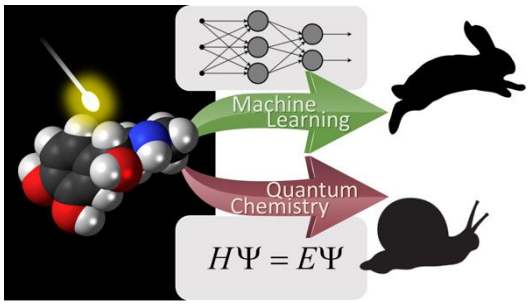
Efficient simulations with machine learning and semiempirical methods

Pavlo O. Dral

[1]State Key Laboratory of Physical Chemistry of Solid Surfaces, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Department of Chemistry, and College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China

dral@xmu.edu.cn



Machine learning (ML) has emerged as a useful tool for accelerating quantum chemical (QC) calculations.[1-2] We exploit its power to speed up calculations of rovibrational,[3] UV/vis absorption spectra,[4] and dynamics[5] with our *MLatom* program package (see MLatom.com).[6]

Figure 1 Machine learning is much faster than quantum chemical methods.[2]

Another active area of our research is developing and using efficient low-cost semiempirical QC methods (SQC) such as ODM2, which are to-date the most accurate NDDO-based SQC methods for a broad variety of properties including ground-state and excited-state properties.[7]

Exploiting the synergy between QC and ML methods, we have developed hybrid ML/QC approaches using ML to improve the accuracy of QC methods via Δ-learning[8] or tuning the semiempirical QC Hamiltonian[9]. This research has led to the development of the general-purpose AIQM1 method approaching coupled-cluster accuracy for ground-state properties of common organic compounds,[10] but also transferable to other properties.

[1] P. O. Dral, *J. Phys. Chem. Lett.* **2020,** *11*, 2336–2347.

[2] P. O. Dral, M. Barbatti, *Nat. Rev. Chem.* **2021,** *5*, 388–405.

[3] P. O. Dral, A. Owens, A. Dral, G. Csányi, *J. Chem. Phys.* **2020,** *152*, 204110.

[4] B.-X. Xue, M. Barbatti, P. O. Dral, *J. Phys. Chem. A* **2020,** *124*, 7199–7210.

[5] P. O. Dral, M. Barbatti, W. Thiel, *J. Phys. Chem. Lett.* **2018,** *9*, 5660–5663.

[6] P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti, *Top. Curr. Chem.* **2021,** *379*, 27.

[7] P. O. Dral, X. Wu, W. Thiel, *J. Chem. Theory Comput.* **2019,** *15*, 1743–1760.

[8] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2015,** *11*, 2087–2096.

[9] P. O. Dral, O. A. von Lilienfeld, W. Thiel, *J. Chem. Theory Comput.* **2015,** *11*, 2120–2125.

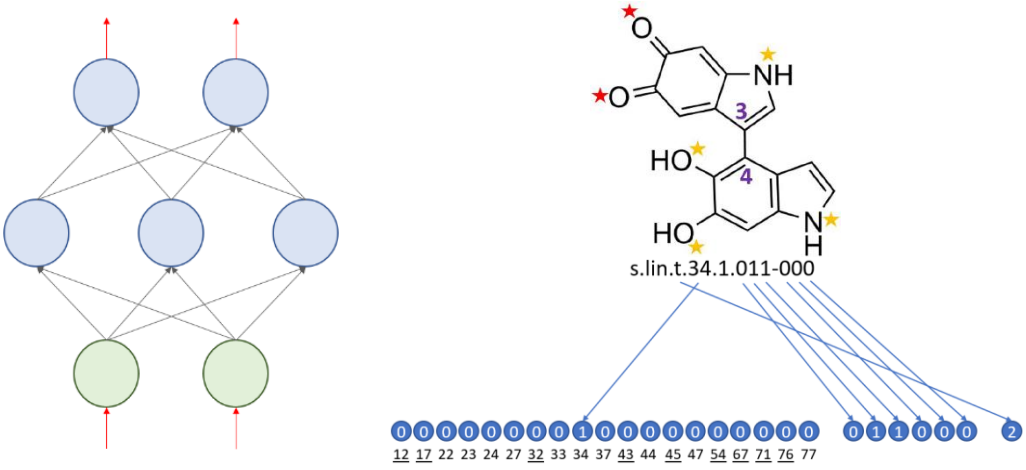
[10] P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *submitted. Preprint on ChemRxiv: https://doi.org/10.33774/chemrxiv-2021-zk477* **2021**.

Modelling DHI dimer stability and optical absorption with neural networks

Daniel Bosch

IQCC – Universitat de Girona

Abstract: Melanin is composed of oligomers made of two basic units, 5,6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-carboxylic acid (DHICA), and melanin oligomers contain 4 to 8 DHI/DHICA units.[1,2] The stability and optical absorption of a comprehensive virtual library of 830 DHI dimers has been studied previously,[3] optimizing the structures with DFT and calculating the vertical absorption with TD-DFT. Considering that structural possiblities grow enormously as more DHI units the oligomer has, we envision the possibility of applying Machine Learning (ML) to calculate the properties of larger systems, and test here the application to the dimers as a proof of concept:[4] we employ Artificial Neural Networks to model both the stability and the optical absorption of the DHI dimers. Since the dataset used is very homogeneous, we use this as an advantage and we codify the molecules according to their diversity using molecular descriptors of structure. We have tested several molecular descriptors and fine-tuned the Neural Networks parameters in order to enhance the performance of the designed models.



[1] H. Okuda, K. Yoshino, K. Wakamatsu, S. Ito, T. Sota, *Pigment Cell Melanoma Res.* **2014**, 27.

[2] A. Pezzella, A. Napolitano, M. dIschia, G. Prota, R. Seraglia, P. Traldi, *Rapid Commun. Mass Spectrom.* **1997**, 11, 368-372.

[3] Blancafort, L., Wang, J., *Angew.Chem. Int. Ed.* **2021**, 60,18800–18809.

[4] Dral, P.O., Barbatti, M. Molecular excited states through a machine learning lens. *Nat Rev Chem* **5**, 388–405 (2021).