doi:10.2533/chimia.2019.983 Chimia 73 (2019) 983–989

Quantum Chemistry Meets Machine Learning

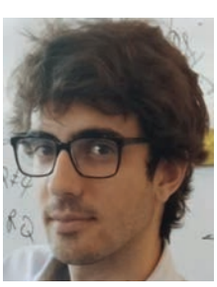
Alberto Fabrizio, Benjamin Meyer, Raimon Fabregat, and Clemence Corminboeuf*

Abstract: In this account, we demonstrate how statistical learning approaches can be leveraged across a range of different quantum chemical areas to transform the scaling, nature, and complexity of the problems that we are tackling. Selected examples illustrate the power brought by kernel-based approaches in the large-scale screening of homogeneous catalysis, the prediction of fundamental quantum chemical properties and the free-energy landscapes of flexible organic molecules. While certainly non-exhaustive, these examples provide an intriguing glimpse into our own research efforts.

Keywords: Catalysis · Free-energy landscapes · Machine Learning · Quantum Chemistry

Originally from France, ***Benjamin Meyer*** received his BSc and MSc in chemistry from Strasbourg University. In 2013, he moved to Nancy to work with Dr. Alessandro Genoni and Dr. Manuel Ruiz-Lopez; in 2016 he received a PhD in computational chemistry from Lorraine University for his work on linear scaling methods and X-ray constrained wave function methods. Benjamin then joined Prof. Clemence Corminboeuf's laboratory at École Polytechnique Fédérale de Lausanne in Switzerland as a postdoctoral research fellow. He is mainly working on molecular scalar fields and machine learning applications in quantum chemistry.

Born in Italy, ***Alberto Fabrizio*** received his BSc and MSc in chemistry from the École Polytechnique Fédérale de Lausanne (EPFL). In 2016, he joined the group of Prof. Clemence Corminboeuf at the EPFL as a PhD student. His current research focuses on characterizing and minimizing the errors of density functional approximations in new chemical situations and exploring the applicability of machine-learning models to access charge density information.



Raimon Fabregat received his BSc in theoretical physics at the University of Barcelona and his MSc at the École Normale Supérieur de Lyon in Computational Physics/Complex Systems Modelling. He joined the Corminboeuf group as a PhD student in 2017. His current research focuses on combining statistical sampling methods with quantum chemistry and machine learning models to speed up the free energy landscape exploration of fluxional organic molecules at high accuracy.

Clemence Corminboeuf started her independent career as a tenure track assistant professor at the EPFL in 2007. She was promoted to associate professor in 2014 and full professor in 2019. In 2010, she received the silver medal at the European Young Chemist Award. She was awarded two ERC grants (starting and consolidator), received the Werner Prize of the Swiss Chemical Society in 2014 and the Theoretical Chemistry Award from the ACS Physical Chemistry Division in 2018. She co-authored more than 150 publications focusing on the development of electronic structure methods as well as conceptual tools targeted for applications in the fields of organic electronics and homogeneous catalysis. Picture credit Alain Herzog, EPFL.

1. Introduction

The holy grail of computational/theoretical chemistry is delivering methods that provide fundamental breakthroughs prior to experiment.[1] Given this consideration, our work is oriented towards developing quantum chemical methods and conceptual tools aimed at discovering or predicting new catalysts and molecular electronic materials. Excitingly, the field of quantum chemistry is currently experiencing various paradigm shifts (*e.g.* machine learning along with big data analysis, GPU accelerated software) while potentially awaiting others (*e.g.,* quantum chemistry on quantum computers). Machine learning (ML) methods are flourishing tremendously in quantum chemistry simulations, inferring predictive models for ground-state molecular properties,[2] potential energy surfaces,[3] molecular forces,[4] infrared spectra,[5] electron densities,[6] density functionals,[7] and response properties like polarizabilities.[8] They promote the routine use of highly-accurate quantum chemical methods[2a] and accelerate the exploration of vast chemical spaces for myriad applications, including the discovery of new potential catalysts,[9] the determination of molecular conformers or polymorphs stability,[10] and the design of novel synthetic pathways.[11]

With a few exceptions[2,4,7] the enormous potential of applications of machine learning models has been relatively slow to impact the field of quantum chemistry. We perceive three main

*Correspondence: Prof. Dr. C. Corminboeuf, E-mail: clemence.corminboeuf@epfl.ch
Laboratory for Computational Molecular Design, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne

reasons for this delay: First is the inevitable byproduct of a community clash between machine learning experts and chemists. The first category of scientists lacks adequate knowledge of the key societal and fundamental chemical problems and do not communicate easily with chemists. Chemists, on the other hand, were not always conscious of the latest data-related technology and do not necessarily understand and appreciate what actually entails a machine learning model and in what precise context one might use it. These cross-community limitations can be successfully overcome through national or international synergic networks reuniting diverse scientific communities. A second reason for the delayed impact of ML-quantum chemistry relates to the reality that many quantum chemists adopt the mindset that new methods must be developed based on fundamental physics. In other words, they naturally disregard (or view with a high degree of skepticism) models developed based on statistics or data-driven approaches. Finally, chemists (both theorists and experimentalists) are used to dealing with a fairly small number of data points (<100) and measurements, which are not readily compatible with ML training procedures.

Below, we outline our recent efforts in bridging fundamental quantum chemistry with machine learning techniques. In particular, we illustrate how supervised learning techniques have transformed the scale, complexity, and the nature of the problems tackled.[1,12–14] We begin by discussing how machine learning models can be exploited to move from small- to large-scale screening of homogeneous catalysts. Our second ML effort demonstrates how fundamental, albeit complex, quantum chemical properties can be obtained at a negligible cost. In the third section, we share our thoughts on future challenges for computational organic chemistry and elaborate on the benefits of using machine learning models to achieve both converged statistical sampling and accurate quantum chemistry. The reader should note that this account should not be considered as an exhaustive review of machine learning in quantum chemistry, but rather as a highlight of our recent efforts aimed at integrating the latest developments into our research over a range of diverse applications.

2. Scaling Up

The search for catalysts with enhanced performance characteristics remains a backbone of modern chemical research. Once a problem tackled solely in the synthetic lab, today modern computational power coupled with advances in quantum chemistry have opened the field of catalysis to computational analysis with our research providing pertinent examples.[13] Computations not only lead to an enhanced understanding of existing reaction systems, but are also used to establish general concepts for the design of new catalysts. Ideally, the discovery of new catalytic reactions should benefit from the latest advances in quantum chemistry and also from machine learning models. Large-scale data analyses and databases resulting from quantum chemistry/ML combinations uncover information and relationships that are likely to be missed with commonly used small screening procedures. One of our research lines consists in developing frameworks that enable these ML-based large-scale screenings in order to assist in the design of molecular catalysts.

Until recently, applications of machine learning to homogeneous catalysis remained exceedingly rare.[15] Examples include the predictions of outcomes for C–N coupling reactions and deoxyfluorinations reaction with random forest models,[16] as well as predictive modeling for the chiral phosphoric acid-catalyzed thiol addition to N-acylimines exploiting deep feed-forward neural networks.[17] Other interesting contributions by Kulik *et al.* use deep neural networks and kernel ridge regression for predicting quantum mechanical properties *(e.g.* spin-state splittings, their sensitivity to HF exchange, spin-state specific bond lengths),[18] metal-oxo formation energies,[19] and electronic structure calculation outcomes.[20] Our recent collaboration with the von Lilienfeld group[9a] also constitutes a pioneering computational illustration merging these two fields (*i.e.* homogeneous catalysis and machine learning).

Our basic premise was to establish a broadly applicable toolkit based on the Sabatier principle,[21] which can be easily combined with machine learning models in order to enable large-scale screening (~10^4 catalysts) virtually instantaneously. Our fundamental toolkit exploits volcano plots, which are traditionally used in fields of heterogeneous catalysis and electrochemistry for identifying highly active catalysts.[22,23] These plots pictorially represent the aforementioned Sabatier's principle,[21] which states that the interaction between a catalyst and a substrate should be neither too weak nor too strong. They rely upon an easily determined (experimentally or computationally) descriptor variable (such as the binding energy between a substrate and a catalyst, x-axis of the volcano plot in Fig. 1) to ascertain a property of interest (overpotential,[24] turnover frequency,[25] *etc.*) related to the performance of different catalysts. The underlying mathematical correlations used to create volcano plots are linear free energy scaling relationships,[26] which arise owing to the dependence of the free energies of

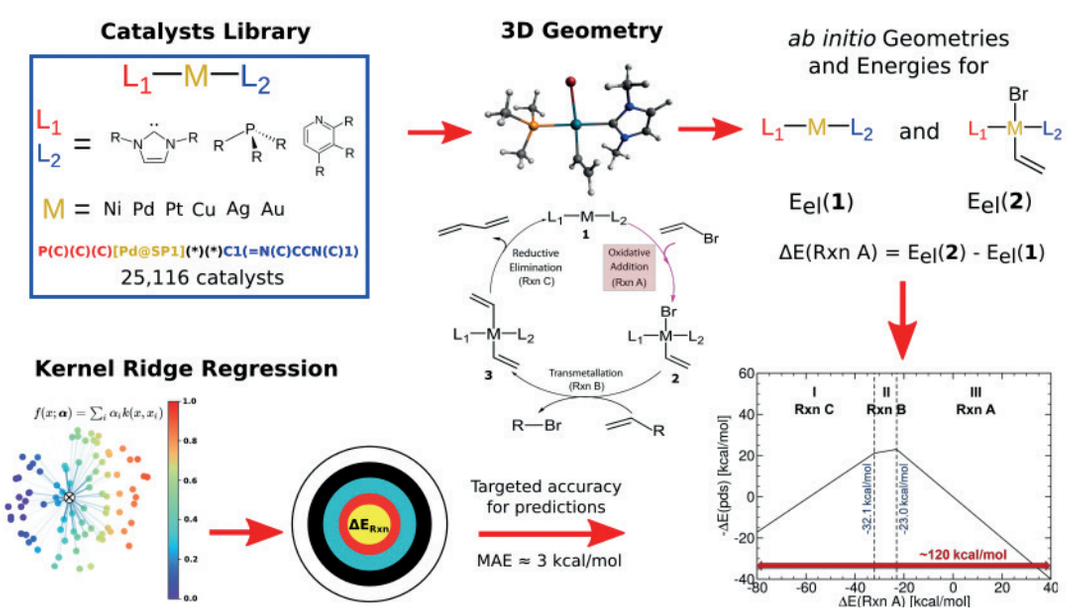


Fig. 1. Catalyst library and schematic representation of the supervised learning strategy for the prediction of 25,116 descriptor variables.

the various catalytic cycle intermediate species on one another.[27] Owing to the power of these scaling relationships, by knowing the value of a single descriptor variable it is possible to know the free energies associated with moving between any two intermediates of the catalytic cycle. In 2015, we transferred the concept, for the first time, into the realm of homogeneous catalysis[13a] and have since proposed further refinements covering many specific aspects of homogeneously catalyzed reactions (*e.g.,* regioselectivity, kinetics, classes of reactions, multidimensional reaction parameters).[13b–f] Using the Suzuki-Miyaura cross-coupling as a prototypical reaction, we originally reproduced experimental known trends for the behavior of different metal–ligand combinations.[13a] The use of these plots in tandem with machine learning models based on kernel ridge regression and different molecular representations (*i.e.,* molecular vectorial representation which encodes the compositional and structural information of a molecule suitable for a kernel function that measures their similarities and relationships) (Fig. 1) allowed us to dramatically increase the scale of the catalyst screening (~10^4)[9b] and to perform big-data type analysis, which uncovered hidden chemical patterns (Fig. 2).[9c] Using the machine learning strategy, we were able to identify promising Suzuki-Miyaura catalysts with a cost lower than $10 (USD) per mmol as well as catalysts suitable for the entire class of existing C–C cross-coupling reactions (*i.e.,* Kumada, Negishi, Stille, Hiyama, *etc*.).[9b,13d] To facilitate the exploration of the catalyst landscapes, we created interactive maps that are freely available to interested readers on the material cloud.[9c]

We are now prioritizing the use of similar workflows for challenging reactions of societal relevance (*e.g.,* hydrogenation of CO_2, aryl ether cleavage) as well as the elaboration of broadly applicable ML frameworks compatible with the more complex situations inherent to organocatalysis (less rigid structures) and to the diversity of descriptors useful to cast the catalytic cycles. Meeting this objective will require the elaboration of novel, more transferable local (*i.e.,* atom-based) representations that allow the reproduction of potentially complex descriptors, models that maximize the diversity of the training set for reducing the number of required instances and the overall cost of the predictive models. As advocated by Ceriotti,[28] these future transferable ML protocols will most likely be rooted in physically-motivated frameworks.

3. Reaching More Complexity

Another source of inspiration for the quantum chemical machine learning community comes from the complexity of the electronic structure problem. Quantum chemists have traditionally followed two conceptually diverging pathways to solve the Schrödinger equation and access molecular properties. The first, and by far the most common, is *deterministic* and consists in the development and use of a hierarchy of physically motivated approximations to the exact solution of the Schrödinger equation, either based on the many-body wavefunction (*i.e.,* Hartree-Fock, post-Hartree-Fock methods) or the electron density (*i.e.,* DFT). The second pathway is *stochastic* and is well represented by the collection of the quantum Monte Carlo techniques (*e.g.,* VMC, DMC, FCI-QMC, *etc.*). In this context, machine learning techniques are complementary to the two traditional approaches and represent a third, *statistical* route to access complex molecular properties. More specifically, this third paradigm of quantum chemistry relies upon the consideration that, given a sufficiently high number of observations, any molecular property of arbitrary complexity can be predicted with only a small degree of uncertainty.

With the current boom of machine-learning applications in quantum chemistry, the complexity of molecular properties able to be predicted has evolved from simple scalar (*e.g.* atomization and isomerization energies[2]) to vector and tensorial quantities (*e.g.* forces,[4] multipole moments,[29] (hyper-)polarizabilities[8]) up to complex functions and fields such as potential energy surfaces,[3] electron densities[6] and many-body wavefunctions.[30]

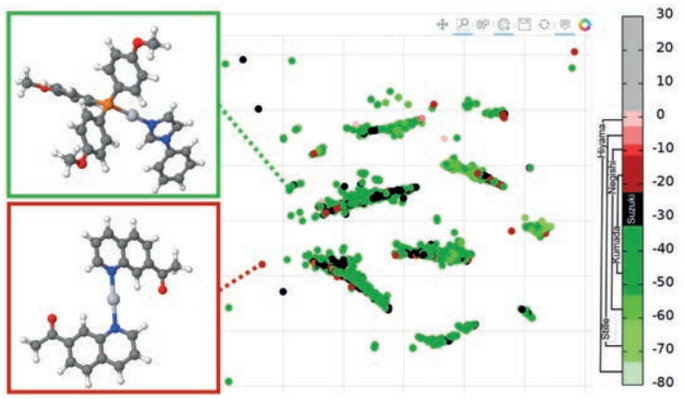


Fig. 2. Interactive map (DOI: 10.24435/materialscloud:2019.0007/v3) obtained after an unsupervised learning procedure (*i.e.* dimensionality reduction algorithm) and used for the identification of C–C cross coupling catalysts with the best thermodynamic profiles.

While both the increase in computing power and the improvement of quantum chemical algorithms contributed in accelerating the acquisition of quantum chemical data necessary to train the ML models, it is the sophistication of local machine learning representations and the development of kernels encoding symmetry conservation laws[28] that enabled the statistical learning of increasingly complex quantum chemical objects. Accessing the electron density [$\rho(\mathbf{r})$] or the many-body wavefunction at a negligible cost is especially appealing given that all the other molecular properties could be derived from them.

$\rho(\mathbf{r})$ is traditionally obtained from *ab initio* computations, but these may become rapidly demanding when targeting thousands of molecules/conformations or very large chemical systems. In addition to established methodologies, such as linear-scaling methods[31] and multipolar fittings of experimental data,[32] machine-learning models are developed to beat the computational cost of *ab initio* approaches by delivering the electron density solely from the nuclear coordinates. Yet, predicting the amplitude of a function at every point in space in a transferable manner is a non-trivial task.

For this purpose, together with Ceriotti and collaborators, we elaborated scalable and transferable machine-learning models of the electron density, based on an atom-centered, local representation of the density field in terms of Gaussian basis function.[6b,d] This framework relies on symmetry-adapted Gaussian process regression (SA-GPR), where the complex features and symmetries of the density field are represented by a hierarchy of kernels based on spherical tensors of increasing order (λ-SOAP).[8b] The predictive power of this technique has been tested both on a conformationally diverse ensemble of increasingly complex hydrocarbons extracted from molecular dynamics simulation and on a chemically diverse ensemble of amino acid side-chain dimers taken from the Sherrill's BioFragment Database (BFDb).[33] From a chemical perspective, the appeal of ML-predicted fundamental quantum chemical quantities are best illustrated by all the accessible qualitative insights provided by, for instance, topological descriptors (*e.g.,* density overlap region indicator DORI[34]), the electrostatic potential (ESP) (see Fig. 3), as well as binding energies readily estimated from the ESP, *etc*. The framework based on atom-centered contributions allows the linear-scaling transfer of the predicted densities from small molecules to more complex chemical systems, as long as the training set contains sufficient chemical diversity. The impressive transferability is best exemplified by the excellent charge density ML-prediction of pentapeptides using a SA-GPR model, trained exclusively on amino acid side-chain dimers. (Fig. 3).

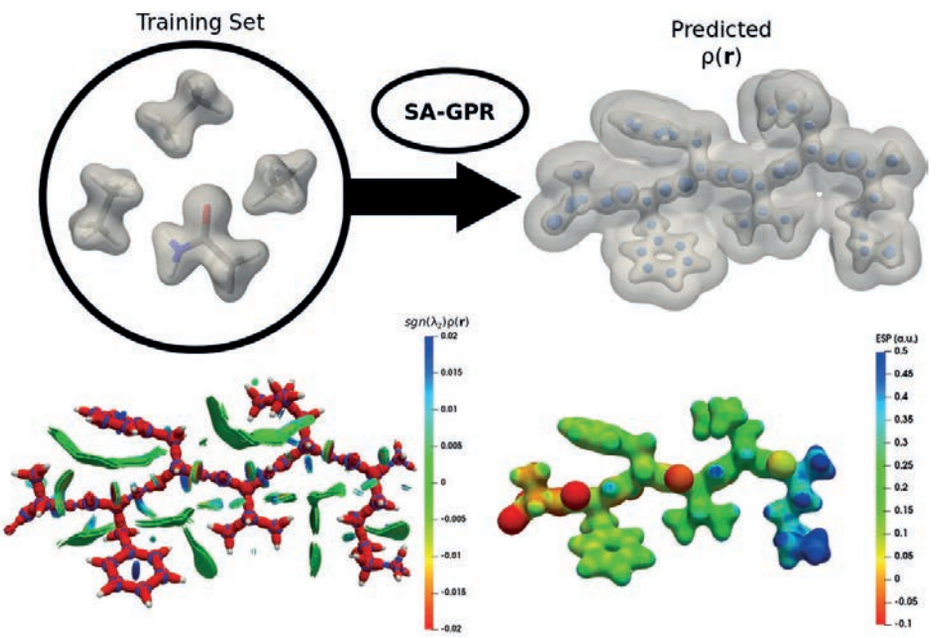


Fig. 3. (*top*) Transferability: electron density of an amyloid forming peptide (PBD ID: 3OW9) predicted by training on small molecule dimers. Electron density is reported at three isovalues: 0.5, 0.1 and 0.001 e^- $Bohr^{-3}$. (*bottom left*) DORI map of the peptide (isovalue: 0.9) colored by $sgn(\lambda_2)\rho(\mathbf{r})$ in the range from –0.02 a.u. (red: attractive interaction) to 0.02 a.u. (blue: repulsive interaction). (*bottom right*) Electrostatic potential (ESP) in Hartree (a.u.) of the peptide computed from the predicted density (isovalue: 0.05 e^- $Bohr^{-3}$). Further details in ref. [6d].

This is a clear quantum chemical example demonstrating how the increase in sophistication of the ML model transforms our way of accessing some of the most fundamental molecular properties. A step further in the evolution towards this complexity would consist of using similar transferable techniques to fully bypass the resolution of the Schrödinger equation and derive robust *statistical* relationship between the electron density and the exchange-correlation energies and, thus, to infer information for the design of improved density functional approximations.[6a]

4. Redefining our Quantum Chemical Objective

Finally, in addition to accelerating the large-scale screening of materials and offering access to the complex electronic structure properties irrespective of the system size, machine learning techniques are increasingly exploited to achieve converged statistical sampling without sacrificing quantum chemical accuracy.[4,5,35] While combining *ab initio* potentials with enhanced sampling techniques has always been a target, the impracticality of obtaining both converged statistical sampling and accurate energetics (*i.e.,* post-Hartree-Fock) has traditionally hampered the ability of improving the quantum chemical description of moderately sized, yet highly flexible molecules that evolve on complex potential energy surfaces. In fact, the most commonly employed approaches in computational organic chemistry[36] routinely ignore [or describe in a simplified (and borderline erroneous) manner] crucial phenomena such as conformational complexity of the species and the subtle interplay between non-covalent interactions, full entropic contributions, and solvent effects. In computational organic chemistry, a tremendous opportunity exists to change the *nature* of the computational problems tackled so far by marrying machine learning-based accelerated sampling approaches while retaining the advantages of molecular quantum chemistry flagship codes. These approaches are especially suited to bring higher levels of sophistication to the energetic description of rather flexible medium sized organo- and photoswitchable catalysts rationally designed based on concepts such as steric hindrance, π-stacking, anion-π interactions, hydrogen bonding *etc.*

Our efforts in this direction[37] consist of combining kernel-based approaches with global or local molecular representations (*i.e.* molecular or atom-based vectorial representation) together with enhanced sampling that use both Hamiltonian and *reservoir* replica exchange (Hres-RE). The so-called Modular Replica Exchange Simulator (MRES, see Fig. 4) dramatically speeds up free energy computations at the target QM level by reusing canonical sampling generated under a different, cheaper, Hamiltonian (*e.g.,* density functional tight binding DFTB).[38]

Fig. 5b illustrates the gain in time and accuracy for computing the *ab initio* free energy landscapes at 300 K of a prototypical system (thieno[2,3-b]thiophene, 40 atoms) possessing regions dominated by different extents of π-stacking interactions and variable fluxionality. The machine learning potentials, which use kernel ridge regression and the spectrum of London Axilrod-Teller-Muto (SLATM) representation,[10b] are trained to predict domain-based local pair natural orbitals DLPNO-CCSD(T)[39]/CBS or DFT energy quality. The 1500 most structurally different dispersion-corrected DFTB conformations out of a total of 43000 were selected[40] and employed for training the ML model reaching an accuracy of 0.6 kcal mol^{-1}. The relatively small number of necessary training points makes it possible to reach the DLPNO-CCSD(T)/CBS target.

The need for accurate quantum chemistry is demonstrated in Fig. 5 showing the rather poor performance of the DFTB semi-empirical potential, which flattens the entire profile reducing the relative energies of the three regions to basically zero. In contrast, DLPNO-CCSD(T)/CBS captures the significant, albeit subtle, differences between a deeper energy region well (Fig. 5c: *disarticulated*) and the entropy-driven open conformational region (Fig. 5c: *open*) dominated by soft vibrational modes and anharmonic effects.

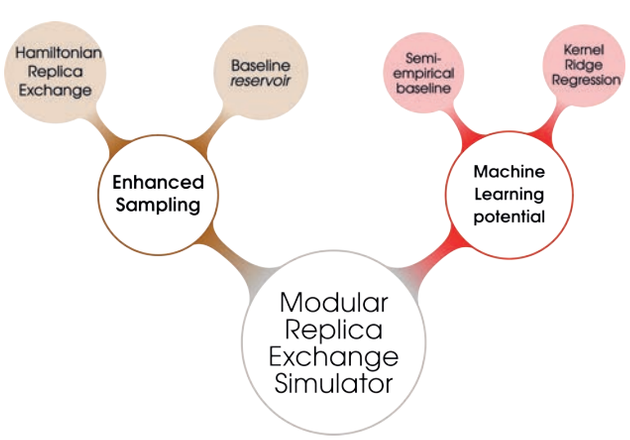


Fig. 4. Illustration of the sampling scheme designed to compute accurate free energies with machine learning potentials using MRES.

The same strategy can be used to predict the free-energy surfaces of chemical diverse systems such as amino acid (AA) dipeptides with the general structure ACE-AA-NME. The performance of the ML potential is illustrated in Fig. 6 with the thorough exploration of the free energy landscape of a flexible lysine dipeptide. At 300 K, three different minima (in trans conformation) characterizing the secondary structure of the dipeptide are identified in the Ramachandran plot. The conformational region A is nearly planar and stabilized by a hydrogen bond between one peptide bond and the carbonyl oxygen of another (see A in Fig. 6) forming a five-membered ring. The minima B and C are stabilized by hydrogen bonds that are characterized by two seven-membered rings of atoms where the side chain is either in the equatorial or in the axial position to this ring (see B and C in Fig. 6, respectively). Yet again, the comparisons between DFTB and a higher level (*i.e.,* PBE-dDsC[41]) stress the necessity of improving upon the DFTB baseline to achieve accurate free energy landscapes. DFTB tends to overstabilize the C7 axial form (*i.e.,* 0.8 kcal/mol for DFTB and 1.9 for DFTB+ΔML) (see region C) and understabilize the quasi planar C5 form (*i.e.,* 0.5 kcal/mol for DFTB and 0.1 for DFTB+ΔML) (see region A) of the lysine dipeptide.

The low-cost ML-based *ab initio* free-energy landscapes obtained for these dipeptides constitute a first step prior to building even more transferable ML-potentials capable of predicting the free-energy profile of any oligopeptides containing up to 10 or 20 peptide bonds and, alternatively, to accelerate the crystal structure prediction of small oligopeptides.

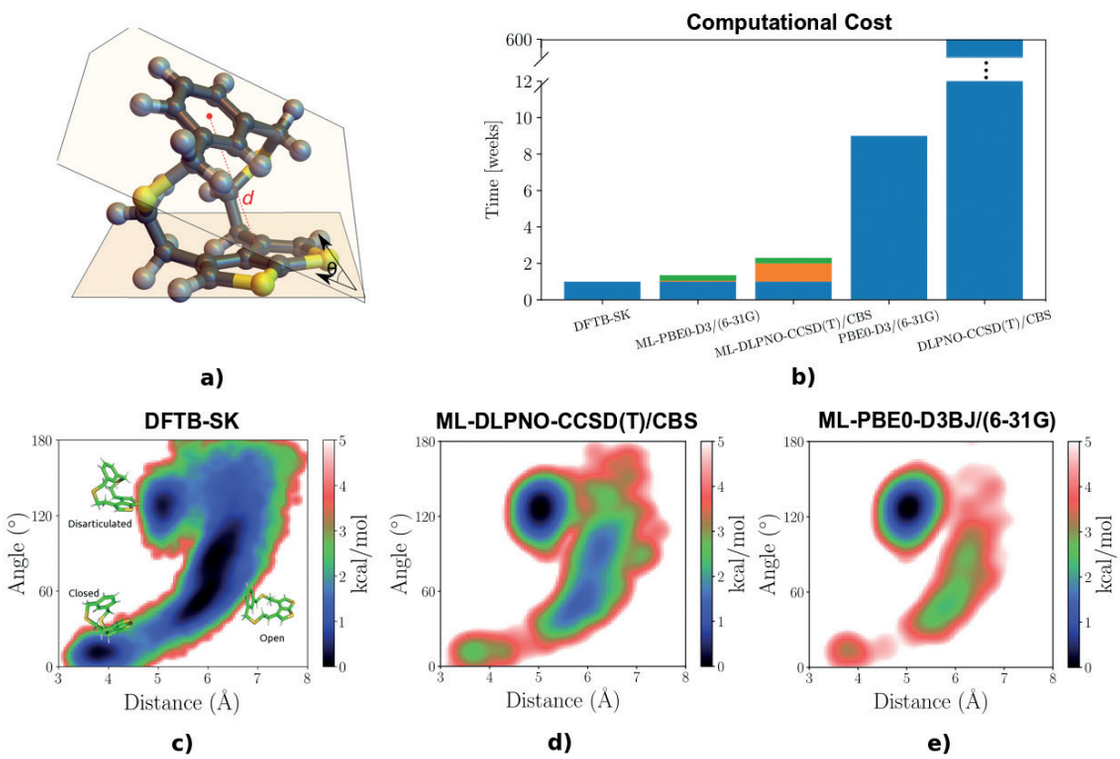
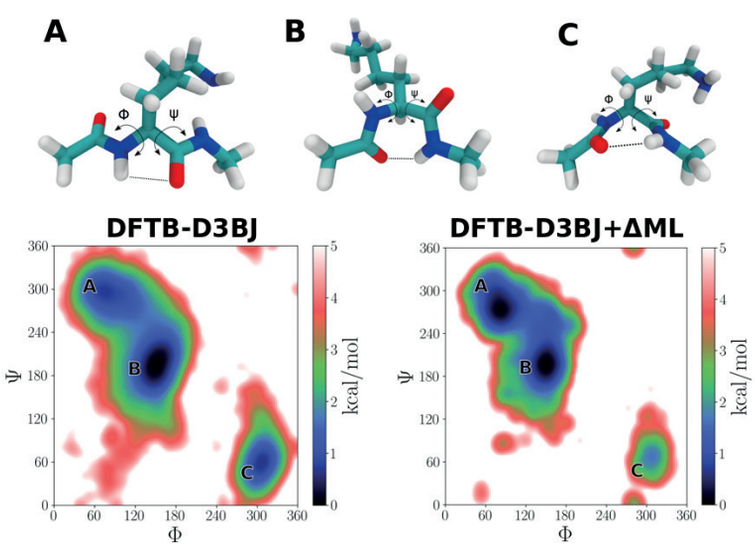


Fig. 5. a) Representation of dithiacyclophane and its collective variables: The distance between the cyclic moities and the angle between the planes that go through them. b) Total computational cost of generating the free energy landscapes with different potentials. The blue color represents time spent doing Temperature RE (T-RE) simulations, the orange color represents time spent doing the single point computations to train the ML models, and the green color represents time spent doing Hres-RE simulations using the results from a T-RE simulation with a cheaper potential. The cost of the free energy landscape generated with DLPNO-CCSD(T)/CBS is an estimation (DFTB-SK). c), d) and e) Free energy landscapes generated with the DFTB-SK, ML-DLPNO-CCSD(T)/CBS, ML-PBE0-D3BJ(6-31G).

Fig. 6. Representation of the three metastable minima of the lysine dipeptide (top) in the form ACE-LYS-NME capped with an acetyl group (ACE) and an N-methyl amide group (NME). The two dihedral angles Ψ and θ are used as collectives variables. Comparison of the Free Energy Surface (FES) of the lysine dipeptide obtained from direct DFTB-D3BJ computations (bottom left-hand corner) and ML-corrected potential (bottom right-hand corner) targeting the PBE-dDsC level.



5. Conclusions and Outlook

In this contribution, we describe our output associated with the use of kernel-based machine learning techniques covering three aspects of quantum chemistry with applicability in the fields of molecular catalysis, fundamental electronic structure theory, and statistical sampling in computational organic chemistry. The methods and applications discussed here belong to next-generation approaches available to the computational chemistry community and constitute part of more comprehensive array of novel ML methods and applications that are currently seeing and will continue to see widespread use.

Acknowledgements

The work highlighted in this article has been supported by EPFL, by the National Centre of Competence in Research (NCCR) 'Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL)' of the Swiss National Science Foundation (SNSF) and the European Research Council (ERC-CoG Grant 817977, PushQChem). The authors thank Michele Ceriotti and O. Anatole von Lilienfeld and their team for the collaboration.

Received: September 18, 2019

[1] a) C. Copéret, M. Chabanas, R. Petroff Saint-Arroman, J.-M. Basset, *Angew. Chemie Int. Ed.* **2003**, *42*, 156, DOI: 10.1002/anie.200390072; b) J. D. A. Pelletier, J.-M. Basset, *Acc. Chem. Res.* **2016**, *49*, 664, DOI: 10.1021/acs.accounts.5b00518.

[2] a) O. A. von Lilienfeld, *Angew. Chem. Int. Ed.* **2018**, *57*, 4164, DOI: 10.1002/anie.201709686; b) M. Rupp, O. A. von Lilienfeld, K. Burke, *J. Chem. Phys.* **2018**, *148*, 241401, DOI: 10.1063/1.5043213; c) M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. Von Lilienfeld, *Phys. Rev. Lett.* **2012**, *108*, 058301, DOI: 10.1103/PhysRevLett.108.058301

[3] a) J. Behler, *Angew. Chem. Int. Ed.* **2017**, *56*, 12828, DOI: 10.1002/anie.201703114; b) J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401, DOI: 10.1103/PhysRevLett.98.146401; c) J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, A. E. Roitberg, *J. Chem. Phys.* **2018**, *148*, 241733, DOI: 10.1063/1.5023802.

[4] a) S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller, *Sci. Adv.* **2017**, *3*, e1603015, DOI: 10.1126/sciadv.1603015; b) S. Chmiela, H. E. Sauceda, K.-R. Müller, A. Tkatchenko, *Nat. Comm.* **2018**, *9*, 3887, DOI: 10.1038/s41467-018-06169-2. c) Z. Li, J. R. Kermode, A. De Vita, *Phys. Rev. Lett.* **2015**, *114*, 096405, DOI: 10.1103/PhysRevLett.114.096405.

[5] M. Gastegger, J. Behler, P. Marquetand, *Chem. Sci.* **2017**, *8*, 6924, DOI: 10.1039/C7SC02267K.

[6] a) F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, K.-R. Müller, *Nat. Comm.* **2017**, *8*, 872, DOI: 10.1038/s41467-017-00839-3; b) A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, M. Ceriotti, *ACS Cent. Sci.* **2019**, *5*, 57, DOI: 10.1021/acscentsci.8b00551; c) A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, R. Ramprasad, *npj Comput. Mater.* **2019**, *5*, 22, DOI: 10.1038/s41524-019-0162-7. d) A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti, C. Corminboeuf, *Chem. Sci.* **2019**, DOI: 10.1039/C9SC02696G.

[7] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke, *Phys. Rev. Lett.* **2012**, *108*, 253002, DOI: 10.1103/PhysRevLett.108.253002.

[8] a) D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, M. Ceriotti, *Proc. Natl. Acad. Sci.* **2019**, *116*, 3401, DOI: 10.1073/pnas.1816132116; b) A. Grisafi, D. M. Wilkins, G. Csányi, M. Ceriotti, *Phys. Rev. Lett.* **2018**, *120*, 036002, DOI: 10.1103/PhysRevLett.120.036002.

[9] a) J. R. Kitchin, *Nat. Catal.* **2018**, *1*, 230, DOI: 10.1038/s41929-018-0056-y; b) B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld, C. Corminboeuf, *Chem. Sci.* **2018**, *9*, 7069, DOI: 10.1039/C8SC01949E; c) B. Sawatlon, M. D. Wodrich, B. Meyer, A. Fabrizio, C. Corminboeuf, *ChemCatChem* **2019**, *11*, 4096, DOI: 10.1002/cctc.201900597.

[10] a) A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, e1701816, DOI: 10.1126/sciadv.1701816; b) B. Huang, O. A. von Lilienfeld, *arXiv:1707.04146*, 2017.

[11] a) B. Maryasin, P. Marquetand, N. Maulide, *Angew. Chem. Int. Ed.* **2018**, *57*, 6978, DOI: 10.1002/anie.201803562; b) J. N. Wei, D. Duvenaud, A. Aspuru-Guzik, *ACS Cent. Sci.* **2016**, *2*, 725, DOI: 10.1021/acscentsci.6b00219; c) P. Swhaller, T. Gaudin, D. Lanyi, C. Bekas, T. Laino, *Chem. Sci.* **2018**, *9*, 6091, DOI: 10.1039/C8SC02339E.

[12] a) [1] T. Bligaard, J. K. Nørskov, S. Dahl, J. Matthiesen, C. H. Christensen, J. Sehested, *J. Catal.* **2004**, *224*, 206, DOI: 10.1016/j.jcat.2004.02.034; b) J. K. Norskov, F. Abild-Pedersen, F. Studt, T. Bligaard, *Proc. Natl. Acad. Sci.* **2011**, *108*, 937, DOI: 10.1073/pnas.1006652108.

[13] a) M. Busch, M. D. Wodrich, C. Corminboeuf, *Chem. Sci.* **2015**, *6*, 6754, DOI: 10.1039/C5SC02910D; b) M. D. Wodrich, M. Busch, C. Corminboeuf, *Chem. Sci.* **2016**, *7*, 5723, DOI: 10.1039/C6SC01660J; c) M. Busch, M. D. Wodrich, C. Corminboeuf, *ACS Catal.* **2017**, *7*, 5643, DOI: 10.1021/acscatal.7b01415; d) M. Busch, M. D. Wodrich, C. Corminboeuf, *ChemCatChem* **2018**, *10*, 1592, DOI: 10.1002/cctc.201701710; e) M. D. Wodrich, B. Sawatlon, M. Busch, C. Corminboeuf, *ChemCatChem* **2018**, *10*, 1586, DOI: 10.1002/cctc.201701709; f) M. D. Wodrich, M. Busch, C. Corminboeuf, *Helv. Chim. Acta* **2018**, *101*, e1800107, DOI: 10.1002/hlca.201800107.

[14] A. Aspuru-Guzik, M.-H. Baik, S. Balasubramanian, R. Banerjee, S. Bart, N. Borduas-Dedekind, S. Chang, P. Chen, C. Corminboeuf, F.-X. Coudert, L. Cronin, C. Crudden, T. Cuk, A. G. Doyle, C. Fan, X. Feng, D. Freedman, S. Furukawa, S. Ghosh, F. Glorius, M. Jeffries-EL, N. Katsonis, A. Li, S. S. Linse, S. Marchesan, N. Maulide, A. Milo, A. R. H. Narayan, P. Naumov, C. Nevado, T. Nyokong, R. Palacin, M. Reid, C. Robinson, G. Robinson, R. Sarpong, C. Schindler, G. S. Schlau-Cohen, T. W. Schmidt, R. Sessoli, Y. Shao-Horn, H. Sleiman, J. Sutherland, A. Taylor, A. Tezcan, M. Tortosa, A. Walsh, A. J. B. Watson, B. M. Weckhuysen, E. Weiss, D. Wilson, V. W. W. Yam, X. Yang, J. Y. Ying, T. Yoon, S.-L. You, A. J. G. Zarbin, H. Zhang, *Nat. Chem.* **2019**, *11*, 286, DOI: 10.1038/s41557-019-0236-7.

[15] a) A. G. Maldonado, G. Rothenberg, *Chem. Soc. Rev.* **2010**, *39*, 1891, DOI: 10.1039/b921393g; b) G. A. Landrum, J. E. Penzotti, S. Putta, *Meas. Sci. Technol.* **2005**, *16*, 270, DOI: 10.1088/0957-0233/16/1/035.

[16] a) M. K. Nielsen, D. T. Ahneman, O. Riera, A. G. Doyle, *J. Am. Chem. Soc.* **2018**, *140*, 5004, DOI: 10.1021/jacs.8b01523; b) D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, *Science* **2018**, *360*, 186, DOI: 10.1126/science.aar5169.

[17] A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow, S. E. Denmark, *Science* **2019**, *363*, eaau5631, DOI: 10.1126/science.aau5631.

[18] a) J. P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin, H. J. Kulik, *Inorg. Chem.* **2019**, *58*, 10592, DOI: 10.1021/acs.inorgchem.9b00109; b) J. P. Janet, H. J. Kulik, *Chem. Sci.* **2017**, *8*, 5137, DOI: 10.1039/C7SC01247K; c) J. P. Janet, L. Chan, H. J. Kulik, *J. Phys. Chem. Lett.* **2018**, *9*, 1064, DOI: 10.1021/acs.jpclett.8b00170.

[19] A. Nandy, J. Zhu, J. P. Janet, C. Duan, R. B. Getman, H. J. Kulik, *ACS Catal.* **2019**, *9*, 8243, DOI: 10.1021/acscatal.9b02165.

[20] C. Duan, J. P. Janet, F. Liu, A. Nandy, H. J. Kulik, *J. Chem. Theory Comput.* **2019**, *15*, 2331, DOI: 10.1021/acs.jctc.9b00057.

[21] a) P. Sabatier, *Ber. Deutsch. Chem. Gesellshaft* **1911**, *44*, 1984: b) P. Sabatier, 'La Catalyse en Chimie Organique Librarie Polytechnique', Paris, **1913**.

[22] See a special issue entitled 'Computational Catalysis for Organic Synthesis' in *Acc. Chem. Res.* **2016**, *49*.

[23] a) H. Dau, C. Limberg, T. Reier, M. Risch, S. Roggan, P. Strasser, *ChemCatChem* **2010**, *2*, 724, DOI: 10.1002/cctc.201000126; b) J. Greeley, N. M. Markovic, *Energy Environ. Sci.* **2012**, *5*, 9246, DOI: 10.1039/c2ee21754f; c) M. T. M. Koper, *Nat. Chem.* **2013**, *5*, 255, DOI: 10.1038/nchem.1600.

[24] a) S. Trasatti, *J. Electroanal. Chem. Interfacial Electrochem.* **1972**, *39*, 163, DOI: 10.1016/S0022-0728(72)80485-6; b) J. Greeley, I. E. L. Stephens, A. S. Bondarenko, T. P. Johansson, H. A. Hansen, T. F. Jaramillo, J. Rossmeisl, I. Chorkendorff, J. K. Nørskov, *Nat. Chem.* **2009**, *1*, 552, DOI: 10.1038/nchem.367; c) M. Chhowalla, H. S. Shin, G. Eda, L.-J. Li, K. P. Loh, H. Zhang, *Nat. Chem.* **2013**, *5*, 263, DOI: 10.1038/nchem.1589; d) A. R. Zeradjanin, J.-P. Grote, G. Polymeros, K. J. J. Mayrhofer, *Electroanalysis* **2016**, *28*, 2256, DOI: 10.1002/elan.201600270.

[25] a) C. J. H. Jacobsen, S. Dahl, B. S. Clausen, S. Bahn, A. Logadottir, J. K. Nørskov, *J. Am. Chem. Soc.* **2001**, *123*, 8404, DOI: 10.1021/ja010963d; b) C. J. H. Jacobsen, S. Dahl, A. Boisen, B. S. Clausen, H. Topsøe, A. Logadottir, J. K. Nørskov, *J. Catal.* **2002**, *205*, 382, DOI: 10.1006/jcat.2001.3442.

[26] I. C. Man, H. Su, F. Calle-Vallejo, H. A. Hansen, J. I. Martínez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov, J. Rossmeisl, *ChemCatChem* **2011**, *3*, 1159, DOI: 10.1002/cctc.201000397.

[27] a) F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. R. Munter, P. G. Moses, E. Skúlason, T. Bligaard, J. K. Nørskov, *Phys. Rev. Lett.* **2007**, *99*, 016105, DOI: 10.1103/PhysRevLett.99.016105; b) F. Calle-Vallejo, J. I. Martínez, J. M. García-Lastra, J. Rossmeisl, M. T. M. Koper, *Phys. Rev. Lett.* **2012**, *108*, 116103, DOI: 10.1103/PhysRevLett.108.116103.

[28] M. Ceriotti, *J. Chem. Phys.* **2019**, *150*, 150901, DOI: 10.1063/1.5091842.

[29] T. Bereau, D. Andrienko, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2015**, *11*, 3225, DOI: 10.1021/acs.jctc.5b00301.

[30] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, R.J. Maurer, *arXiv:1906.10033,* **2019**.

[31] a) P. D. Walker, P. G. Mezey, *J. Am. Ceram. Soc.* **1993**, *115*, 12423, DOI: 10.1021/ja00079a025; b) T. E. Exner, P. G. Mezey, *J. Phys. Chem. A* **2002**, *106*, 11791, DOI: 10.1021/jp0263166; c) H. Stoll, G. Wagenblast, H. Preuss, *Theor. Chim. Acta* **1980**, *57*, 169, DOI: 10.1007/BF00574903.

[32] a) F. L. Hirshfeld, *Acta Crystallogr. B* **1971**, *27*, 769, DOI: 10.1107/S0567740871002905; b) R. F. Stewart, *Acta Crystallogr. A* **1976**, *32*, 565,

DOI: 10.1107/S056773947600123X; c) N. K. Hansen, P. Coppens, *Acta Crystallogr. A* **1978**, *34*, 909, DOI: 10.1107/S0567739478001886.
[33] L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz, C. D. Sherrill, *J. Chem. Phys.* **2017**, *147*, 161727, DOI: 10.1063/1.5001028.
[34] P. de Silva, C. Corminboeuf, *J. Chem. Theory Comput.* **2014**, *10*, 3745, DOI: 10.1021/ct500490b
[35] a) H. E. Sauceda, S. Chmiela, I. Poltavsky, K. R. Müller, A. Tkatchenko, *J. Chem. Phys.* **2019**, *150,* 114102*,* DOI: 10.1063/1.5078687; b) K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller A. Tkatchenko, *Nat. Commun.* **2017**, *8*, 13890, DOI: 10.6084/m9.figshare.978904.
[36] a) R. Petraglia, A. Nicolaï, M. D. Wodrich, M. Ceriotti, C. Corminboeuf, *J. Comput. Chem.* **2016**, *37*, 83, DOI: 10.1002/jcc.24025; b) S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847, DOI: 10.1021/acs.jctc.9b00143.
[37] R. Fabregat, A. Fabrizio, B. Meyer, C. Corminboeuf, *in preparation*.
[38] a) D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, R. Kaschner, *Phys. Rev. B* **1995**, *51*, 12947, DOI: 10.1103/PhysRevB.51.12947; b) M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260, DOI: 10.1103/PhysRevB.58.7260.
[39] a) C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101, DOI: 10.1063/1.4821834.1; b) C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106, DOI: 10.1063/1.4773581.
[40] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, M. Ceriotti, *J. Chem. Phys.* **2018**, *148*, 241730, DOI: 10.1063/1.5024611.
[41] a) S. Steinmann, C. Corminboeuf, *J. Chem. Theory Comput.* **2010**, *6*, 1990, DOI: 10.1021/ct1001494; b) S. Steinmann, C. Corminboeuf, *J. Chem. Phys.* **2011**, *134*, 044117, DOI: 10.1063/1.3545985; c) S. Steinmann, C. Corminboeuf, *J. Chem. Theory Comput.* **2011**, *7*, 3567, DOI: 10.1021/ct200602x.