Journal of Computer-Aided Molecular Design (2021) 35:397–398
https://doi.org/10.1007/s10822-021-00373-x

EDITORIAL

Editorial special issue on "Quantum Mechanics in Industry"

Andreas H. Göller[1]

Accepted: 12 January 2021 / Published online: 15 February 2021

When being invited to guest edit a special journal issue on *Quantum Mechanics in Industry* I spontaneously agreed to the task to promote the visibility and usage of an arsenal of methods that is often seen as costly, complicated and too slow for routine application but has the potential to explain many chemical phenomena where empirical methods and machine learning fail. The contributions to this issue confirmed and even surpassed my expectations and I want to express my gratitude to the authors for their papers and also the reviewers for their helpful feedback that even enhanced the manuscripts.

Computational chemistry started with the advent of quantum mechanics in the first quarter of the nineteenth century, based on the ground-breaking works of Born, Einstein, Heisenberg, Pauli and Schrödinger, just to name some of the major contributing scientists. For more than four decades QM had to stay a theoretical concept due to the inherent computational costs. In the seventies of the last century semiempirical and low-level Hartree–Fock methods hit the stage. Though computers were still big and slow, they finally arrived in pharmaceutical industry, and chemists interested in understanding and rational design started to become computational chemists.

The use of *Quantum Mechanics in Industry* was and is determined by the Bauhaus principle of "Form follows Function", i.e. in pharmaceutical and agrochemical industry computational chemists focus on solving the problem at hand. Application of QM is to a certain extent industrialized and workflow driven with the constraints of fit to project timelines, applicability to larger compound series, stability of computational processes, and typically applying the most cost-efficient lowest-level method to arrive at the required scientific information.

Nevertheless, this special issue on "Quantum Mechanics in Industry" due to the various scientific challenges addressed, provides a broad spectrum of quantum chemical methods. They range from semiempirical and density-functional to coupled-cluster methods and even to machine learning for the prediction of any thinkable quantum mechanical calculation output ranging from energies to dipole moments or spectra, circumventing the high costs of direct QM calculations.

The spectrum of problems addressed in this special issue ranges from structural and energetic characterization of the ligands like isomerism, conformer, protomer and tautomer states and energies, to chemical reactivity, ligand binding interactions with their targets and solvation effects.

Quantum mechanics is the only computational method able to describe chemical reactivity, which is by the way distinct to machine learning of retro-synthesis pathways. Therefore, energetics, stability and reaction pathway elucidation became the first application of QM by chemists, as exemplified by three articles in this special issue.

The first article is concerned with density functional theory calculations on the rate limiting steps of product formation for oxidation reactions by *Flavin-containing Monooxygenase* and glucuronidation reactions by *UDP-glucuronosyltransferase* (https://doi.org/10.1007/s10822-020-00321-1), both important enzymes in drug metabolism. The learning from this basic research study will then in the next step be incorporated into machine learning models for site-of-metabolism prediction algorithms. The second study deals with the application of conceptual density functional theory, namely the application and limits of the *electrophilicity index* (https://doi.org/10.1007/s10822-020-00342-w), for the prediction of the inherent reactivity of *covalent binders*. The authors, apart from this do's and don'ts article, have recently published a machine learning model based on quantum mechanically derived synthetic data for training. So called synthetic data, i.e. computational results as input as substitute for missing experimental data, is a recent trend in machine learning, as exemplified also in the EC-RISM publication of this special issue and laid out in the review on QM machine learning. The third study describes the *derivation of a binding mode hypothesis* for prothioconazole to its

✉ Andreas H. Göller
andreas.goeller@bayer.com

[1] Bayer Pharma AG, Wuppertal, Germany

target *CYP51* (https://doi.org/10.1007/s10822-020-00331-z) in the absence of a complex crystal structure. The topology of binding with significant differences for R- and S-enantiomers is characterized by analysis of electron densities (QTAIM) and the difference of binding enthalpy at coupled cluster [DLPNO-CCSD(T)] level of theory.

Central to the small molecules in pharmaceuticals and agrochemicals are also isomerism, conformation, tautomer and protomer state, all of them crucially depending on solvation. Whereas the generation of molecular conformations is long possible via stochastic or systematic empirical algorithms, the identification of low-lying conformations in different embeddings like solvent, membrane or binding site, requires a more rigorous treatment than the empirical methods for conformer, tautomer or protomer generation can provide. The article on ReSCoSS provides a *QM based workflow for the identification of relevant solution conformers* (https://doi.org/10.1007/s10822-020-00337-7) which identifies low-energy solvent-dependent and target binding conformations. Minimum energy conformations also play a major role in a study on the relative enantiomer free energies applying double-hybrid density functional theory for the *rationalization of the synthetic inaccessibility of certain enantiomers of hydroxyestradienone* (https://doi.org/10.1007/s10822-020-00353-7), one of the starting materials for the synthesis of a promising drug candidate, as required by regulatory authorities. A similar problem was addressed in a study to *identify potential Z-isomers of neonicotinoids* (https://doi.org/10.1007/s10822-020-00326-w) by DFT potential energy scans combined with coupled-cluster CCSD(T) energy corrections of the isomerization reaction paths. This, in combination with quantum chemical NMR calculations and low-temperature NMR experiments provided evidence for the existence of Z-clothianidin, and the data were accepted by regulatory authorities.

Further to molecular conformation, the identification of embedding-dependent correct tautomer and protomer states is crucial for the design of new drugs and agro substances. Starting point in any target-based molecular modeling project is the experimental crystal structure determination of a protein ligand complex. Refinement of the diffraction data is typically performed by empirical force fields that fit the sequence to electron density. The study on *X-ray refinement and accurate tautomer and protomer states* (https://doi.org/10.1007/s10822-020-00354-6) provides a route to incorporate explicit terms for electrostatics, polarization, dispersion, hydrogen bonds, and other key interactions that are missing in empirical approaches.

The topic of *ring-chain tautomers* (https://doi.org/10.1007/s10822-020-00334-w), not well-addressed in cheminformatics codes and challenging from an energetic standpoint due to effects like change in conjugation or aromaticity, is taken up by applying quantum-mechanical scoring after micro-p*K*a predictions of potential protonatable sites, thus avoiding pre-defined tautomerization patterns.

Any biological process is governed by interactions of the molecules with the surrounding, namely binding pocket, membranes, tissues or complex solvents. Proper description of solute–solvent interactions after more than three decades is still an ongoing research topic. One decade of joint academia–industry method *development and application of the embedded cluster reference interaction site model* (EC-RISM) is reviewed in this special issue. Emphasis is put on the results from participating in several SAMPL (Statistical Assessment of Modeling of Proteins and Ligands) blind prediction challenges (https://doi.org/10.1007/s10822-020-00347-5). The story provided here is not a complete success story, but an honest and clear description of the current state and the long way still to go.

Any continuum solvation model suffers from the issue not correctly describing strong and highly directed interactions of polar ligand functional groups with solvent molecules. The strategy here is so-called *microsolvation* (https://doi.org/10.1007/s10822-020-00366-2), i.e. the placement of explicit solvent molecules and the otherwise continuous description of solvation. The authors here describe concepts and the Ansatz they chose, but at the same time emphasize also that they are not there yet.

Knowing that conformer, tautomer and protomer states are closely connected and dependent on the molecular embedding, and that many of the weak non-bonding interactions are not well-described by empirical methods, quantum mechanics is the only alternative. Furthermore, the *time-dependent fluctuations of molecules at ambient temperatures* require sampling by molecular dynamics. The solution would be the combination of both worlds, in the form of *QM-MD*, which is an active field of research but still too costly for the description of realistic biological systems in the condensed phase.

The most promising candidate for a solution of the computational cost dilemma evolving now is *machine learning of potential energy surfaces* (https://doi.org/10.1007/s10822-020-00346-6) and almost every other property that can be derived from QM calculation results. The state of the art is covered in an extensive review as part of this special issue, which I foresee as an outlook for the *bright future of* "*combining the best of all worlds*", *namely cheminformatics*, *molecular mechanics*, *quantum mechanics and machine learning*.