for water dimer to approximate energy values from density functional theory calculations. 100000 configurations covering the energy range ~600 kcal/mol were used for model training and cross-validation. A total of ~54000 descriptors initially calculated was reduced to ~6000 descriptors selected as most relevant to the model. Testing of our model with an independent set of 100000 configurations yielded root mean square error of prediction smaller than 0.3 kcal/mol.

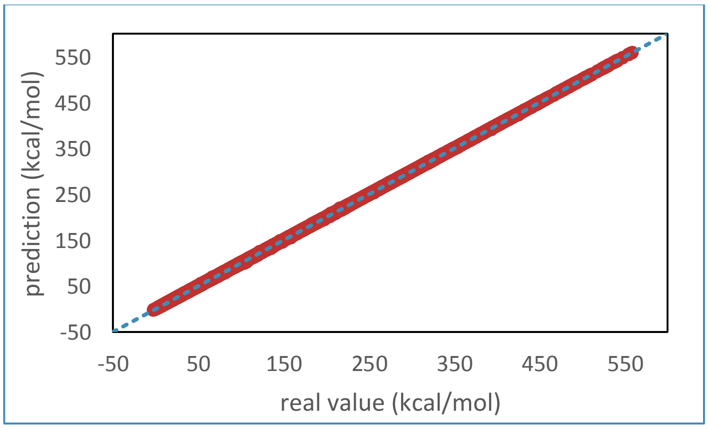


Figure 1. Scatter plot of predicted vs. B3LYP/6-31G* energy values for 100000 water dimer geometries of external test set. Energies are given relative to twice the energy of an optimized water molecule.

References:

1. Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. *J. Chem. Phys.*, *148*, **2018**, 241727.
2. Wang, X.; Leng, C. *J. R. Statistic. Soc. B*, *78*, **2016**, 589-611.
3. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. *Mach. Learn.*, *46*, **2002**, 389-422.

Solvent Effects on Linear Cu(I)-CAAC Complexes

Nora Lüdtke, Jelena Föeller and Christel M. Marian
Department of Chemistry, Heinrich Heine University, Düsseldorf, Germany
nora.luedtke@uni-duesseldorf.de

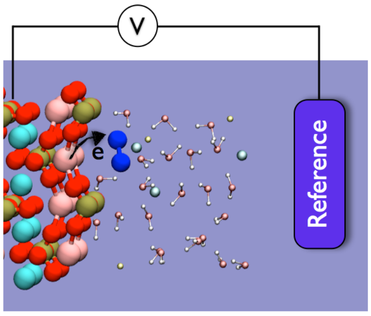
Since Organic Light Emitting Diodes (OLEDs) have gained a lot of importance in light and display technology, the demand for new and better OLED emitters has strongly increased. In recent years, the search for potent electroluminescent transition metal complexes was based on abundant metals such as copper to replace rare and precious metal centers. While investigations were mainly focused on tetragonally and triagonally coordinated copper complexes, linear copper complexes synthesized by Gernert et al.[1] yielded good internal quantum efficiency. This work builds on the investigations by Gernert et al. by presenting quantum chemical investigations of the spectroscopic properties of $Cu(CAAC^{Me})Cl$ and the $Cu(CAAC^{Me})_2^+$. Because of the charge-transfer (CT) character of its S_0-S_1 and S_0-T_1 transitions, solvation is expected to have considerable impact on the spectroscopic properties of $Cu(CAAC^{Me})Cl$. By means of combined density functional theory (DFT) and multireference configuration interaction (MRCI) methods, absorption and emission wavelengths were calculated in vacuo and in THF. Moreover, Lewis bases have been indicated to form exciplexes with low-valent copper complexes. To consider their impact, an explicit THF molecule and an implicit solvent shell were applied for calculating the photophysics of $Cu(CAAC^{Me})Cl$. The solvent environment causes a huge blue shift of the absorption wavelength of the chloride complex. In contrast, the solvation only has a small impact on the spectroscopic properties of the symmetric $Cu(CAAC^{Me})_2^+$ cation. Calculated rate constants of spin-allowed and spin-forbidden radiative transitions and of (reverse) intersystem crossing processes complement the picture and enable the assignment of the emission spectra.

References

[1] Gernert, M., Müller, U., Haehnel, M., Pflaum, J. and Steffen, A.: A Cyclic Alkyl(amino)carbene as Two-Atom π-Chromophore Leading to the First Phosphorescent Linear Cu(I) Complexes, *Chemistry*, **2017**, *23*, 2206-2216.

Constant potential thermodynamics and kinetics in electrochemical environments

Marko Melander
Department of Chemistry, University of Jyväskylä, Finland
`marko.m.melander@jyu.fi`



Abstract

Using the electrode potential to control reaction kinetics and thermodynamics forms the very core of all electrochemical systems. To provide a satisfactory atomistic model, a proper description of the electrochemical interface needs to include the electrode, electrolyte, and their interface as a function of the electrode potential. The inherent complexity of the electrochemical interface presents a yet unsolved challenge for atomistic simulations.

For thermodynamics the most elegant approach is to perform the simulations in a grand canonical (GC) ensemble where both the electronic and ionic chemical potentials can be fixed as done in experiments; by controlling these variables we can extract atomic-level information about the electrocatalytic and charge transfer reactions at the electrode interface. I will present a general GC multi-component density functional theory (GC-DFT) framework to model the electrochemical interface. The new theoretical approach is able to treat the thermodynamics of the interface, including the solvent, ionic double layer and explicit electrode potential.

The developed GC-MCDFT framework is also extended to treat electrocatalytic reaction kinetics at fixed potentials. I have extended thermal rate theory to GC ensemble to treat i) adiabatic reactions within GC transition state theory, ii) adiabtic Marcus-type treatment of electron and proton transfer, iii) computation of non-adiabatic electron, proton, and proton-coupled electron transfer kinetics, and iv) interpolation between non-adiabatic – adiabatic – solvent controlled kinetics. While the standard toolbox of DFT and transition state can access adiabatic inner-sphere simultaneous proton-coupled electron transfer reaction, the new approach allows treatment of pure (outer-sphere) electron transfer, decoupled proton-electron transfer and nonadiabatic effects. The work extends the scope of electrocatalytic reactions amenable to an atomistic computational description.

The new models enable the of study atomistic details of a wide array of reaction mechanisms and environment effects. In my presentation I will introduce theory behind GC-DFT and its use to describe thermodynamics and kinetics of electrocatalytic reactions.

References:

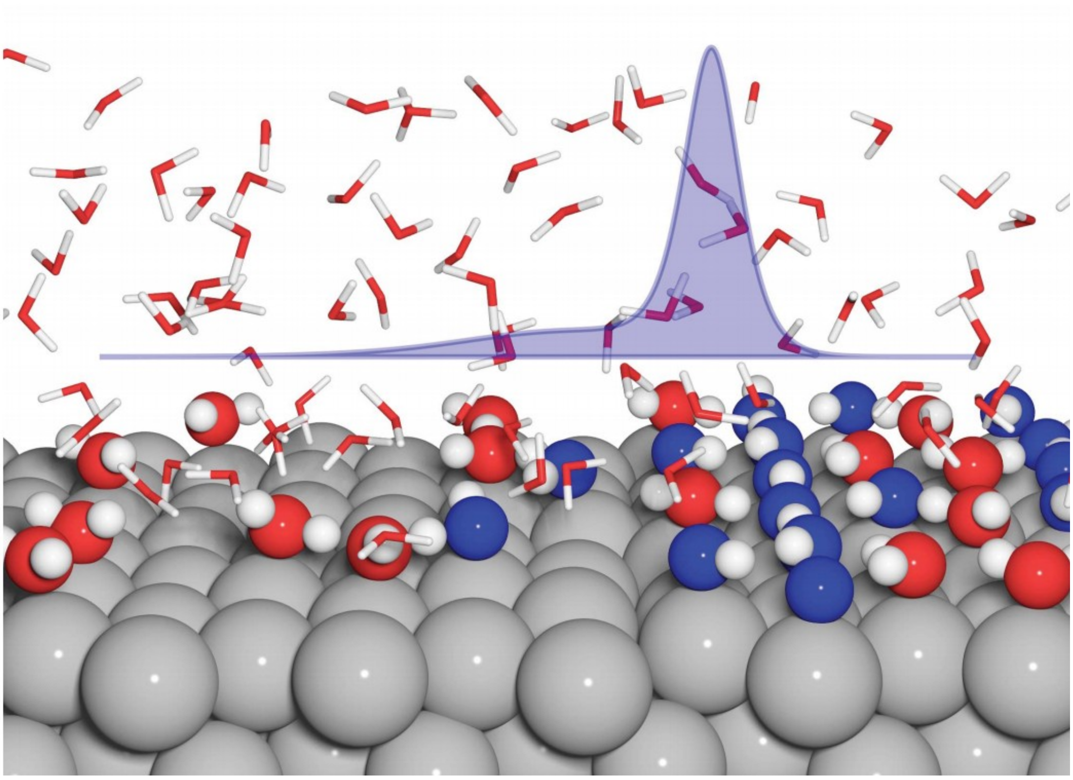
M. Melander, M. Kuisma, T. Christensen, K. Honkala, J. Chem. Phys. 150, 041706 (2019)
M. Melander, *in preparation*

Modelling the Water/Pt(111) interface using Machine Learning

August. E.G. Mikkelsen[1], Ivano E. Castelli[1], T. Vegge[1]

[1]Atomic Scale Modelling and Materials, Department of Energy Conversion and Storage, Technical University of Denmark, Denmark

Abstract: The water/Pt(111) interface is one of the most important interfaces in the field of electro-catalysis and has been studied in great detail using both static water-layer models[1] and computationally intensive force-models based on density functional theory[2].
This work will describe current efforts towards machine-learning a force-field that can describe the complicated interface. The data set used is based on the results of[2], who modelled the interface using ab initio molecular dynamics.



Taken from [2]: doi.org/10.1039/C8SC02495B

[1] J. Rossmeisl, J. K. Nørskov, C. D. Taylor, M. J. Janik and M. Neurock, J. Phys. Chem. B, 2006, 110, 21833–21839.
[2] H. H. Kristoffersen, T. Vegge, H. A. Hansen, Chem. Sci., 2018, 9, 6912-6921

Calculations of the ^{125}Te NMR chemical shift of diphenyl ditelluride

C. Muraro [a], M. Bortoli [a], M. Dalla Tiezza [a], G. Saielli [b], L. Orian [a]

[a] Dipartimento di Scienze Chimiche, Università degli Studi di Padova, Via Marzolo 1, 35131 Padova, Italy
[b] CNR Institute on Membrane Technology, Unit of Padova, Via Marzolo 1, 35131 Padova, Italy

E-mail: cecilia.muraro@studenti.unipd.it

Aromatic ditellurides are regarded as promising molecules that might serve as efficient antioxidants compounds in regulating the oxidative stress level inside the cell, that is they can mimic the activity of the ubiquitous selenoenzyme glutathione peroxidase. [1] However, an accurate analysis of their reactive mechanisms has yet to be completed and should be, together with a thorough investigation of their acute and long-term toxicity, carried out to obtain an exhaustive picture on how to exploit the potential of these molecules in small and easily synthesizable compounds for pharmacological and medicinal use. [2] Diphenyl ditelluride has shown the most promising features making it one of, if not the, most studied molecule among Te-based GPx mimics. [3] Nevertheless, investigation on this compound is computationally not trivial as this molecule presents easy conformational freedom (two conformers are shown in Figure 1), due to the presence of dihedral angles with low rotational barriers. In this work, the relevance of the different structures found in the conformational space spanned by diphenyl ditelluride is investigated with a computational approach based on scalar relativistic DFT. In particular, the focus is centered on the variations on the electron density around the tellurium atom upon modification of key geometrical parameters. These effects are assessed thorugh state-of-the-art *in silico* ^{125}Te NMR, using different computational approaches to establish how structural changes affect the chemical shift of ^{125}Te. Additionally, this investigation serves also as a benchmark to obtain a robust protocol for the calculation of accurate ^{125}Te NMR chemical shifts organotellurides exhibiting a high degree of flexibility.

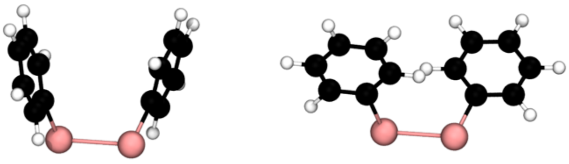


Figure 1 Two different stable conformations of diphenyl ditelluride.

[1] C. W. Nogueira, G. Zeni, J. B. T. Rocha, *Chem. Rev.* **2004**, *104*, 6255–6285.
[2] L. Orian, S. Toppo, *Free Radic. Biol. Med.* **2014**, *66*, 65–74.
[3] M. Bortoli, F. Zaccaria, M. Dalla Tiezza, M. Bruschi, C. Fonseca Guerra, F. M. Bickelhaupt, L. Orian, *Phys. Chem. Chem. Phys.* **2018**, *20*, 20874–20885.

Orbital optimization in coupled cluster theory

Rolf H. Myhre[a], Benedicte Ofstad[b] and Simen Kvaal[b]
[a]Department of Chemistry, NTNU, 7491 Trondheim, Norway
[b]Hylleraas Centre for Quantum Molecular Sciences,
Department of Chemistry, University of Oslo, 0315 Oslo, Norway

Coupled cluster (CC) theory is characterized by an exponential parametrization of the electronic wavefunction. This results in superior accuracy compared to a linear parametrization, but makes a variational formulation exceedingly complex and one instead adopts a bivariational formulation. Orbital optimized CC [1] and CC in an optimized active space [2] were developed in the late 90s. These models combine CC with a variational orbital optimization and have seen some success. A few years later, the bivariational nonorthogonal orbital optimized CC (NOCC) model was introduced [3]. However, it received little attention until the recent introduction of the orbital adapted CC (OACC) method [4]. This is a CC model in an active space with bivariational orbital optimization.

Implementing NOCC doubles (NOCCD) is relatively straightforward with the wavefunction parameters solved similarly to CCSD with T_1-transformed integrals and slightly more involved equations for the orbital parameters than for the singles amplitudes in CCSD. Consequently, NOCC converges rapidly using the standard DIIS solver and the results are very similar to CCSD. From limited testing, NOCC appears to converge about twice as fast as OCC.

The OACC model, on the other hand, is much more involved. Because it includes rotations between virtual orbitals, one cannot approximate the inverse Hessian with the difference between the diagonal elements of the Fock matrix. Furthermore, the magnitude of the diagonal elements of the Hessian can span four to five orders of magnitude, making preconditioning necessary. Adopting the GDM algorithm [5] to approximate the orbital-orbital block of the Hessian saw some success. However, convergence was still slow. The problem is similar to MCSCF and the orbital-wavefunction blocks of the Hessian are probably the most important for optimization of the virtual active space. This indicates that the full Hessian, or at least an approximation, is necessary for rapid convergence for OACC.

[1] C. D. Sherrill et al. *J. Chem. Phys.* 109.11 (1998), pp. 4171–4181. DOI: `10.1063/1.477023`.

[2] A. I. Krylov et al. *J. Chem. Phys.* 109.24 (1998), pp. 10669–10678. DOI: `10.1063/1.477764`.

[3] T. B. Pedersen, B. Fernández, and H. Koch. *J. Chem. Phys.* 114.16 (2001), pp. 6983–6993. DOI: `10.1063/1.1358866`.

[4] S. Kvaal. *J. Chem. Phys.* 136.19 (2012), p. 194109. DOI: `10.1063/1.4718427`.

[5] T. van Voorhis and M. Head-Gordon. *Mol. Phys.* 100.11 (2002), pp. 1713–1721. DOI: `10.1080/00268970110103642`.

Molecular Pseudo-potentials for the study of chemical properties and reactions

Alexander Punter,[a] **Yannick Carissan,**[a] **Paola Nava.**[a]

a) Aix-Marseille Université, Marseille, France;

E-mail: alexander.punter@etu.univ-amu.fr

Keywords: pseudo-potentials; π-systems; sp^2 carbon.

Pseudo-potentials are an established method of simplifying atomic and molecular calculations, and consist of replacing explicitly-treated particles, atoms or even groups of atoms with simple potential functions.

We aim to expand the range of uses of pseudo-potential methods available to reproduce the physical properties of molecules. We expect that such methods will be of benefit firstly through the significant reduction in complexity of molecular calculations, particularly for larger molecules, and secondly by allowing users to examine the behaviour of systems in which only certain electrons are explicitly treated e.g. π-orbital systems.

We have optimised a series of potential sets for a variety of simple organic molecular moieties, and find that many basic properties are well-replicated by the pseudo-system (HOMO energies, 1st ionisation energies and 1st excitation energies). These simple structures transfer well to much larger molecules. We find also that more advanced feats, including reasonable geometric optimisation, can be achieved with these pseudo-systems. Recently we have observed that these pseudo-systems can generate reasonable absorption spectra, in which individual excitations are clear and well-reproduced. Increasing the accuracy and versatility of these absorption spectra is our current focus.

Predicting new, simple, molecular inorganic species

Pekka PYYKKÖ[a]

[a]Department of Chemistry, University of Helsinki, POB 55 (A. I. Virtasen aukio 1),
FIN–00014 Helsinki, Finland. E-mail: *Pekka.Pyykko@helsinki.fi*

For predicting new, simple molecular species, one does not necessarily need artificial intelligence. A combination of chemical intuition, the isoelectronic principle, and any normal quantum chemical method will get one quite far.

As reviewed in [1], some successful examples were:

1. New transition-metal hydrides, MH_n.
2. New multiply-bonded species for Periods 2 to 3 of types A≡B, A=B=C, A=B=C=D or A≡B-C≡D, and A=B=C=D=E. These include the new cations N_5^+ and $OCNCO^+$.
3. New members of the uranyl isoeectronic series, such as NUO^+.
4. The uranyl-like series of $OUIr^+$. Observed [2]!
5. Systems with noble-metal–noble-gas chemical bonds, such as $AuXe^+$.
6. The first argon compound, HArF.
7. The cluster series of WAu_{12}.
8. Transition-metal centred polyazide ions.
9. Covalent molecules with a central -Zn-Zn- bond.
10. Tetrahedral clusters of zinc and cadmium.

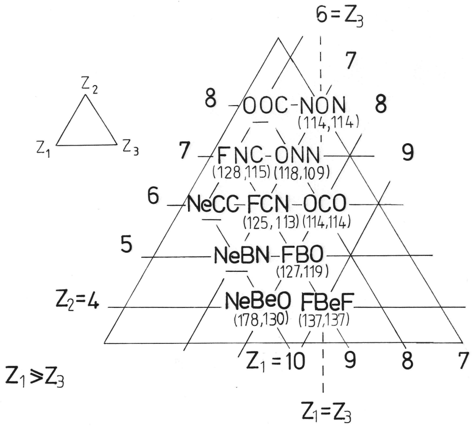


Figure 1: For three-component systems, ABC, complete maps can be made.

[1] 1 P. Pyykkö, *Phys. Chem. Chem. Phys.*, 2012, **14**, 14734-14742.

[2] 2 M. Santos et al., *Eur. J. Inorg. Chem.*, 2006, 3346-3349.

Theoretical study on the electronic structure and stability of different clusters of $(Sb_2O_5)nH_2O$ and $(Nb_2O_5)nH_2O$ coordinated to the glucose used or probed in antilieshmanial activity

Hassan Rabaa[1], Sara Elmoutaoikel[1], Dage Sundholm[2] and Andriy Grafov[3],

[1]Department of Chemistry, ESCTM, University Ibn Tofail, Po Box 133, 14000 Kenitra, Morocco.
[2,3] Department of Chemistry, University of Helsinki, P.O.B. 55 (A.I. Virtasen aukio 1), FIN-00014, Finland

Abstract:

Transition metal oxide clusters have drawn great attention due to their fundamental importance as well as the possibility of constructing nanostructure materials and devices using clusters as building blocks. In the current study we focused on elucidation of the dependence of stability of different hydroxylated Antimony or Niobium oxides clusters size and their interaction with organic molecules (glucose).

Molecular structures and relative stabilities of the different $(Sb_2O_5)nH_2O$ and $(Nb_2O_5)nH_2O$ clusters were investigated by *ab initio* calculations at MP2 and TPSS levels An extended basis sets for the treatment of Sb and Nb atoms with an effective core potentials available in the literature were used. Hydroxylated Niobium clusters are predicted to be more stable. Analysis of ΔG, ΔS and binding energies shows high interaction strengths of Sb-O and Nb-O with Cyclic Glucose system.

Descriptor design and selection for hydrogen evolution reaction

Yashasvi S. Ranawat[1], **Marc O. J. Jager**[1], **Eiaki V. Morooka**[1], **and Adam Foster**[1]

[1] Department of Applied Physics, Aalto University, Espoo, 02150, Finland

ABSTRACT

Catalyst design is a crucial aspect of hydrogen evolution reaction (HER). It is aided by rigorous simulations of catalytic action on suitable candidates. However, this is non-trivial due to an infinite search space, and computationally intensive DFT simulations. Here, machine-learning aided approaches can be applied to augment the database of catalyst, and also select the relevant catalyst. Since computers cannot inherently understand the rotational, translational, or permutational invariance of atomic representation, various representations of chemical environment — descriptors — are introduced[1]. The study compares the descriptors: SOAP[2], an electronic description based on SOAP and local adaptation of MBTR[3], to outline their applicability in predicting adsorption-energy for HER.

References

1. Jäger, M. O., Morooka, E. V., Canova, F. F., Himanen, L. & Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Comput. Mater.* **4**, 37 (2018).
2. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
3. Huo, H. & Rupp, M. Unified representation for machine learning of molecules and crystals. *arXiv preprint arXiv:1704.06439* (2017).

Using an Autoencoder for Dimensionality Reduction in Quantum Dynamics

Sebastian Reiter, Thomas Schnappinger, Regina de Vivie-Riedle

Department of Chemistry, Ludwig-Maximilians-Universität,
Butenandtstr. 5-13, Munich, Germany

Quantum dynamics is an important tool in the investigation of ultrafast (photo)chemical processes. The time-dependent Schrödinger equation is solved on a discrete spatial grid of nuclear coordinates. One of the main limitations of this approach is the fact that the number of grid points scales exponentially with the number of degrees of freedom of the molecule.

We can tackle this so called *curse of dimensionality* with a variety of coordinate reduction techniques, where the goal is to find a low-dimensional subspace of internal coordinates that describes the process in question. The most common way to construct such a subspace is to manually select suitable coordinates by chemical intuition. This can be achieved by employing either pure normal modes or linear combinations thereof as basis vectors for a low dimensional subspace. These basis vectors have to be carefully selected and their construction requires a high amount of *a priori* knowledge of the molecular system. Therefore automatic coordinate generation can be a helpful alternative to human intuition.

We present two different automatic workflows to construct linear [1] and non-linear [2] reduced dimensional subspaces for applications in grid-based quantum dynamics. In both methods, the first step is to sample the configuration space of the desired reaction pathway using semiclassical trajectories. The obtained data set can either be processed with a principal component analysis (PCA) to constructed linear coordinates or it can be used as input to an autoencoder network (see Figure 1) to find non-linear coordinates. As first examples we focus on proton transfer reactions in small molecules in the ground state.

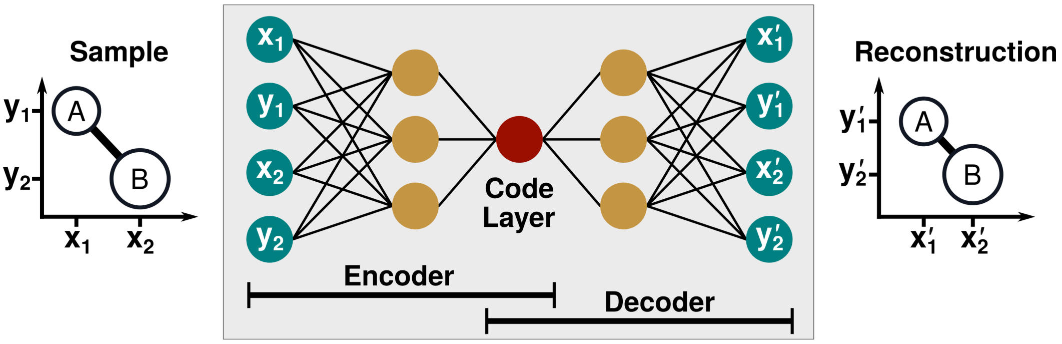


Figure 1: Schematic representation of an autoencoder used in coordinate reduction.

[1] J. P. P. Zauleck, S. Thallmair, M. Loipersberger, R. de Vivie-Riedle, *J. Chem. Theory Comput.* **2016**, *12*, 5698–5708.

[2] J. P. P. Zauleck, R. de Vivie-Riedle, *J. Chem. Theory Comput.* **2017**, *14*, 55–62.

Advancing structure-based drug-design with novel alchemical free energy calculation methods

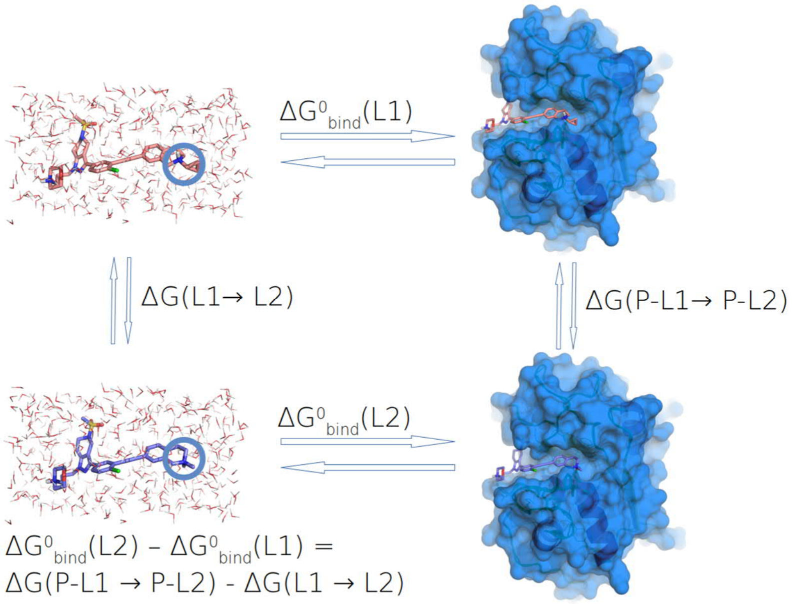
Jenke Scheen[a] , **Antonia Mey**[a], **Mark Mackey**[b], **Julien Michel**[a]

[a]*EaStChem School of Chemistry, Joseph Black Building, University of Edinburgh, Edinburgh, EH9 3FJ, United Kingdom.*
e-mail: j.scheen@sms.ed.ac.uk
[b]*Cresset Group®, New Cambridge House, Bassingbourn Road, Litlington, Cambridgeshire, SG8 0SS, United Kingdom.*

Keywords: free energy perturbation, molecular dynamics, protein-ligand binding

In efforts to effectively quantify drug-receptor affinities *in silico*, computational chemists possess and continue to develop a wide, diverse range of molecular modeling techniques. However, an ever-present caveat remains the trade-off between computational expensiveness (processing time) and prediction accuracy. Alchemical free energy (**AFE**; or free energy perturbation, **FEP**) methods are increasingly used in academia and industry to support the ligand optimisation problem; their scope, however, has been largely limited to small libraries in hit-to-lead optimisation scenarios due to their computational cost and the accuracy limitations of classical force-fields used to compute drug-receptor energetics.

A recent publication by our group showed that even a simple regression model can serve as a corrector to FEP calculations in order to increase accuracy[1]. The goal of this project is to develop and apply a DNN-based model that serves as a diverse correction term to AFE quantification steps. Empirical correction terms will be learned by mining previously generated AFE datasets. With this, we aim to deliver a methodology that combines the best aspects of physics-based and data-driven methods to predict binding affinities in a hybridised model.



Visual representation of a thermodynamic cycle that free energy perturbation (FEP) methods rely on. Illustrated are a given ligand 1 (L1, top-left) and the target receptor (P, top-right) and a second ligand (L2, bottom-left) and the target receptor (P, bottom-right). Instead of calculating the L1 or L2 binding energy ($\Delta G^0_{bind}(L1,2)$) directly, the perturbation energies of adding a cyclopropyl to L1 to obtain L2 in both solution (left column) and complex (right column) can produce an accurate binding energy approximation for either L1 or L2 binding P, at a fraction of the computing cost.

[1] Papadourakis, M., Bosisio, S. & Michel, J. J Comput Aided Mol Des (2018) 32: 1047. https://doi.org/10.1007/s10822-018-0154-6

Electronic excitation spectra calculation of large systems (> 1000 atoms) with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB).

Jakob Seibert[1]**, Christoph Bannwarth**[1]**, Stefan Grimme**[1]

[1]*MCTC, University of Bonn, Beringstr. 4, Bonn, Germany*

The computation of electronic absorption or circular dichroism (CD) spectra of molecules with a system size of 1000 atoms or more is challenging. However, CD spectroscopy is an important tool for structure elucidation of biomolecular systems or characterization of transition metal complexes. We previously developed the simplified variants of Tamm-Dancoff approximated (sTDA)[1] and time-dependent (sTD)[2] density functional theory which are based on a regular ground state Kohn-Sham determinant (obtained with hybrid or range-separated hybrid functionals[3]). This approach enables the computation of electronic spectra with TD-DFT accuracy for molecules up to a few hundred atoms. The computation of the ground state is the computational bottleneck. Our recently developed tight-binding method using an extended AO basis set (xTB)([4]) solves this problem and enables the ultra-fast computation of electronic spectra by sTDA-xTB. Computed spectra for biomolecular sytems (> 1000 atoms), e.g, for entire exemplary proteins and DNA fragments, as well as large transition metal complexes are presented.

[1] S. Grimme, *J. Chem. Phys.* **138**, 244104 (2013).

[2] C. Bannwarth, S. Grimme, *Comput. Theor. Chem.* **45**, 10401041, (2014).

[3] T. Risthaus, A. Hansen, S. Grimme, *Phys. Chem. Chem. Phys.* **16**, 14408 (2014).

[4] S. Grimme, C. Bannwarth *J. Chem. Phys.* **145**, 054103 (2016).

High-throughput forecasting of molecular properties in solution

Philipp Eiden, Andrew Palmer, Volker Settels

Digitalization in Research & Development, BASF SE, Carl-Bosch-Str. 38,
67056 Ludwigshafen am Rhein, Germany

Computational modelling of chemical behavior is a fundamental part of materials research and development at BASF. It is key, not only to designing novel high-performing chemistries with the specific physical requirements desired by our customers but also to rationally developing the efficient processes steps needed for economic production. As most products of BASF are applied in liquids or amorphous materials, QSPR models have proven to be very potent, which rely on computed molecular properties in solution.

The value of QSPR models originates from massive *in silico* compound screening, which requires large databases. We constrain our molecule catalogue to those small molecules which form the building blocks of commercially available chemicals, in contrast to databases of hypothetical molecules with no clear synthetic route. This still provides a rich diversity of chemistry with additional information such as price, availability and toxicity. Molecules are selected for calculation using diversity criteria to ensure that we achieve high coverage of the molecular space.

For each molecule in the database, the following computational procedure is followed: First, an extensive conformational search is performed on DFT level of theory. Free Gibbs energies in solution are computed by COSMO-RS theory for many solvents. Afterwards, a conformer ensemble is selected, which includes all conformers important in at least one solvent. Solvent-dependent molecular properties are computed ranging from polarizabilities and multipole moments to H-bond interaction enthalpies and partition coefficients. To date, our database comprised more than 100 000 molecules in over 60 different surroundings. Approximately 8 000 properties are calculated per molecule. We believe this is the largest and most diverse database ever generated for quantum chemically computed properties in solution.

However, our chemical space contains millions of molecules, and we would like to screen all available chemistries. Therefore, we have performed a large-scale evaluation of existing machine learning models on computed molecular properties in our database including cutting-edge deep learning approaches [1]. We find that the deep learning approach outperforms baseline fingerprint based methods and we routinely achieve regression determination coefficients which comes close to the accuracy of the underlying computed properties compared to experimental properties. This result enables us to extend our QSPR models to provide comprehensive molecules forecasts across our chemical space.

[1] Wengong Jin, Regina Barzilay, Tommi Jaakkola Junction Tree Variational Autoencoder for Molecular Graph Generation https://arxiv.org/abs/1802.04364

Accelerating geometry optimisations with machine learning

Robert A. Shaw, Adam N. Hill, and J. Grant Hill

Department of Chemistry, University of Sheffield, Sheffield, S3 7HF, UK

rashaw1@sheffield.ac.uk

A standard procedure in quantum chemical calculations is to build an initial guess of a molecule, or molecular system, then optimise the geometry. This process can be time consuming and frustrating, and as increasingly large systems are looked at, can lead to serious problems. Moreover, the iterations involved in such optimisations can be an order of magnitude more computationally expensive than a single-point energy calculation, especially if analytic derivatives are not available. It would therefore be very useful to be able to streamline this process, both more efficiently generating an initial guess, and minimising the number of optimisation steps required.

Large quantities of high-level structural data for small molecules are becoming increasingly easy to obtain [1, 2]. Here, we use machine learning approaches to generate 3D geometries from molecular graphs with only connectivity and bond-order information. We demonstrate the proof-of-concept for small organic molecules, showing extensibility to larger versions of similar systems. With further work, this approach could provide a useful tool, both for practical and outreach purposes.

References

1. M. Rupp *et al.,* Phys. Rev. Lett. 108, 58301 (2012)
2. J. S. Smith, O. Isayev, and A. E. Roitberg, Scientific Data 4, 170193 (2017)

HPC-Europa3: Travel and collaborate with EC funding

A. Sillanpää
CSC - IT Center for Science Ltd., P.O. Box 405, 02101 Espoo, Finland
e-mail: atte.sillanpaa@csc.fi

The HPC - Europa3 (HPCE3) programme funds researcher visits, provides computational resources and support from the supercomputing centre of the target country (see map). Visitors can come from any EU country or associated state and a small quota is available also for researchers from elsewhere. HPCE3 is an excellent opportunity to start new or strengthen existing collaboration. Reasonable travel and accommodation costs are reimbursed including a small daily allowance. The visits can last 3–13 weeks.

HPCE3 will continue until April 2021. The goal is to support more than one thousand visits. Based on recent feedback the amount of bureaucracy is low and the acceptance ratio has been high: In the first five calls all applications to visit a research group in Finland have been accepted. More than half of these visits have concerned chemistry or material science.

The programme is open to researchers of any level, from PhD students to full professors, in any field requiring computational resources. Generally, the visitor will not work in the target country's computing center but will be hosted by a research group. Applications always concern academic research. The researchers of universities, universities of applied sciences and research institutes, but also small and medium size companies can apply for a visit or act as a host.

Applications can be done at any time via the on-line portal[1] and are processed four times a year. The call deadlines can be found at the HPC-Europa3 website[2]. About a thousand host groups have already signed up and can be browsed (by country) in the application form. If you want to receive a visitor, sign up as a host at the on-line portal for hosts[3].

Applications are evaluated for suitability for HPCE3 (visit brings additional value wrt. using resources remotely, it aims at starting new or strengthening existing collaboration, research needs HPC or resources not available locally, ...) and for scientific quality of the research plan. If you are uncertain whether your project matches the criteria, please contact us.

HPC-Europa3 receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No.730897

References

[1] https://hpc-europa.cineca.it/researcher/login
[2] http://www.hpc-europa.eu/
[3] https://hpc-europa.cineca.it/host/login

Machine learning and materials science: an overview of applications

M. Sluydts[1,2], M. Larmuseau[1,2], T. Crepain[1], K. Dumon[1], K. Lejaeghere[1,3] and S. Cottenier[1,2]

[1] Center for Molecular Modeling, Ghent University, Belgium
[2] Department of Electrical Energy, Metals, Mechanical Constructions & Systems, Ghent University, Belgium
[3] Department of Applied Physics, Ghent University, Belgium
email: michael.sluydts@ugent.be

Material science heads evermore towards the study of materials both complex in structure and properties. Studying these new materials requires an efficient combination of theoretical and experimental research. One way of achieving this is by using machine learning methods which allow us to extract and utilize essential information from both types of data. On this poster we will discuss several examples of how machine learning can help us accelerate materials research.

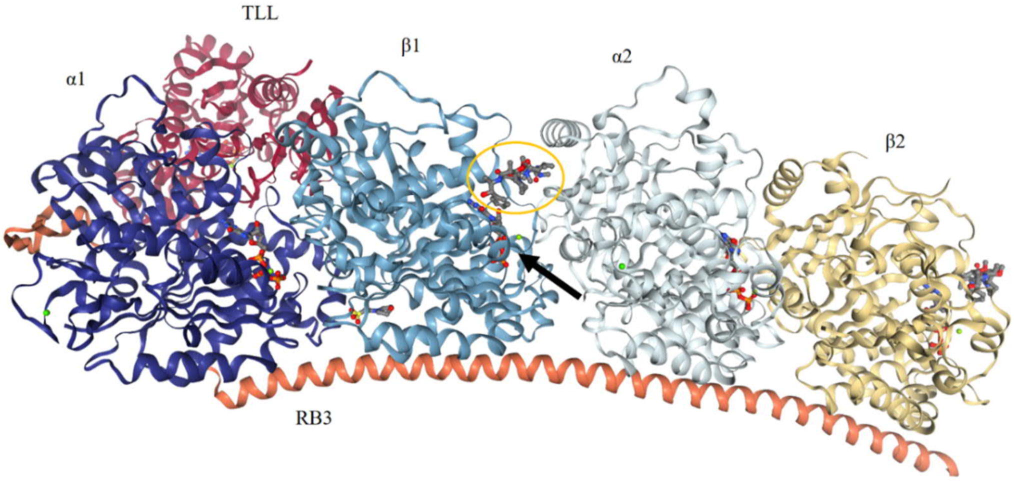
In theoretical research we are faced with an eternal tradeoff between computational resources and the detail with which we wish to study our atomistic systems. For example, we may have to choose between studying a single system with compositional variations, such as functional groups or defects, or studying the time evolution of one of these systems to obtain time and temperature-dependent properties. Regardless of the choice made, the computational time will often be the bottleneck of what we can achieve. Machine learning offers several ways of reducing the computational time required and thus extending what ab initio methods can achieve. First we discuss the case of high-throughput screening where a compositional space is explored to obtain either trends or find individual interesting materials. Predictions are made for desired properties by combining automated ab initio screening with live training of a surrogate machine learning model. Often calculations for a mere 10% of the screening space are already sufficient to obtain the desired information. A second application investigates the use of geometric deep learning models to process larger datasets and create atomistic potentials which can be for geometric relaxation and molecular dynamics.

Machine learning can also be used in experimental research. We provide an illustration of this in the form of microstructural analysis through optimal and scanning electron microscopy. Individual images are clustered into material classes using both a classical convolutional neural network approach as well as a more novel triplet network. In both cases, a compact representation of the image is constructed which can be used either as tool for visualization or as input for property models. In the future this type of information reduction may also allow us to combine theoretical and experimental data in single models. Regardless, the future of materials science and machine learning looks bright.

Improving the efficiency of the auristatin cancer-drug family

Iris Sokka,[1] Filip Ekholm,[1] and Mikael P. Johansson [1,2]
[1]Department of Chemistry, University of Helsinki, Finland
[2]Helsinki Institute of Sustainability Science (HELSUS), Finland
iris.sokka@helsinki.fi

The drug molecules monomethyl auristatin E (MMAE) and monomethyl auristatin F (MMAF) target, specifically, cancer cells, and work by preventing mitosis. In solution, the auristatins exist in an equal mixture of two conformers, of which only one is biologically active [1]. Isomerisation between the two forms is slow, which diminishes both the efficiency and safety of the drugs. Here, we show that a simple one-atom replacement shifts the isomer equilibrium almost completely towards the active conformation [2].



[1] M.P. Johansson, H. Maaheimo, F. Ekholm, "New insight on the structural features of the cytotoxic auristatins MMAE and MMAF revealed by combined NMR spectroscopy and quantum chemical modelling", Sci. *Rep*. **7** (2017) 15920.
[2] I. Sokka, F. Ekholm, M.P. Johansson, in preparation

GFN2-xTB: An extended Tight-Binding Quantum Chemical Method for Structures, Frequencies and Noncovalent Interactions across the Periodic Table

S. Spicher[1] , **C. Bannwarth**[2], **S. Ehlert**[1], **S. Grimme***[1]

[1]*University of Bonn, Mulliken Center for Theoretical Chemistry,*
Beringstr. 4, 53115 Bonn, Germany
[2]*Stanford University, Department of Chemistry,*
333 Campus Drive, Stanford, CA 94305, USA

A new semiempirical tight-binding model is presented, which is primarily designed for the fast calculation of structures and noncovalent interaction energies for molecular systems with roughly 1000 atoms. The essential novelty in this GFN2-xTB [1] scheme is the inclusion of anisotropic second order density fluctuation effects via cumulative atomic multipole moments and the atomic partial charge dependent D4 London dispersion model. Without a significant increase in the computational demands, this results in a less empirical and overall more physically based method, which does not require any additional halogen or hydrogen bonding corrections. The GFN2-xTB method performs overall excellent for the desired "target" properties including geometry optimizations, vibrational frequency calculations and noncovalent interactions. It often shows even lower errors for "off-target" properties such as reaction barrier heights and molecular dipole moments. This reflects the less empirical nature and the overall improved physics of the GFN2-xTB method, which exclusively relies on element-specific and global parameters. As with the precursor GFN-xTB [2], all elements up to radon (Z = 86) are currently covered.

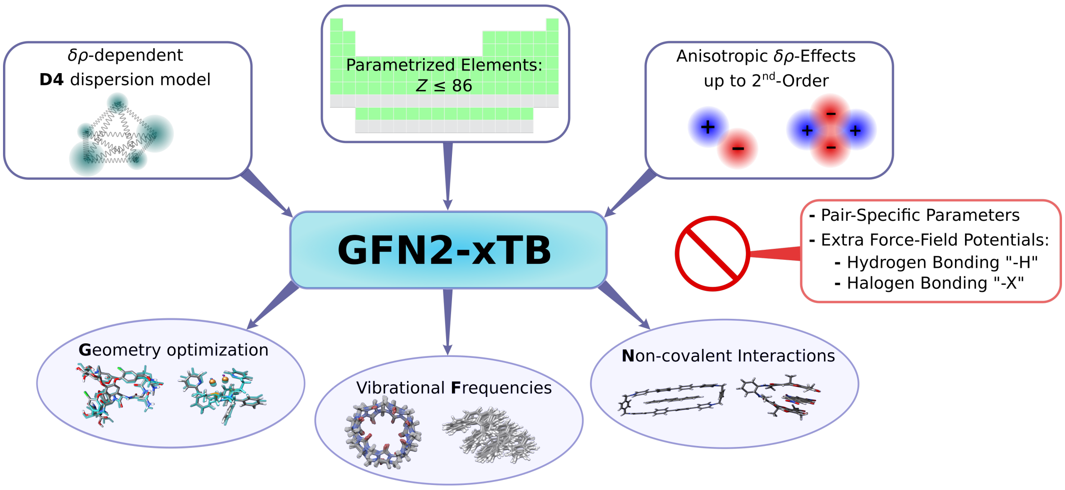


Figure 1: GFN2-xTB, a semiempirical tight-binding method for geometry optimizations, vibrational frequencies and noncovalent interactions.

[1] C. Bannwarth, S. Ehlert, S. Grimme *J. Chem. Theory Comput.*, **2018**, submitted.

[2] S. Grimme, C. Bannwarth, P. Shushkov *J. Chem. Theory Comput.*, **2017**, *13*, 1989-2009.

Surface Chemistry Predictions by DFT-Descriptors and Machine Learning - Research Outline

Joakim Halldin Stenlid, Lars GM Pettersson
Department of Physics, Stockholm University, Stockholm, Sweden
email: joakim.halldin-stenlid@fysik.su.se

We aim to make accurate but computationally inexpensive predictions in surface chemistry through machine learning on a set of materials-properties readily computed by DFT. These properties include new and established “reactivity descriptors” belonging to the MSPA (Molecular Surface Property Approach) family. The MSPA descriptors based on, e.g., energy-weighted densities-of-state and the electrostatic potential, have been found to possess excellent predictive capabilities for chemical interactions including reactions when used on congeneric series of molecules, nanoparticles and materials surfaces. The local nature of the descriptors allows predicting the chemical behavior of multiple sites of a compound from one single DFT calculation. Thus, fewer computations are needed and larger systems can be studied compared to traditional approaches. To fully capture a chemical interaction, it is in many cases necessary to combine two or more properties. In the simplest cases, for systems and processes of high similarity, a linear combination is often sufficient. However, the fundamental nature of the properties suggest that general relations should exist over a larger span of compounds and processes. These relations are more complex, but we anticipate that they can be established by a machine-learning approach. Here we summarize previous work on MSPA descriptors and outline our intended research direction in connection with machine learning. We anticipate that our work will be of use in for instance heterogeneous catalysis.

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Computational Investigation of boron-Subphathalocyanines for Dye Sensitized Solar Cell application

F. E. Storm, T. Hansen, and K. V. Mikkelsen

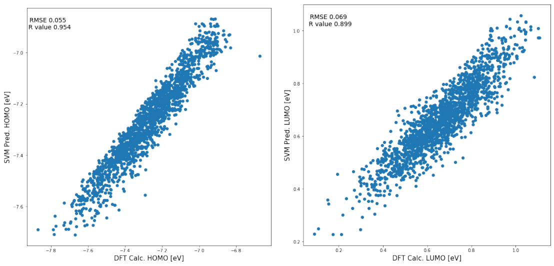
University of Copenhagen, Department of Chemistry, Center for Exploitation of Solar Energy, Denmark

`freja.storm@chem.ku.dk`

The goal of the presented work is to use machine learning for predicting promising boron-Subphthalocyanine (SubPc) candidate molecules for Dye-Sensitized Solar Cell (DSSC) application. To have a high efficiency of the DSSC, the dye must have HOMO/LUMO levels that are well aligned with the redox energies of the electrolyte and the conduction band of the semiconductor conduction. Furthermore, the dye must of cause have a high-intensity absorption in the visible part of the spectrum.
We show how the combination of Supported Vector Machine (SVM) and the electrotopological state index (E-state index) can be used to predict the HOMO and LUMO energies of SubPc based dyes.
A database of 10.000 SubPc structures functionalized in the periphery has been created via PM6 and DFT calculations. From the TD-DFT calculation, the HOMO orbital energy and longest wave-length transitions are used as labels in the SVM treatment of the structures.

SVM learning of HOMO energies

The E-state index was calculated via RdKit (Open-source cheminformatics; http://www.rdkit.org). From the original E-state index, empty columns were removed, and this reduced vector was used as the feature vector.



Since it is well known that the LUMO orbital often is poorly represented with DFT molecular orbital energies, the LUMO is here defined as the HOMO energy + TD-DFT excitation energy.

Conclusion

We have constructed a database of 10.000 PM6 optimized SubPc based dyads and triads. Via DFT the physical properties of these structures have been calculated. Based on 8.900 structures a preliminary investigation of the applicability of machine learning for predicting the HOMO and LUMO energies of the triads have been investigated. Via the SVM formalism, the connection between the predicted and DFT calculated molecular orbital energies are most reliable for the HOMO energies.

QSAR modelling of HIV-1 attachment inhibitory activity of azaindole derivatives

Joanna Stoycheva[1], Ismail Hdoufane[2], Katarina Josifovska[3], Menče Najdoska-Bogdanov[3], Galia Madjarova[1], Jane Bogdanov[3], Alia Tadjer[1], Driss Cherqaoui[2]
[1] Faculty of Chemistry and Pharmacy, Sofia University "St. Kliment Ohridski"
[2] Department of Chemistry, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco
[3] Institute of Chemistry, Faculty of Natural Science and Mathematics, Ss. Cyril and Methodius University, Skopje, Macedonia

Human immunodeficiency virus (HIV-1) glycoprotein 120 (gp120) is one of the key targets for Acquired immunodeficiency syndrome (AIDS). A large number of inhibitors are being designed for this target in order to find safe and effective drugs.

In the present study, we have developed QSAR models for HIV-1 attachment inhibitors based on data for 128 azaindole analogs using suitable molecular descriptors calculated by Dragon software. Chemometrics methods including artificial neural network (ANN) and support vector machine (SVM) were used to set up QSAR models in order to explain the structural requirements of HIV-1 gp120 inhibitory activity. The contribution of each descriptor to the structure-activity relationships was evaluated. The correlation coefficients for both the training set and test set reveal that the SVM model is slightly superior to the ANN one. The results obtained indicate that both models can be successfully used to predict the anti-HIV-1 activity of the studied compounds. These computational studies have potential application in the prediction of biological activity for the newly synthesized compounds.

Semiclassical dynamics with constant wavelength: ML opportunities

Jiří Suchan[1], Petr Slavíček[1]

[1]Department of Physical Chemistry, University of Chemistry and Technology, Prague, Czech Republic

Ab initio photodynamical simulations gradually become a mature field, allowing for a routine simulations of light-induced processes. Most of the photodynamical simulations are performed within the framework of semiclassical dynamics. Setting the correct initial conditions for the atomic nuclei is then critical for obtaining meaningful results. Usually, the simulations are based on sampling the initial state e.g. within Wigner transform of the initial vibrational function. This sampling corresponds to a vertical excitation of the initial wavepacket, multiplied by transition dipole moment. Such simulations protocol, however, corresponds to photochemical process triggered by ultrashort laser pulse, which is typically not a realistic situation for most of the experiments. In fact, most of the experiments are performed with a specific wavelength and a specific eigenstate on the excited state potential energy surface should thus be formed.

We propose a new method for sampling initial nuclear densities in case of continuous wave (CW) laser experiments. Molecular dynamics with constrain of constant excitation energy is introduced by using the means of Lagrange multipliers. Target excitation energy then corresponds to the laser wavelength. The method is tested on CCl_2F_2 molecule also known as Freon-12. Kinetic energy spectra of the Cl fragment are computed using surface hopping simulations using initial conditions provided by our method. The results closely match the experimental data.

We further discuss the opportunities to employ the machine learning (ML) techniques. One possibility is to utilize the kernel ridge regression to correct for the dynamical correlation contribution to the potential energy surface, affecting especially the dissociation energetics. This could be implemented by learning the difference between cheap and high-end *ab initio* methods. Another one is to automatically recognize the main geometric features, which influence the excitation energies of a molecule.

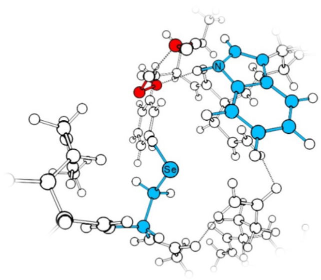
The oxidative step in the catalytic mechanism of glutathione peroxidases: insights from quantum chemistry

Marco Dalla Tiezza,[a] *F. Matthias Bickelhaupt,*[b,c] *Laura Orian*[a]

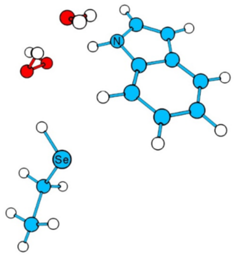
[a] *Dipartimento di Scienze Chimiche Università degli Studi di Padova Via Marzolo 1 35131 Padova*
[b] *Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling (ACMM), Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam (The Netherlands)*
[c] *Institute for Molecules and Materials (IMM), Radboud University, Heyendaalseweg 135, 6525 J Nijmegen (The Netherlands)*

An ineffective control of the concentration of reactive oxigen and nitrogen species (ROS and RNS) has been proved to be one of the main causes of severe pathologies. Commonly, the so-called oxidative stress, is kept at bay by several enzymes among which, the ubiquitous glutathione peroxidases (GPx).[1] These proteins are able to process and reduce hydroperoxides using glutathione (GSH) as cofactor. A peculiar feature of GPx is its redox-active residue, i.e. a selenocysteine. In particular, its initial selenol form is oxidized to selenenic acid with subsequent reduction of the harmful hydroperoxide substrate. The regeneration of the initial reduced form of GPx takes place consuming two GSH equivalents and producing water and the oxidized form of glutathione (GSSG). The key role of Se is fundamental since the enzymatic activity of the GPx sulfur mutant is drastically compromised compared to the selenium based GPx.[2-3] Unfortunately, the advantage of selenium over sulfur is still not clear, although it is likely rooted in the different redox properties of the two chalcogens. The aim of this work is to unravel an important issue in the oxidative step of the GPx mechanism, i.e. the oxidation of selenol to selenenic acid using a modern and accurate computational approach. The whole enzyme has been shrinked to contain the higly conserved residues and few other fundamental functional groups in models of the GPx catalytic pocket of increasing complexity. Thermodynamics and kinetics data have been collected and are discussed to better understand the intriguing difference in activity of GPx and its S mutant.

GPx4



GPx4 catalytic pocket



Model system of the GPx4 catalytic pocket

References:

(1) L. Orian, G. Cozza, M. Maiorino, S. Toppo, F. Ursini, *Glutathione*, CRC Press, 2018, pp. 53–66.
(2) Toppo, S.; Flohé, L.; Ursini, F.; Vanin, S.; Maiorino, M. Biochim. Biophys. Acta BBA - Gen. Subj. **2009**, 1790 (11), 1486–1500.
(3) Orian, L.; Mauri, P.; Roveri, A.; Toppo, S.; Benazzi, L.; Bosello-Travain, V.; De Palma, A.; Maiorino, M.; Miotto, G.; Zaccarin, M.; Polimeno, A.; Flohé, L.; Ursini, F. Free Radic. Biol. Med. **2015**, 87, 1–14.

Machine learning for excited-state molecular dynamics

Julia Westermayr[1], Michael Gastegger[2], Maximilian, F. S. J. Menger[1,3], Sebastian Mai[1], Leticia González[1], Philipp Marquetand[1]

[1]Institute of Theoretical Chemistry, University of Vienna, Währinger Straße 17, 1090 Vienna, Austria
[2]Machine Learning Group, Technical University of Berlin, 10587 Berlin, Germany
[3]Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via G. Moruzzi 13, 56124 Pisa, Italy

Excited state ab-initio molecular dynamics (AIMD) is a powerful tool to investigate photo-induced processes in molecules, but it is seriously limited by the cost of the underlying on-the-fly quantum chemical calculations, hampering its application for long timescales. In this work, we face this bottleneck by combining conventional AIMD with machine-learning algorithms to enable accurate photodynamics on nanosecond time scales, which are otherwise out of reach with contemporary approaches [1].

In detail, we employ the surface hopping molecular dynamics program suite SHARC (Surface Hopping including ARbitrary Couplings) [2, 3], in which the demanding quantum chemistry calculations have been replaced with predictions made by multi-layer feed-forward neural networks (NNs). We generate training data with quantum chemistry programs via grid sampling and an adaptive sampling scheme [4]. This procedure is applicable to any molecule and does not need any information about excited state properties or critical points in advance. To guarantee that NNs predict properties correctly during dynamics simulations, we apply an active learning approach similar to reference [5] by using two NNs and comparing their outputs on-the-fly. Results are shown for an analytical model and a test molecule with 12 degrees of freedom.

[1] J. Westermayr, M. Gastegger, M. F. S. J. Meger, S. Mai, L. González, P. Marquetand, arXiv:1811.09112, Nov. 2018.
[2] M. Richter, P. Marqeutand, I. S. Jesús González-Váquez, and L. González *J. Chem. Theory Comput*. **2011**, *7*, 1253-1258.
[3] S. Mai, P. Marquetand, and L. González L. *WIREs Comput Mol Sci.* **2018**, *8*, e1370.
[4] M. Gastegger, J. Behler, and P. Marquetand *Chem. Sci.* **2017**, *8*, 6924–6935.
[5] J. Behler *Int. J. Quantum Chem.* **2015**, *115*, 1032-1050.