



 Cite this: *Phys. Chem. Chem. Phys.*,
2023, 25, 26415

Benchmark experiments for numerical quantum chemistry

 Ricardo A. Mata, ^a Anne Zehnacker-Rentien ^{ab} and Martin A. Suhm ^a

DOI: 10.1039/d3cp90186f

rsc.li/pccp

Driven by impressive advances in both hard- and software as well as a steady flow of new concepts, quantum chemical predictions go from strength to strength. Nonetheless, the benefits from this progress will starkly depend on the quality of data we provide or are provided with. Quantum chemical methods can replace and generate novel research workflows, but these will only be useful with a clear knowledge of their accuracy and scope. Benchmarking of electronic structure methods has been a common practice ever since the 1990s, partly as a response to the plethora of density functional methods introduced every year. With a somewhat bloated toolbox available to the computational chemist, it became more relevant to know which is the right tool, rather than to add further alternatives into the mix. This has also translated into the nagging question: “Which functional should I use?” Many authors have answered the call and have compared time and time again different DFT methods for different properties, sometimes as a main focus, others as a side task to justify the choice of theory level. Whatever the case may be, the shift clearly moved from questioning whether

one can compute it to how one should compute it.

Most of the benchmarking practices to date tend to focus on comparing relative electronic energies, chiefly between the methods in test and a reference (expensive) quantum chemical approach (*e.g.*, the CCSD(T) gold standard). This is a straightforward computational protocol, but fails to cover the complexity involved in a large number of chemical observables. Experiment remains the ultimate adjudicator about the suitability of theoretical models and protocols. This usually requires combined solutions for the electronic structure and for the dynamics of the nuclei. To avoid misleading error compensation in theory and misleading experimental references, great care in the design of benchmark experiments is required. This includes sufficiently large databases, multi-experimental cross-validations, and the organization of blind challenges for unbiased predictions. On the theory side, it is crucial to have approaches at hand which minimize error in either electronic or nuclear dynamics problems, such that rigorous lessons can be learned for the other component of the numerical challenge. On the experimental side, gas-phase experiments, often at low temperature, obtained by spectroscopic techniques with high accuracy, are the natural comparison with theoretical data. In this respect, supersonic expansions or cryogenic ion traps allow studying cold isolated

molecules or weakly bound complexes with unprecedented precision. High-resolution rovibrational or rotational spectroscopy are available for small systems, while larger systems are often studied at vibrational resolution. Besides these experiments providing structures and nuclear motion information, crystallographic data (<https://doi.org/10.1039/D2CP04098K>), luminescence (<https://doi.org/10.1039/D2CP01623K>), mass spectrometry experiments such as photon- or collision-induced dissociation or ion mobility (<https://doi.org/10.1039/D2CP01414A>), NMR spectroscopy (<https://doi.org/10.1039/D2CP04092A> or <https://doi.org/10.1039/D2CP03992C>) or X-ray scattering (<https://doi.org/10.1039/D2CP02933B>) give valuable information on electronic effects. A wealth of experimental methods allow studying reactive systems in solution (<https://doi.org/10.1039/D2CP03937K>) such as electrochemistry or luminescence, and provide new areas of benchmarking.

This themed collection, *Benchmark Experiments for Numerical Quantum Chemistry*, of more than 40 articles (about one third being classified as hot articles) addresses different aspects of this endeavour, bringing quantum theory and experiment together at suitable meeting points, for the mutual benefit of both communities. Two perspectives in the field of non-covalent interactions address the theoretical advances in fully coupled, numerically exact rovibrational states (<https://doi.org/10.1039/D2CP04005K>) and

^a *Institute of Physical Chemistry, University of Goettingen, Tammannstr. 6, 37077 Goettingen, Germany*

^b *Institut des Sciences Moléculaires d'Orsay (ISMO), CNRS, rue André Rivière, Université Paris-Saclay, F-91405 Orsay, France*

how to organize a particular blind challenge on hydrate vibrational shifts from the experimental side (<https://doi.org/10.1039/D2CP01119K>). The latter also invites less exact quantum approaches and machine learning (for the outcome, see <https://doi.org/10.1039/D3CP01216F>), to be put under scrutiny in later rounds. For somewhat more rigid molecular systems, a review (<https://doi.org/10.1039/D2CP04706C>) demonstrates how closely the Born–Oppenheimer concept of molecular equilibrium structure and the experimentally observable rotational constants can be brought together. Formic acid is reviewed as an elementary example for the vibrational characterization of a bistable molecule (<https://doi.org/10.1039/D2CP04417J>). Finally, a tutorial review addresses how reactivity scales help in structuring and overcoming challenges in kinetics benchmarking (<https://doi.org/10.1039/D2CP03937K>). The remaining articles in the themed collection elaborate on similar problems while expanding to other areas. On the theory side, this includes uncertainty quantification in rolling benchmarks (<https://doi.org/10.1039/D2CP01725C>) and nuclear quantum effects in reaction kinetics (<https://doi.org/10.1039/D2CP03809A>) as well as multi-level schemes for larger system sizes (<https://doi.org/10.1039/D2CP05056K>). In the gas phase as a natural benchmarking habitat, high resolution rotational spectroscopy is certainly among the toughest experiments to be met by theory and it goes far beyond just providing rotational constants (<https://doi.org/10.1039/D2CP05774C>, <https://doi.org/10.1039/D2CP04825F>, <https://doi.org/10.1039/D2CP04067K>, <https://doi.org/10.1039/D2CP03897H>), often with several research groups joining forces (<https://doi.org/10.1039/D2CP04663F>, <https://doi.org/10.1039/D2CP04060C>, <https://doi.org/10.1039/D2CP03962A>). The open questions

encompass the description of quadrupole coupling, the proper description of conformational flexibility at an acceptable computational cost, and the application of rotational spectroscopy for chirality analysis. Describing the coupling between rotation and vibration, especially large amplitude motion, is still a grand challenge (<https://doi.org/10.1039/D2CP03897H>). Size and conformer-selective characterisation of the excited-state deactivation pathways or processes involved in host–guest interactions or molecular recognition are used for assessing the validity of excited-state descriptions or complex potential-energy surfaces commonly used by experimentalists. (<https://doi.org/10.1039/D2CP04570B>, <https://doi.org/10.1039/D2CP03796C>, <https://doi.org/10.1039/D2CP03953B>, <https://doi.org/10.1039/D2CP01414A>, <https://doi.org/10.1039/D2CP03110H>). Several papers extend the applicability range for molecular benchmarking, to radical and biradical complexes (<https://doi.org/10.1039/D2CP04092A>, <https://doi.org/10.1039/D3CP01156A>, <https://doi.org/10.1039/D2CP04101D>, <https://doi.org/10.1039/D2CP03889G>), to non-standard electronic transitions (<https://doi.org/10.1039/D3CP00160A>, <https://doi.org/10.1039/D2CP01623K>) or to X-ray scattering off small molecules (<https://doi.org/10.1039/D2CP02933B>). Other contributions revisit previous benchmarking efforts, such as for intermolecular balances (<https://doi.org/10.1039/D2CP03907A>, <https://doi.org/10.1039/D2CP05141A>), formic acid complexes (<https://doi.org/10.1039/D2CP03893E>, <https://doi.org/10.1039/D2CP04176F>) or for micro-hydration (<https://doi.org/10.1039/D2CP04174J>), and even for elementary diatomic molecules (<https://doi.org/10.1039/D2CP03964H>). New benchmark data sets are presented and used for practical purposes (<https://doi.org/10.1039/D2CP04049B>, <https://doi.org/10.1039/D2CP03992C>, <https://doi.org/10.1039/D2CP04052B>),

with data sizes up to several million (<https://doi.org/10.1039/D2CP03966D>). The extension of studies towards complex systems has resulted in a diversity in the problems tackled experimentally, such as the study of the crystalline phase or complex protein environments (<https://doi.org/10.1039/D2CP04098K>, <https://doi.org/10.1039/D2CP00184E>). The studies are here extended to electrostatic properties (<https://doi.org/10.1039/D2CP04052B>), or to metal surface adsorption (<https://doi.org/10.1039/D2CP04398J>). While biomolecular docking processes (<https://doi.org/10.1039/D2CP04671G>) represent a relatively mature area of benchmarking practice, AI-based approaches are more recent (<https://doi.org/10.1039/D3CP01216F>). The ultimate goal must be to bring these different areas together to better assert the robustness of methods and avoid depending on error cancellation (<https://doi.org/10.1039/D2CP04098K>).

Benchmarking should be a continuous activity, keeping our theoretical models grounded to the highest standard: empirical validation. It requires an incessant review and expansion of references, a critical eye to mismatches and shortcomings plus the insight to propose new theories/approximations which effectively overcome the latter. Bringing different communities together generates common data points to everyone involved, fostering interdisciplinarity. It can also build moments of respite away from the individualistic “publish-or-perish” culture, by sharing experience and data for the greater good.

We thank all colleagues for their illuminating scientific contributions, Vikki Pritchard and Izzy Darlington from PCCP for the management of the themed collection and the Göttingen research training group BENCH for triggering this timeless and still timely topic.