

# Benchmarking Quantum Chemical Methods: Are We Heading in the Right Direction?

Ricardo A. Mata\* and Martin A. Suhm\*

benchmarking · experiments · quantum chemistry · theory

Over the last years a subtle but profound change has taken place in chemical research. Electronic structure calculations have become ubiquitous, with much of the work published in the field today making use of theoretical results. The chemistry curricula have also been accompanying this change, incorporating a growing number of courses on quantum chemistry and computational chemistry, slowly but surely. The reasons behind this quiet revolution are all well known. The computing hardware has undergone significant developments, such that even a mobile phone surpasses the performance of computing clusters from 30 years back. This allows us to compute more and in less time, translating into larger systems, longer timescales. Furthermore, the computer algorithms have improved significantly, expanding the realm of application and the quality of the calculations performed.

We find ourselves at a perceived turning point where quantum chemical calculations are believed by many to be on par with experimental methods. Even if the reader may not share this view, it is not difficult to find reasons why others would. Already in the 1960s we observe examples of theory competing with experiment in terms of accuracy. A seminal example is the  $H_2$  adiabatic dissociation energy computed by Kolos and Wolniewicz.<sup>[1]</sup> The authors carried out a variational calculation and found their value ( $36117.4 \text{ cm}^{-1}$ ) to exceed the best experimental estimate at the time ( $36113.6 \pm 0.6 \text{ cm}^{-1}$ ).<sup>[2]</sup> Given that the theoretical estimate would necessarily give an upper bound to the true energy of  $H_2$ , and thus a lower bound to its dissociation energy, the experimental value was questioned. The episode was only concluded after a further experiment confirmed the theoretical result.<sup>[3]</sup> Since then, there have been several cases where theory has made

predictions which were only later confirmed by experiment,<sup>[4–9]</sup> and many more where theory played an essential role in the interpretation of experimental results.

Already the early  $H_2$  story points to a fascinating aspect, in particular now that the agreement between theory and experiment has progressed to eight significant digits.<sup>[10]</sup> In contrast to other fields of science, where a model is mostly used as a framework to break down and understand complex systems, electronic structure methods are regarded as quantitative-data providers. This is a completely different scenario from that faced in biology or economics, just to name two examples. In many fields, the ruling equations are still unknown, or the conditions are ill-defined. An exact monthly weather forecast is unthinkable, but we chemists dare to believe that our theoretical models will be able to provide us with a reaction mechanism involving hundreds of atoms. This line of thought is fertile ground for illusions of grandeur, and calls for a critical analysis of the relation between experiment and theory in chemistry (Figure 1).

[\*] Dr. R. A. Mata, Dr. M. A. Suhm  
Institut für Physikalische Chemie  
Universität Göttingen  
Tammannstrasse 6, 37077 Göttingen (Germany)  
E-mail: rmata@gwdg.de  
msuhm@gwdg.de

ORCID The ORCID identification number(s) for the author(s) of this article can be found under:  
<http://dx.doi.org/10.1002/anie.201611308>.

© 2017 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited, and is not used for commercial purposes.



Figure 1. A false change of paradigm. Experience has taught us that one should be critical on both ends.

Since we are not philosophers of science (nor pretend to be), we will focus on a pragmatic view and only briefly touch on fundamental issues. In the language of Thomas Kuhn, quantum chemistry is largely in a rather unspectacular state of normal science.<sup>[11]</sup> Apart from subtle issues such as parity violation,<sup>[12]</sup> the Schrödinger equation and its relativistic variants are assumed to be essentially correct and complete for all practical purposes.<sup>[13]</sup> Currently the cutting edge of quantum chemistry is simply modeling larger and larger systems in an increasingly quantitative way.

Our concern is thus evolutionary rather than revolutionary. Beyond the usual method validations within the theoretical community, we address the relation between quantum chemical calculations and experiments. In order to guarantee a fruitful interplay, we need to define how to experimentally evaluate and benchmark the numerous methods in the best way. We argue that this subject is frequently neglected, and that this neglect leads to a slower development of quantum chemistry and the field of chemistry as a whole. While we are unable to provide unique and final recipes, we would like to share some of our thoughts and discussions with the community and conclude with some recommendations. Without prejudice against any of the methodologies covered under the general designation of theoretical/computational chemistry,<sup>[14]</sup> this essay deals exclusively with electronic-structure methods, given their particular appeal and promise towards a first-principles description of chemical systems. Mechanical force fields and topological methods, for example, have a more pragmatic relation to experiment, and do not warrant the same type of discussion. The dynamics and statistics needed to connect the electronic-structure information to the real world would also be an important and critical topic of its own. This is beyond our scope, just like conceptual models and other qualitative theories.



Ricardo Mata has been a research group leader at the University of Göttingen since 2009, working in the field of theoretical chemistry. His main topics of interest include weak molecular interactions, incremental/fragment-based approaches to wave function methods, and applications in enzymatic catalysis. Among some of his main concerns are the applicability and accuracy of computational approaches moving from small to complex chemical systems.



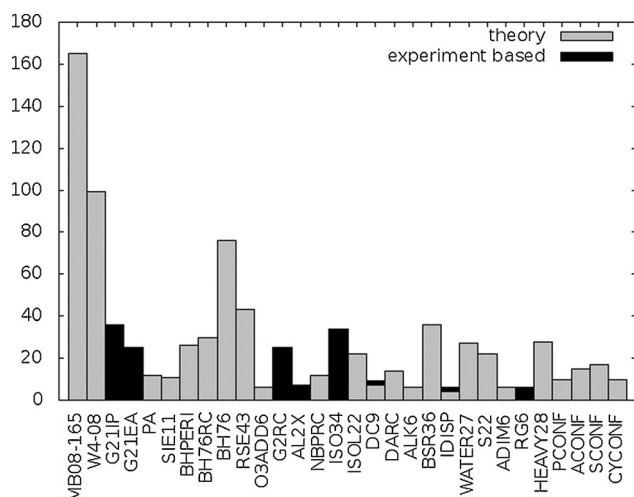
Martin Suhm likes to bring experiment closer to theory in the field of intermolecular interactions, to assist in identifying weaknesses and strengths in standard quantum chemical approaches. Over the last two decades, his Göttingen-based group has combined linear IR and Raman spectroscopy tools with supersonic jet expansions to unravel the structure, vibrational dynamics, and energetics of flexible molecules and van der Waals complexes. He is a member of the Editorial Board of *Angewandte Chemie*.

## Theory Benchmarking Theory

One permanent issue in quantum chemistry is to establish a practical overview of the many different approximations to the Schrödinger equation. Literally hundreds of electronic-structure methods are available at our fingertips, and it can be a cumbersome task to compare them or even to select one for a particular application. One way to deal with this question has been to establish a hierarchy of models. In the case of wave-function methods, the process can be rather straightforward. For single-reference correlated methods there are widely accepted orderings. The latter can be better understood through the use of diagrammatic representations. Generally, the method that includes the greater number of terms lies higher in the ranking. However, this is not necessarily true, since the nature of each term is different along with the impact on the overall performance. While a ranking for coupled cluster (CC) methods such as  $\text{CCSD} < \text{CCSDT} < \text{CCSDTQ} < \text{CCSDTQP}$  is rather straightforward, the ordering within the Møller–Plesset perturbation series (MP2, MP3, MP4, ...) will depend on the system being studied, since it is not necessarily convergent.<sup>[15]</sup> And although CCSDT is a more expensive approach, the CCSD(T) method is for most application purposes a more robust method than the full triples variant. This results from a quite favorable cancellation of error between the overestimated triples contribution and the neglect of quadruple excitations.<sup>[16]</sup>

Also in the case of density functional theory (DFT) there have been attempts at establishing similar hierarchies. The most well-known example is the DFT Jacob's Ladder, proposed by Perdew,<sup>[17]</sup> which defines a set of steps featuring the different levels of approximation in the exchange-correlation kernels. Even for artificial molecules, such hierarchies have been shown to shine through.<sup>[18]</sup> It is, however, relatively easy to find individual examples where a functional from a lower rank may be able to outperform methods higher up in the ladder<sup>[19]</sup> and truth be told, this was not the intended purpose of Perdew's proposal.

Although such hierarchies can be questionable, they provide a very welcome order amidst the chaos of modern-day quantum chemistry toolboxes. They have also created the possibility to carry out theory benchmarks with theory as reference. The quality of a model is thereby no longer measured through any relation to experiment, but purely to the similarity to another model. This practice has become so popular that many manuscripts dedicated to quantum chemistry benchmarks do not feature a single experimental result.<sup>[20]</sup> Sometimes, the word “experiment” is missing from the entire manuscript, or may at best be found once in the outlook. The current most complete set of benchmark sets, the GMTKN30,<sup>[21]</sup> includes only a small amount of experimental reference data (Figure 2). The acceptance of CCSD(T) as a “gold standard”<sup>[22]</sup> has been a particular encouragement to the practice, from DFT benchmarking<sup>[23,24]</sup> up to theory-only blind-test challenges.<sup>[25]</sup> As further evidence, if one looks again at the GMTKN30 database, 14/30 sets use as reference data estimated CCSD(T)/CBS limits. Even improvements beyond CCSD(T) are sometimes judged without any reference to experiment.<sup>[26]</sup> This is certainly related to the



**Figure 2.** Reference points in the GMTKN30 database, classified according to their origin (pure theoretical values or experiment-based).

fact that most such calculations refer to a fixed geometry, although the importance of structural relaxation at least for selected degrees of freedom has been emphasized.<sup>[27]</sup>

There are good reasons for theory-only benchmarking. Even weather forecasters can develop a reasonable sense for their simulation error by just comparing the results of disparate numerical models for the same starting conditions. Another obvious application is the comparison of different codes and numerical strategies within a model family.<sup>[28]</sup> For the identification of isolated experimental database errors, sometimes even superficial comparisons between different computational predictions suffice.<sup>[29]</sup> One may also be interested in a single quantity which is ill-defined or hard/impossible to measure experimentally (e.g., harmonic spectra,<sup>[30]</sup> or transition-state structures<sup>[31,32]</sup>) so that the best (or only) reference is the result of another calculation. Or there may simply be no satisfactory experimental data available or in reach for a specific class of compounds.<sup>[33]</sup> Current state-of-the-art quantum chemical methods depend on a wide variety of approximations. They range from more technical issues such as density fitting<sup>[34,35]</sup> and numerical grids,<sup>[36]</sup> up to developments such as explicit correlation,<sup>[37]</sup> trying to reach the complete basis set limit. All these require the comparison to internal standards.

However, the literature is filled with examples where the comparison to experiment is completely disregarded for no particularly good reason. There are many causes for this fault. On one hand, we have the focused education of theoretical chemists, who dedicate several years to equations and little time to lab experience. In other cases, one may find it difficult to establish relations between what has been measured in experiment and the results of calculations, or even in understanding the approximations the experiment itself involves. More trivially, insufficient time is spent in searching the literature due to the all too well known publication pressure.<sup>[38]</sup> All of these issues can result in misguided comparisons (oranges and apples) leading to frustration or flawed benchmarking, or to a complete neglect of the published experimental data.

There are also cases where the appropriate experimental literature is cited but not really digested by the theory/theory benchmark. As an example, we pick the ethanol dimer, which was shown semi-experimentally to revert the conformational preference of its monomer (*trans*) to a homochiral double-*gauche* arrangement in the lowest-energy pair structure.<sup>[39,40]</sup> The driving force is obviously a compact, dispersion-optimizing packing. A recent computational study claiming accurate ( $1 \text{ kJ mol}^{-1}$ ) ab initio approximations for longer-chain alcohol clusters<sup>[41]</sup> uses ethanol dimer as a stepping stone for the validation of their methods, which is not an unreasonable approach. Based on the results of a series of calculations of perceived increasing accuracy, the authors argue that in contrast to all the cited experimental and high-level computational evidence, the double-*trans* dimer is systematically the most stable dimer. This is due to a range of misconceptions, not the least of which is the overlooked difference between homo- and heterochiral pairings and the resulting apparently decreased importance of dispersion corrections. Even if some of the targeted cluster quantities happen to be close to older thermodynamic data, this is clearly no match for the right reason. Careful comparison to available spectroscopic data would immediately have revealed the flaws in this laborious study.

We end this section with a little story told by Coulson<sup>[42]</sup> about an exhibition of quantum chemistry in Paris after World War II. There were lovely diagrams of resonance structures of benzene and excellent numerical illustrations of the lowering in energy produced by them. But Linus Pauling, as he went round that exhibition and came to these diagrams, said, “Why don’t you put a bottle of the stuff by the side of the diagrams?”

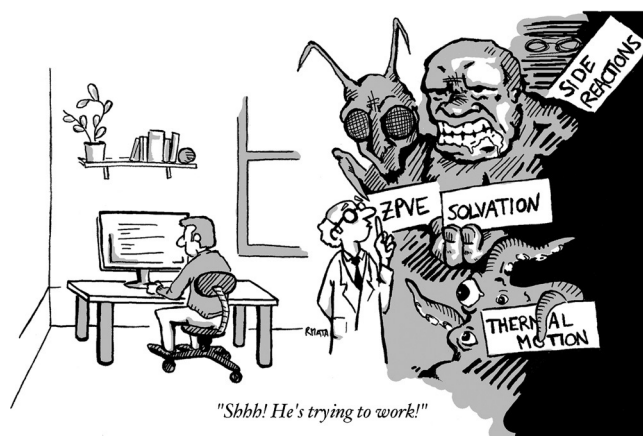
### Experiment Benchmarking Theory

At the end of the day, any weather-forecasting model has to be tested against reality, and success for one season or region is no guarantee for further successful predictions. Chemists also want reliable forecasts for exotic conditions,<sup>[43]</sup> not only for well-trod paths. And even for those, there may be occasional surprises. In the end, there is no way around some experimental benchmarking. The benefits of this practice are obvious, when one looks back at the development of electronic-structure methods. Even though quantum chemists will often find shortcomings in their methods when performing theory/theory benchmarks, it is usually the hard test against experiment that brings about a change in practice. A recent example of this is the description of dispersion forces. Although conventional DFT functionals were well known to fail in the description of weakly bound van der Waals complexes,<sup>[44]</sup> until about ten years ago a large community regarded this as a minor inconvenience for computational studies on “realistic” systems. Looking at the interaction curves of rare-gas dimers was not convincing enough, and it was uncertain whether these small contributions would cancel out or add up. A real change only came about when a series of benchmark studies with increasing system sizes, supported by experimental evidence, unequivocally showed that the DFT



functionals in use at the time were systematically biased, and that the errors were in fact adding up.<sup>[45–51]</sup>

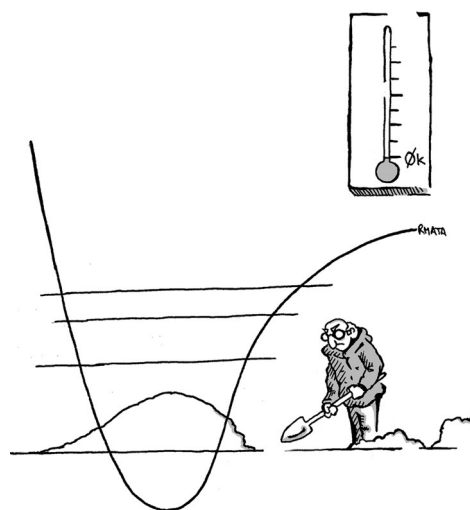
Obviously, experiments aiming to benchmark theory need to be carefully designed, and the right questions have to be posed. For the macroscopic properties of a multilayer insulation material, quantum chemistry may at the start not be more quantitative than models in biology or economics. Reductionism is called for. Experimentalists must provide answers to popular excuses from the theory side (Figure 3): System too big? Let's boil it down to the essentials to meet theory.<sup>[52]</sup> Thermal motion? Let's cool down the sample for a start.<sup>[53]</sup> Environmental effects? Let's move into the gas phase for a reliable comparison.<sup>[54]</sup> Experimentalists offer lots of elegant solutions for the apparently contradicting requirements of gas-phase isolation and low temperature. Other complications, such as relativistic or quantum-electrodynamic effects, cannot be removed by the experimentalist.



**Figure 3.** Experimentalists can also play a role in removing complexity in different chemical systems and providing the best conditions for benchmarking.

There is something less exotic which is hard for experimentalists to get rid of: zero-point vibrational energy (ZPVE), or separated quantum delocalization of the nuclei, an artifact of the Born–Oppenheimer approximation which is so fruitful in electronic structure theory (Figure 4). At least for the study of intermolecular interactions, where theory/theory benchmarking is very popular, this can limit the system size more than one would wish for. Progress is definitely needed from both sides in this area and it is helpful to have careful analyses on how to best bridge this.<sup>[55,56]</sup>

Experimental benchmarking requires a special state of mind. The first goal is to identify any possible weakness in the experimental approach chosen to test quantum chemical methods. This requires phases of intense interaction with theory but also events of blind competition,<sup>[57]</sup> because, as Feynman justly emphasized:<sup>[58]</sup> *The first principle is that you must not fool yourself—and you are the easiest person to fool.* If the experimental result is known beforehand to the theoretician or the theoretical prediction known to the experimentalist, there is a non-zero risk of mutual influence, however rationally the scientists may approach the challenge.



**Figure 4.** A resilient obstacle between theory and experiment: the ZPVE.

Another useful factor is molecular diversity, something chemists are very good at generating. A single match between theory and experiment is almost worthless, because it could always be the result of fortuitous error cancellation. A broad, disparate data set is usually needed, and this means a lot of work at the laboratory bench. Because it has become widespread to train density functionals on such benchmark data sets (so far mostly theoretical ones, we regret), there is a continuous demand for new, reliable experiments.

While the credo of the theoretical benchmarker is “calculate observables!”, the motto of the benchmark experimentalist should be “measure fundamental quantities!”. The hyperpolarizability of a nonlinear optical material or the third virial coefficient of a polar gas may be highly relevant quantities to which quantum mechanics contributes importantly, but whether or not a particular quantum chemical method captures the essentials can often be seen most transparently by simply looking at structures<sup>[56]</sup> and energies,<sup>[59,60]</sup> properties of the stationary points on the potential energy hypersurface. This explains why spectroscopic methods play a central role in benchmarking. No higher-order property can be predicted correctly for the correct reason, if the predicted structure or energetics is far off. Therefore, benchmarking experimentalists must sometimes make an effort to move away from their favorite observables and system conditions and try to address more basic ones. In this way, the possibilities for error compensation are minimized. A meaningful approach to the judging bench usually means more effort at the laboratory bench. This also allows much higher levels of computation to be used and validated.<sup>[61]</sup> Ideally, these could subsequently serve as secondary standards for theory/theory comparisons.<sup>[62]</sup>

In organic chemistry, absolute energies are often less important than conformational preferences. Therefore, experiments on such energy rankings of different structures for a given system may also be useful for benchmarking purposes. Again, it is beneficial for the comparison to quantum chemistry to remove thermal excitation and the environment.

For intramolecular isomerism, the case of glucose may be mentioned,<sup>[54]</sup> although one should remember that the global-minimum structure for C<sub>6</sub>O<sub>6</sub>H<sub>12</sub> is a completely different one.<sup>[63]</sup> Solvation preferences can be studied by intermolecular balance experiments as a function of chemical substitution.<sup>[64]</sup> It turns out the balance between  $\pi$ -cloud and oxygen solvation of anisoles is not captured correctly by the M06-2X functional, whereas the B3LYP-D3 functional performs satisfactorily. Whether this is for a good reason can be found out by extending the data base.

Numerous good-practice examples of benchmark comparisons to experiment may be found in the literature and only a few recent examples can be mentioned here. The Hättig group compared different methods for calculating adiabatic electronic excitation energies based on a large high-resolution experimental database compiled by Leutwyler.<sup>[65]</sup> Vibrational frequencies can be compared rather rigorously to experiment, at least for small molecules, and their benchmarking<sup>[66,67]</sup> contributes to a resolution of the ZPVE problem. Grimme's group compared experimental rotational constants (after ZPVE correction) to theoretical predictions<sup>[68]</sup> and also cautiously explored back-corrected experimental noncovalent interactions for very large systems.<sup>[69]</sup> Such attempts to bring experiment and theory together despite environmental effects can be challenged in a subsequent step by other theoretical methods,<sup>[70]</sup> leading to a healthy competition between different approaches. A recurrent conclusion from such scientific dialogues is the need for more accurate low-temperature gas-phase reference data for complex formation energies.<sup>[53]</sup> Thermodynamic data in the gas phase provide a particularly rich and frequently explored source for stable-molecule energetics.<sup>[71]</sup> On the other end, periodic solids offer access to interaction energies through experimental sublimation enthalpies.<sup>[72]</sup>

Chemistry is about change and it would be far too restrictive for benchmarking activities to stay at the static level of structures and energies. It is worth mentioning the counterpart to the H<sub>2</sub> energy challenge in reaction dynamics and kinetics: the H + H<sub>2</sub> reaction. Ever since the archetypical representation of its reaction surface<sup>[73]</sup> it has been the subject of countless detailed experimental benchmarking efforts.<sup>[74]</sup> Scattering experiments and time-resolved spectroscopy continue to provide the most valuable reference data for chemical dynamics and thus the time-dependent Schrödinger equation. This also brings us to another point. Advances are not only made through benchmark databases. Exemplary comparisons between theory and experiment for selected molecules are also valuable, because they have the potential to push the frontiers of methodology. Impressive examples include alanine,<sup>[75]</sup> butane,<sup>[76]</sup> anionic gold clusters,<sup>[77]</sup> and even condensed water.<sup>[78]</sup> It then remains to investigate whether such frontier methods/approaches are generally applicable to other chemical systems.

In this context, the role of benchmarking experiments is occasionally reversed.<sup>[79]</sup> Rather than challenging theory, it may be that a persistent gap turns out to be an experimental misinterpretation, or even measurement error. When this happens, quantum chemists have done their homework with careful model assessment, or experimentalists have simply

been too optimistic in their assumptions. The latter is more often the case for old experimental reference values,<sup>[80,81]</sup> which are easily shown to be inconsistent with state-of-the-art quantum chemical predictions.<sup>[82,83]</sup> Today, countless experimental misinterpretations never surface because they are resolved before publication through intense interaction with theory, which is of course an excellent practice beyond benchmarking. Sometimes, there is a very fundamental and illuminating reason why theory and experiment cannot match, such as for the gap between electronic states of weakly bound dimers.<sup>[84]</sup> More often, the opposite situation is found: Experiment and theory agree perfectly with each other for a specific quantity, but for no good reason, because two or more errors cancel.<sup>[85]</sup> A recent example is the switch from stretched linear alkanes to dispersion-driven hairpin conformations as a function of chain length. The MM2 force field prediction is within experimental error,<sup>[86]</sup> but only because it overestimates the conformational folding energy and the dispersion attraction between the two chain ends by similar amounts.<sup>[89]</sup> Anyone can be lucky in predicting snow for next Christmas! Different variants of CCSD(T) calculations also agree with experiment<sup>[86–88]</sup> but now for much better reasons. It is actually quite possible that these calculations are more accurate than the experiment, which retains an estimated error bar of  $\pm 1$  chain segments.<sup>[89]</sup> This remains a possibility despite multiple spectroscopic evidence and the ZPVE bottleneck in the calculations. But we will probably only know for sure once a fundamentally improved experiment becomes available—possibly some kind of clever action spectroscopy on ultracold trapped molecules.

## Conclusions

Like powerful action spectroscopies which have to be sufficiently validated by linear spectroscopies with more limited application range, powerful approximate models have to be carefully checked by higher-quality (but size-limited) approximations in quantum mechanics. In view of such important and straightforward intradisciplinary calibration activities, the more sophisticated and more tedious cross-disciplinary benchmarking between theory and experiment tends to be neglected. This must change for the good of science, because both disciplines profit enormously from periodic cross-fertilization. We advocate that it would be very helpful if junior quantum chemists and experimentalists were co-educated for the activity of benchmarking in research-training groups and research units. Weather forecasters should learn to check the real weather and weather-dependent activities profit from a deeper understanding of the forecasting models.

In the end, we are all sinners, but in order to provide the best possible conditions for quantum chemistry to improve, some guidelines could perhaps be of use. Based on our views and the recent history detailed above, we would like to suggest:

- Theoreticians should focus on measurable quantities, and/or a well-defined connection between the computed

quantity and the experimental data. They should keep in touch with experimental colleagues and seek advice.

- In the case of theory/theory benchmarks, reviewers should also play a role in verifying whether pertinent experimental data are missing. Editors can contribute by regularly assigning an extra experimental reviewer.
- Experimentalists should contact theory colleagues and work collaboratively to establish realistic theory/experiment comparisons. They should insist that theoretical results and methods can and should be experimentally tested and in some cases even disqualified.
- Theoreticians should challenge published experimental results whenever the results of their calculations are in contradiction although they have strong reasons to believe that they should not be. The ensuing discussion can only benefit both communities.
- Regular blind challenges should be established on selected experimental benchmark results.
- To help in all the above mentioned points, there should be an effort to establish experimental-benchmark data sets for use in the quantum chemistry community, rather than only the popular theory-based data sets. This effort should be supported by scientific chemical societies.
- National institutes of standards should play an active role in the latter point, managing cross-checking and critical reviews, supporting efforts in interrelating databases from different subfields of chemistry. Crystallographic databases may serve as good-practice examples.

Apart from the benchmark studies, theoreticians and experimentalists alike could also contribute by changing some of their habits, so as to lower existing barriers:

Theoreticians should:

- not only mention the “successful” methods, but also take the opportunity to address the approaches that fail, in particular those with wide acceptance in the community (Reviewers should be open to this type of additions.)
- document their work properly (They should provide all needed material to ensure reproducibility and make proper use of the Supporting Information (provide structures, energies). This has to be supported by the editors of scientific journals. There is no current guideline on what a theoretical study should provide in the Supporting Information, and there is little involvement of the editors/referees in defining what should be included there. A universal file format for storing energy and coordinates would be helpful.)
- provide whenever possible estimates and comments on error sources and crucial approximations in the calculations carried out.

Experimentalists should

- promote to the quantum chemistry community results that are particularly amenable to computation and benchmarking
- promote to the quantum chemistry community conflicting experimental results where theory (or adjudicating experiments) could help in a decision

- present talks or posters on specialized theoretical chemistry conferences (WATOC, Symposium für Theoretische Chemie, ...) to support the former points
- rectify outdated experimental data by the available publication instruments of revisit, critical review, erratum, retraction, comment.

This discussion leaves out, nonetheless, perhaps the biggest threat to benchmarking activities in quantum chemistry: the growing political pressure in several countries to focus only on immediately useful, applied science. Experimentalists and theoreticians alike have to fight for the freedom to contemplate the fundamental meeting points between their methods. And on this note we conclude, with the reminder that fortuitous success will not take us very far.

### Acknowledgements

This essay was triggered, co-designed, and enriched by a great experimentalist, critical reviewer, and persistent advocate of benchmark experiments with excellent connections to quantum chemistry. We dedicate it to Prof. Samuel Leutwyler on the occasion of his 65th birthday. We also thank our departmental and methodical colleagues for enlightening discussions, as well as the Stiftung der Georg-August-Universität Göttingen for financial support.

### Conflict of interest

The authors declare no conflict of interest.

**How to cite:** *Angew. Chem. Int. Ed.* **2017**, *56*, 11011–11018  
*Angew. Chem.* **2017**, *129*, 11155–11163

- 
- [1] W. Kolos, L. Wolniewicz, *J. Chem. Phys.* **1968**, *49*, 404–410.
  - [2] G. Herzberg, A. Monfils, *J. Mol. Spectrosc.* **1961**, *5*, 482–498.
  - [3] G. Herzberg, *J. Mol. Spectrosc.* **1970**, *33*, 147–168.
  - [4] N. Shenvi, S. Roy, J. C. Tully, *Science* **2009**, *326*, 829–832.
  - [5] N. Bartels, K. Golibrzuch, C. Bartels, L. Chen, D. J. Auerbach, A. M. Wodtke, T. Schäfer, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17738–17743.
  - [6] A. Jain, Y. Shin, K. A. Persson, *Nat. Rev. Mater.* **2016**, *1*, 15004.
  - [7] P. Pyykkö, *Phys. Chem. Chem. Phys.* **2012**, *14*, 14734–14742.
  - [8] S. Grimme, *Chem. Eur. J.* **2004**, *10*, 3423–3429.
  - [9] H. Wolf, D. Leusser, M. R. V. Jørgensen, R. Herbst-Irmer, Y.-S. Chen, E.-W. Scheidt, W. Scherer, B. B. Iversen, D. Stalke, *Chem. Eur. J.* **2014**, *20*, 7048–7053.
  - [10] J. Liu, E. J. Salumbides, U. Hollenstein, J. C. J. Koelemeij, K. S. E. Eikema, W. Ubachs, F. Merkt, *J. Chem. Phys.* **2009**, *130*, 174306.
  - [11] T. S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, **1962**.
  - [12] M. Quack, *Angew. Chem. Int. Ed.* **2002**, *41*, 4618–4630; *Angew. Chem.* **2002**, *114*, 4812–4825.
  - [13] P. A. M. Dirac, *Proc. R. Soc. London Ser. A* **1929**, *123*, 714–733.
  - [14] W. Thiel, *Angew. Chem. Int. Ed.* **2011**, *50*, 9216–9217; *Angew. Chem.* **2011**, *123*, 9382–9384.
  - [15] F. H. Stillinger, *J. Chem. Phys.* **2000**, *112*, 9711–9715.
  - [16] T. Helgaker, W. Klopper, A. Halkier, K. L. Bak, P. Jørgensen, J. Olsen in *Understanding Chemical Reactivity*, Vol. 22 of *Quan-*



- tum-Mechanical Prediction of Thermochemical Data* (Ed.: J. Cioslowski), Kluwer, Dordrecht, **2001**, pp. 1–30.
- [17] “Jacob’s ladder of density functional approximations for the exchange-correlation energy”: J. P. Perdew, K. Schmidt in *Density Functional Theory and Its Applications to Materials* (Eds.: V. E. Van Doren, K. Van Alseoy, P. Geerlings), AIP, New York **2001**.
- [18] M. Korth, S. Grimme, *J. Chem. Theory Comput.* **2009**, *5*, 993–1003.
- [19] S. F. Sousa, P. Alexandrino, M. J. Ramos, *J. Phys. Chem. A* **2007**, *111*, 10439–10452.
- [20] A Web of Science author search (November 2016) for the topic “benchmark” in the category “Chemistry Physical” returns something like a who’s who in theoretical chemistry. None of the top 60(!) entries according to publication count would consider themselves to be primarily experimentalists. With all reservation about bibliometrics and buzzwords—could there be stronger statistical evidence that experiment is desperately needed in this field?
- [21] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 291–309.
- [22] It is somewhat difficult to trace back to the first discussion of the term. One of the earliest citations is found in J. F. Stanton, J. Gauss, *Adv. Chem. Phys.* **2003**, *125*, 110, in which personal communication with J. M. L. Martin is referenced. Martin, in turn, credits Thom Dunning in a manuscript published in the same year: M. A. Iron, M. Oren, J. M. L. Martin, *Mol. Phys.* **2003**, *101*, 1345–1361. We leave the rest to the science historians.
- [23] T. Anacker, J. Friedrich, *J. Comput. Chem.* **2014**, *35*, 634–643.
- [24] A. M. Teale, O. B. Lutnaes, T. Helgaker, D. J. Tozer, J. Gauss, *J. Chem. Phys.* **2013**, *138*, 024111.
- [25] D. E. Taylor, J. G. Ángyán, G. Galli, C. Zhang, F. Gygi, K. Hirao, J. W. Song, K. Raul, O. A. von Lilienfeld, R. Podeszwa, I. W. Bulik, T. M. Henderson, G. E. Scuseria, J. Toulouse, R. Peverati, D. G. Truhlar, K. Szalewicz, *J. Chem. Phys.* **2016**, *145*, 124105.
- [26] J. Rezac, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- [27] J. Witte, M. Goldey, J. B. Neaton, M. Head-Gordon, *J. Chem. Theory Comput.* **2015**, *11*, 1481–1492.
- [28] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iusan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepnik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, S. Cottenier, *Science* **2016**, *351*, 1415.
- [29] A. L. Hickey, C. N. Rowley, *J. Phys. Chem. A* **2014**, *118*, 3678–3687.
- [30] J. C. Howard, G. S. Tschumper, *J. Chem. Theory Comput.* **2015**, *11*, 2126–2136.
- [31] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 6908–6918.
- [32] L. Simón, J. M. Goodman, *Org. Biomol. Chem.* **2011**, *9*, 689–700.
- [33] A. Karton, S. Daon, J. M. L. Martin, *Chem. Phys. Lett.* **2011**, *510*, 165–178.
- [34] V. D. Domínguez-Soria, G. Geudtner, J. L. Morales, P. Calaminici, A. M. Köster, *J. Chem. Phys.* **2009**, *131*, 124102.
- [35] T. Hrenar, G. Rauhut, H.-J. Werner, *J. Phys. Chem. A* **2006**, *110*, 2060–2064.
- [36] J. M. L. Martin, C. W. Bauschlicher, Jr., A. Ricca, *Comput. Phys. Commun.* **2001**, *133*, 189–201.
- [37] H.-J. Werner, G. Knizia, T. B. Adler, O. Marchetti, *Z. Phys. Chem.* **2010**, *224*, 493–511.
- [38] M. Quack, M. A. Suhm, *Theor. Chim. Acta* **1996**, *93*, 61–65.
- [39] C. Emmeluth, V. Dyczmons, T. Kinzel, P. Botschwina, M. A. Suhm, M. Yáñez, *Phys. Chem. Chem. Phys.* **2005**, *7*, 991–997.
- [40] J. Paul, I. Hearn, R. V. Cobley, B. J. Howard, *J. Chem. Phys.* **2005**, *123*, 134324.
- [41] M. Umer, W. A. Kopp, K. Leonhard, *J. Chem. Phys.* **2015**, *143*, 214306.
- [42] C. A. Coulson, *Rev. Mod. Phys.* **1960**, *32*, 170–177.
- [43] G. N. Simm, M. Reiher, *J. Chem. Theory Comput.*, **2016**, 2762–2773.
- [44] S. Kristián, P. Pulay, *Chem. Phys. Lett.* **1994**, *229*, 175–180.
- [45] L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, *J. Chem. Phys.* **2000**, *112*, 7374–7383.
- [46] T. M. Gilbert, *J. Phys. Chem. A* **2004**, *108*, 2550–2554.
- [47] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463–1473.
- [48] E. I. Izgorodina, M. L. Coote, L. Radom, *J. Phys. Chem. A* **2005**, *109*, 7558–7566.
- [49] X.-Q. Yao, X.-J. Hou, H. Jiao, H.-W. Xiang, Y.-W. Li, *J. Phys. Chem. A* **2003**, *107*, 9991–9996.
- [50] C. E. Check, T. M. Gilbert, *J. Org. Chem.* **2005**, *70*, 9828–9834.
- [51] M. D. Wodrich, C. Corminboeuf, P. von R. Schleyer, *Org. Lett.* **2006**, *8*, 3631–3634.
- [52] M. Heger, R. A. Mata, M. A. Suhm, *Chem. Sci.* **2015**, *6*, 3738–3745.
- [53] S. Maity, R. Knochenmuss, C. Holzer, G. Féraud, J. Frey, W. Klopper, S. Leutwyler, *J. Chem. Phys.* **2016**, *145*, 164304.
- [54] J. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.
- [55] C. Puzzarini, J. F. Stanton, J. Gauss, *Int. Rev. Phys. Chem.* **2010**, *29*, 273–367.
- [56] M. K. Kesharwani, B. Brauer, J. M. L. Martin, *J. Phys. Chem. A* **2015**, *119*, 1701–1714.
- [57] A. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylisma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio, Jr., A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meeke, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, C. R. Groom, *Acta Crystallogr. Sect. B* **2016**, *72*, 439–459.
- [58] R. P. Feynman, *Surely You’re Joking, Mr. Feynman*, Unwin Paperbacks, Boston, **1985**, p. 343.
- [59] K. Peterson, T. H. Dunning, Jr., *J. Chem. Phys.* **1995**, *102*, 2032–2041.
- [60] J. A. Frey, C. Holzer, W. Klopper, S. Leutwyler, *Chem. Rev.* **2016**, *116*, 5614–5641.
- [61] J. A. Frey, S. Leutwyler, *J. Phys. Chem. A* **2006**, *110*, 12512–12518.

- [62] S. Kozuch, D. Gruzman, J. M. L. Martin, *J. Phys. Chem. C*, **2010**, 20801–20808.
- [63] J. Altnöder, K. Krüger, D. Borodin, L. Reuter, D. Rohleder, F. Hecker, R. A. Schulz, X. T. Nguyen, H. Preiß, M. Eckhoff, M. Levien, M. A. Suhm, *Chem. Rec.* **2014**, *14*, 1116–1133.
- [64] H. C. Gottschalk, J. Altnöder, M. Heger, M. A. Suhm, *Angew. Chem. Int. Ed.* **2016**, *55*, 1921–1924; *Angew. Chem.* **2016**, *128*, 1955–1959.
- [65] N. O. C. Winter, N. K. Graf, S. Leutwyler, C. Hättig, *Phys. Chem. Chem. Phys.* **2013**, *15*, 6623–6630.
- [66] D. P. Tew, W. Klopper, M. Heckert, J. Gauss, *J. Phys. Chem. A* **2007**, *111*, 11242–11248.
- [67] J. C. Howard, J. L. Gray, A. J. Hardwick, L. T. Nguyen, G. S. Tschumper, *J. Chem. Theory Comput.* **2014**, *10*, 5426–5435.
- [68] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.
- [69] T. Risthaus, S. Grimme, *J. Chem. Theory Comput.* **2013**, *9*, 1580–1591.
- [70] A. Ambrosetti, D. Alfè, R. A. DiStasio, Jr., A. Tkatchenko, *J. Phys. Chem. Lett.* **2014**, *5*, 849–855.
- [71] B. Ruscic, D. Feller, K. A. Peterson, *Theor. Chem. Acc.* **2014**, *133*, 1415.
- [72] A. M. Reilly, A. Tkatchenko, *J. Chem. Phys.* **2013**, *139*, 024705.
- [73] M. Polanyi, H. Eyring, *Z. Phys. Chem. Abt. B* **1931**, *12*, 279–311.
- [74] K. Koszinowski, N. T. Goldberg, J. Zhang, R. N. Zare, F. Bouakline, S. C. Althorpe, *J. Chem. Phys.* **2007**, *127*, 124315.
- [75] H. Jaeger, H. F. Schaefer III, J. Demaison, A. G. Császár, W. D. Allen, *J. Chem. Theory Comput.* **2010**, *6*, 3066–3078.
- [76] D. Barna, B. Nagy, J. Csontos, A. G. Császár, G. Tasi, *J. Chem. Theory Comput.* **2012**, *8*, 479–486.
- [77] M. P. Johansson, A. Lechtken, D. Schooss, M. K. Kappes, F. Furche, *Phys. Rev. A* **2008**, *77*, 053202.
- [78] T. Morawietz, A. Singraber, C. Dellago, J. Behler, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8368–8373.
- [79] J. Řezáč, D. Nachtigallová, F. Mazzoni, M. Pasquini, G. Pietraperzia, M. Becucci, K. Müller-Dethlefs, P. Hobza, *Chem. Eur. J.* **2015**, *21*, 6740–6746.
- [80] E. A. Walters, J. R. Grover, M. G. White, *Z. Phys. D* **1986**, *4*, 103–110.
- [81] A. Bizzarri, S. Stolte, J. Reuss, J. G. C. M. Van Duijneveldt-van De Rijdt, F. B. Van Duijneveldt, *Chem. Phys.* **1990**, *143*, 423–435.
- [82] S. Haldar, R. Gnanasekaran, P. Hobza, *Phys. Chem. Chem. Phys.* **2015**, *17*, 26645–26652.
- [83] M. Heger, M. A. Suhm, R. A. Mata, *J. Chem. Phys.* **2014**, *141*, 101105.
- [84] P. Ottiger, S. Leutwyler, H. Köppel, *J. Chem. Phys.* **2012**, *136*, 174308.
- [85] R. Kalescky, E. Kraka, D. Cremer, *J. Chem. Phys.* **2014**, *140*, 084315.
- [86] N. O. B. Lüttchwager, T. N. Wassermann, R. A. Mata, M. A. Suhm, *Angew. Chem. Int. Ed.* **2013**, *52*, 463–466; *Angew. Chem.* **2013**, *125*, 482–485.
- [87] J. Byrd, R. J. Bartlett, J. A. Montgomery, Jr., *J. Phys. Chem. A* **2014**, *118*, 1706–1712.
- [88] D. Liakos, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 2137–2143.
- [89] N. O. B. Lüttchwager, M. A. Suhm, *Soft Matter* **2014**, *10*, 4885–4901.

Manuscript received: November 21, 2016

Revised: December 20, 2016

Final Article published: April 28, 2017