

Reliable Gas-Phase Tautomer Equilibria of Drug-like Molecule Scaffolds and the Issue of Continuum Solvation

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

Bayer (Germany)

Research Article

Keywords:

Posted Date: April 5th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1505283/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Accurate calculation of relative tautomer energies in different environments is a prerequisite to many parameters of relevance in drug discovery. This work provides a thorough benchmark of the semiempirical methods AM1, PM3 and GFN2-xTB, the force-field OPLS4, Hartree-Fock and HF-3c, the density functionals PBEh-3c, B97-3c, r2SCAN-3c, PBE, PBE0, TPSS, r2SCAN, ω -B97X-V, M06-2X, B3LYP, B2PLYP, and MP2 versus the coupled-cluster gold-standard DLPNO-CCSD(T) with def2-QZVPP basis sets applied. The outperforming method identified is M06-2X, whereas r2SCAN-3c is the best-performing method in the set of cost-optimized ones. Application of the identified methods on a challenging subset from the SAMPL2 challenge provides evidence that deviations from experiment is caused by deficiencies of current continuum solvation methods.

Introduction

Tautomers are the chameleons of chemistry. They change dependent on the environment from one structure to another, resulting in prominent changes of their molecular properties like shape, electron density distribution and polarization, hydrogen bonding, lipophilicity, solubility, or pK_a . The general scheme for tautomerism is



The most common case with G being hydrogen is prototropic tautomerism. For X and Y being heteroatoms like nitrogen, oxygen or sulfur, the activation barriers are typically low and the interconversions between the tautomeric forms is more or less instantly when transferring the molecule between different environments, with the observed overall rates depending on the rates for forward and reverse rates. If both are similar in magnitude, the observed rate $k_{obs} \geq 10^6 \text{ s}^{-1}$ is quite high and rises towards the relative diffusion limit as the imbalance increases. Due to the fast equilibrium, there is no way to isolate one tautomer.

Tautomerism strongly affects the apparent potency of compounds if the dominant tautomer is not the target-bound form.⁴ Therefore, the energy difference between the predominant form in water and the one in the binding site with lower but not exactly quantifiable dielectric constant will have to be subtracted from the theoretical binding affinity of the ligand. This energy difference stems from a delicate mix of change in conformation - which is actually an ensemble of accessible conformer states, the change in polarization due to the solvent respectively protein environment, and in many cases the protonation state.⁵

Estimations of the numbers of compounds undergoing tautomerism differ dependent on the dataset considered. Sitzmann et al. identified 67% of the molecules in the NCI Chemical Structure DataBase to have more than one tautomer, whereas for the marketed drugs as of 2009 only 26% existed in more than one tautomeric form.

Predominant tautomers in different environments can be identified and to a certain extent quantified by experiments in gas-phase, different solvents and solids, and the interested reader is referred to the monography edited by Antonov. The experimental techniques described in depths therein are absorption UV-vis spectroscopy, stationary and time-resolved fluorescence spectroscopy, femtosecond pump-probe spectroscopy, or NMR, also by utilizing isotope effects. By X-ray diffraction fixated tautomers as well as thermally or photo-induced tautomerism can be detected.

Nevertheless, a major source of insights into tautomerism is by computation, and here dominated by quantum-chemical methods, but also empirical approaches⁵ and to a certain extent reactive force fields² and quite recently by the combination of machine learning and relative alchemical free energy methods.

Computation consists of two steps: tautomer generation and energy calculation. Tautomer generation is straightforward and many algorithms were implemented in commercial software starting with the Daylight toolkit in 1999 and public tools like rdkit or ambit. Cheminformatics-based tautomer generators systematically apply transformation rules iteratively with pruning

based on search-depth and identicals identification, without guarantee of completeness, and without providing relative tautomer energies.

The second step, the energy calculation, requires quantum mechanical calculations to be performed. In the last 20 years, there were many studies that applied certain combinations of QM method, basis set, implicit solvation model, on single conformers or on ensembles, but there were no systematic benchmarks on the performances of the settings used. Regrettably, the GTN30 QM benchmark set does not contain a tautomer dataset, otherwise this study probably would not have been needed to be performed. The most comprehensive comparison was performed in the SAMPL2 challenge in 2010 that had 20 submissions from 7 participants. In conclusion the organizers stated that for errors of lower than 1.5 kcal/mol one should aim for at least MP2/pVDZ level of theory (though it is known that double-zeta basis sets or below border line for perturbation theory). And finally, Fogarassi showed for the four non-ring systems considered that the convergence with respect to the computational level was always different, without conclusion on a method of choice.

In this publication I therefore focus on benchmarking approximate semiempirical and rigorous quantum chemistry as well as density functional methods on their performance on a set of common tautomeric ensembles, knowing that there exist many more transformations like the ones described in comprehensive publication by Sayle.¹ Aim is to identify a reliable and stable work-horse for daily computation in a drug discovery setting. The best combinations of method and basis sets are then applied to a challenging subset from the SAMPL2 challenge.

Methods

Tautomer structures were generated with Pipeline Pilot 21.2 using the component Enumerate Tautomers³ followed by 3D structure generation with corina. The imine double bond isomers which are distinct tautomers had to be created manually from the result set from Pipeline Pilot.

Hartree-Fock, GFN2-xTB, density-functional theory, Moeller-Plesset and the DLPNO approximation to coupled-cluster CCSD(T) calculations were performed with orca 5.0. Semiempirical AM1, PM3 and OPLS4 force field calculations were performed in Schrödinger maestro 2021-4.

Structures were pre-optimized with GFN2-xTB and optimized with PBE0 with D3 dispersion correction applying Becke-Johnson damping (BJ) and def2-TZVP with def2/J auxiliary bases.

Tautomer energies were obtained using the quadruple-zeta quality basis set Def2-QZVPP¹² and def2/J and def2-QZVPP/C auxiliary functions. Eight different density functionals of rising complexity, the general gradient approximation functional PBE, the meta-GGAs PBE0,^{8,9} TPSS and r2SCAN, the range-separated ω -B97X-V, the hybrids M06-2X and B3LYP, and the double-hybrid B2PLYP were applied. The same basis set was used as for our gold-standard CCSD(T) and for MP2 as a well-established standard for the perturbative treatment of electron correlation. Hartree-Fock baseline energies were determined with def2-SVP basis set.

The approximate respectively low-cost methods tested are the general all-atom force field OPLS4¹¹, the semiempirical methods AM1, PM3 and GFN2-xTB, and the Grimme zoo of "3c" methods with HF-3c, PBEh-3c, B97-3c and r2SCAN-3c.

Solvation free energies in solvent water were added applying the default continuum solvation models of the software used. SMD continuum solvation was applied for HF, HF-3c, MP2, CCSD(T), all DFT and DFT-3c methods. The analytical linearized Poisson-Boltzmann (ALPB) was used for GFN2-xTB, COSMO for AM1 and PM3, and an analytical Generalized-Born/Surface-Area (GB/SA) model for OPLS4.

For extrapolation to basis set limit, the scheme of Zhong et al. was applied for SCF and the scheme from Helgaker et al. for the couple-cluster correlation energy,⁴¹ with the family of correlation-consistent basis sets. Thermochemical corrections to free

energies were added from optimization and calculation of the second derivatives using PBE0-D3(BJ) and the def2-TZVP basis set.

Data Sets

The benchmark dataset was taken from the publication by Milletti et al. They had identified tautomer preferences in PDB and CSD crystal structures and additionally experimental and computed relative energies in water, organic solvents and gas-phase. It comprises of 13 small compounds that are typical fragments found in many drug-like molecule cores and substituents, thus representing common tautomeric moieties.

Focus of the original publication was on hit enrichments in docking studies by inclusion of either all or the low-energy tautomers. The respective tautomer energy predictions there stemmed from a cheminformatics tool that takes into account known tautomer patterns, pKa calculations and matches to generic fragments from prevalent tautomers. According to the authors, the 13 compounds (see Scheme 1) split into two groups: compounds **1–7** undergoing annular tautomerism resulting in relatively low ΔG , whereas compounds **8–13** having higher energy differences ΔG .

Since this work is a benchmark of computational methods, we are not only interested in the two lowest states but in the complete sets. Six of the molecules have only 2 or 3 tautomer states, whereas the other seven have multiple states, including imine cis-trans double bond tautomerism.

The second “Experimental” dataset is taken from the SAMPL2 challenge from 2010. The original set comprises of 34 compounds in three subsets, the explanatory set for which experimental values were given upfront, the obscure set with values disclosed only after closing, and the investigatory set for which no experimental values exists. In this study a subset of 14 challenging tautomer equilibria is used based on the selection by Wieder et al.¹³ with structures given in Scheme 2.

Results

Identification of a Reliable Method with the Benchmark Dataset

Coupled-cluster theory with single, double and perturbative triple excitations is used as the accepted gold-standard for reference energies, with the polarized quadruple-zeta quality basis set def2-QZVPP to rule out significant effects due to basis set size. Basis set effects are probed for by application of double- and triple-zeta basis set and by a correction scheme to complete basis set limit. Finally, a test on the necessity to add thermal corrections to free energies is performed.

The following sections we will first provide results for various density functional methods and second-order Moeller-Plesset theory MP2, applying the same polarized quadruple-zeta basis set def2-QZVPP, followed by pure Hartree-Fock with double-zeta basis set def2-SVP and the HF-3c approximation, three semiempirical methods AM1, PM3 and GFN2-xTB, three low-cost density functionals PBEh-3c, B97-3c and r2SCAN-3c, and finally a state-of-the-art force field with OPLS4 that is widely used in molecular design. The benchmarks are performed for gas phase and implicit solvation in water.

Tautomer state energies from Coupled-Cluster Theory

Before we start with the comparison of the various methods we take a look at the CCSD(T) results. Just as a comment. Since the states in Scheme 1 were defined arbitrarily in a systematic way, we can not per se assume state 1 to be the lowest energy state.

Table 1 provides a summary of the benchmark data. The dataset comprises of five molecules with just two tautomers, but also five molecules with, including imine double bond stereoisomers, up to 12 tautomers. As we will see, imine double bond isomers are tautomers with distinct tautomer energies.

Table 1

Tautomer states and CCSD(T) relative state energy differences in kcal/mol between the lowest to second lowest or lowest to highest states. For molecules with only two tautomers the values in the lower part of the table have been omitted for clarity.

Molecule number	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of states	2	12	8	2	2	9	9	2	6	5	3	2	3
dE gas (2nd-min)	4.56	8.95	1.04	4.16	5.92	4.10	1.81	2.10	10.26	14.41	14.05	7.70	8.87
dE water (2nd-min)	3.33	1.17	0.25	1.23	2.66	1.15	4.37	3.21	8.18	11.39	12.38	11.95	8.65
ddE (gas-water)	1.23	7.78	0.79	2.93	3.26	2.95	-2.56	-1.11	2.08	3.02	1.67	-4.25	0.22
lowest state gas	1	1	2	2	1	1	6	2	1	1	1	1	1
2nd lowest gas	2	2	1	1	2	3a	3b	1	5	3	2b	2	2a
lowest state water	1	1	1	2	1	1	1	1	1	1	1	1	1
2nd lowest water	2	2	2	1	2	2	2	2	2	3	2b	2	2a
dE gas (max-min)		30.72	21.72			26.72	20.23		24.72	31.58	16.79		12.09
dE water (max-min)		18.43	14.74			22.74	20.94		15.60	24.23	13.22		9.64
ddE		12.29	6.98			3.98	-0.71		9.11	7.35	3.56		2.45
lowest state gas		1	2			1	6		1	1	1		1
highest state gas		5a	4			4a	4b		3	5	2a		2b
lowest state water		1	1			1	1		1	1	1		1
highest state water		5a	3			4a	4b		3	5	2a		2b

As stated by the dataset authors, the split into compounds **1–7** undergoing annular tautomerism and **8–13** with tautomerism not located in the ring system, results in two sets with low and higher ΔG between the first and second state, but with one prominent exception, namely 2-pyridone/2-hydroxypyridine **8** that is in the low energy group, too.

Solvation in most of the cases lowers the energy gaps between the lowest to second states and the lowest to highest states compared to gas-phase except for **7**, **8**, **12**. Compound **13** shows this lowering for lowest to highest but about identical gaps for lowest to second.

Solvation also changes the order of states for compounds **3**, **6**, **7**, **8**, **9** in case of lowest to second state and for compounds **3** and **7** in case of lowest to highest states. The change in tautomer preference for compound **8**, 2-hydroxypyridine being the preferred state in gas-phase and 2-pyridone the preferred one in water, is a well-known and prominent example. And as we will see, also a challenging case for most computational methods.

Density Functional Theory and MP2

The benchmarked density functional methods include the generalized gradient approximation functional PBE, the meta-GGA functionals PBE0, TPSS and r2SCAN, the range-separated ω -B97-X, hybrids M06-2X and B3LYP, and the double-hybrid B2PLYP. Solvation free energies were calculated with the SMD solvent model and the solvent water.

First, we look into the rank ordering of the tautomers. Figure 1 provides the profiles for three representative examples for solvated (a-c) and gas-phase (d-f) molecules. Plots for all molecules can be found in Supporting Information Figures S1 and S2.

In case of water solvation all methods rank the tautomer states correctly for 10 out of 13 molecules, whereas MP2 and M06-2X rank-order all states for all molecules correctly. For molecule **3** we find wrong orders for tautomers **3.4**, **3.7** and **3.8**, for **7** inconsistent energies for **7.3b** and **7.6**, and for molecule **9** for **9.3** and **9.6**. Imine double bond stereoisomerism for almost every pair leads to different tautomer energies in molecules **2** (3 out of 4 pairs), **6** (2 out of 3), **7** (2 out of 3) and **11** (1 out of 1).

The gas-phase rank-orders are correctly reproduced by all methods for all molecules except **2** (2.2 and 2.3), **7** (7.1, 7.3b, 7.6), **8** (1 and 2), **9** (9.2 and 9.6). The observation of generally larger energy gaps in gas-phase compared to solvation seen with coupled-cluster for most compounds is reproduced by all methods, being larger in gas phase for eight molecules and lower for three compounds, namely **7**, **8**, and **12**. Again, the double bonds isomers have distinct energies for most of the pairs.

Comparing now water and gas-phase, we find the same ordering of all states for only six molecules as there are **1**, **4**, **5**, **11**, **12**, **13** that have two tautomers and compound **10** having five tautomers. For compounds **2**, **3**, **6**, **7**, **8**, **9** not even the two lowest energy states are conserved between water and gas-phase. Most prominent example is 2-pyridone **8**. In water the 2-pyridone state is predicted to be the preferred state by all methods, but with dE between 2 kcal/mol by MP2 and around 6 kcal/mol by PBE, TPSS and B3LYP, compared to the coupled-cluster value of 3.21 kcal/mol. Contrary, in gas-phase the 2-hydroxypyridine state is preferred by 2.1 kcal/mol according to CCSD(T). Again, PBE, TPSS and B3LYP predict the other state or no energy difference. Only M06-2X reproduces the gold-standard dE's with 3.41 and -2.69 kcal for water and gas-phase, respectively.

Whereas it is encouraging that rank-ordering works mostly for all the methods and MP2 and M06-2X even do a perfect job in this respect, the deviations for the energy differences per each state are significant but are important to identify the method of choice.

The curves in Fig. 1 (and the full set of curves in Supporting Information S1 and S2) already provide some indication that not all methods perform well in this respect. At least half of the lines for the states show significant kinks, i.e. deviations from the reference energies. Figure 2 provides bar charts per state and per method for molecules **2**, **7** and **8** in water. Bar charts for all molecules are provided in Supporting Information Figures S3 and S4 for water and gas-phase environment, respectively.

Table 2

Mean unsigned errors MUE for the individual molecules and mean, median and maximum MUE values and standard deviations in kcal/mol calculated as scaled absolute deviations between CCSD(T) and respective method. Top half of the table reports results in water and bottom half in gas-phase. The basis set used throughout is def2-QZVPP. Best performing method is in bold and italic, MUE equal or below 0.1 kcal/mol in bold.

	PBE	PBE0	TPSS	R2SCAN	wB97-X	M06-2X	B3LYP	MP2	B2PLYP
Continuum solvation									
1	0.21	0.15	0.10	0.11	0.52	0.31	0.19	0.48	0.02
2	0.74	0.30	0.65	0.31	0.16	0.12	0.35	0.57	0.25
3	0.98	0.84	1.16	0.85	1.20	0.29	1.16	0.37	0.75
4	0.06	0.12	0.12	0.03	0.05	0.02	0.05	0.15	0.03
5	0.08	0.36	0.01	0.25	0.10	0.20	0.03	0.20	0.04
6	0.72	1.05	0.97	1.10	0.90	0.18	0.99	0.37	0.81
7	1.44	1.64	1.56	1.72	1.47	0.66	1.63	0.39	1.26
8	1.41	0.99	1.50	1.26	0.86	0.10	1.28	0.61	0.80
9	1.09	1.00	1.45	1.14	1.41	0.30	1.49	0.53	0.92
10	0.42	0.70	0.90	0.60	1.37	0.51	1.15	0.07	0.82
11	1.04	0.24	0.92	0.57	0.09	0.05	0.58	1.22	0.07
12	0.50	0.86	0.29	0.80	0.17	0.44	0.77	0.34	0.42
13	0.51	0.26	0.52	0.27	0.40	0.17	0.66	1.01	0.07
Mean	0.71	0.65	0.78	0.70	0.67	0.26	0.80	0.49	0.48
Median	0.72	0.70	0.90	0.60	0.52	0.20	0.77	0.39	0.42
Stddev	0.45	0.44	0.53	0.49	0.53	0.18	0.52	0.31	0.41
Max	1.44	1.64	1.56	1.72	1.47	0.66	1.63	1.22	1.26
Gas-phase									
1	0.25	0.21	0.12	0.09	0.64	0.45	0.20	0.55	0.02
2	0.88	0.54	0.71	0.48	0.63	0.43	0.31	0.64	0.40
3	0.86	0.43	0.94	0.60	0.81	0.35	0.70	0.29	0.47
4	0.12	0.14	0.17	0.01	0.14	0.06	0.01	0.27	0.02
5	0.02	0.37	0.07	0.25	0.21	0.24	0.01	0.26	0.08
6	0.88	1.16	1.10	1.25	1.13	0.36	1.14	0.59	1.06
7	2.30	1.94	2.27	2.29	1.63	0.64	1.95	0.81	1.66
8	1.49	0.73	1.47	1.11	0.52	0.29	0.98	0.48	0.66
9	1.09	0.71	1.25	0.95	1.06	0.74	0.99	0.33	0.67
10	0.63	0.85	1.02	0.85	1.57	0.45	1.13	0.35	0.92
11	0.90	0.09	0.73	0.28	0.34	0.30	0.22	1.36	0.21

	PBE	PBE0	TPSS	R2SCAN	wB97-X	M06-2X	B3LYP	MP2	B2PLYP
12	0.75	0.98	0.43	0.97	0.12	0.56	0.93	0.42	0.53
13	0.55	0.20	0.51	0.15	0.40	0.15	0.71	1.52	0.08
Mean	0.82	0.64	0.83	0.71	0.71	0.39	0.71	0.61	0.52
Median	0.86	0.54	0.73	0.60	0.63	0.36	0.71	0.48	0.47
Stddev	0.58	0.50	0.59	0.60	0.49	0.18	0.54	0.39	0.46
Max	2.30	1.94	2.27	2.29	1.63	0.74	1.95	1.52	1.66

As can be seen from the three example molecules already, the functionals PBE, PBE0, TPSS, r2SCAN and ω B97-X perform weaker than the others where the picture looks more mixed. The direction of the errors is different between methods for molecule **2**, always positive for all except the lowest energy state for **7**, and positive for molecule **8** except for MP2. Comparing the three plots in Fig. 2, (the y-axis has the same range for all molecules in Fig. 2 but not for the plots in Supporting Information, to allow for optimal representation of the data) one can see that the errors are quite small for all methods in case of **2** but typically 2 kcal/mol or more in case of the other two compounds, with the exception of the functional M06-2X.

Generally speaking, the observations for the calculations with continuum solvation are confirmed by the gas-phase data, but there are differences between molecules for the two settings as shown in the following.

Method performance is quantified in Table 2 that reports mean unsigned errors MUE (aka mean absolute error) per molecule and method, calculated by the sum of the absolute errors of the states between CCSD(T) and the respective method and divided by the number of states.

The observations deduced from the three examples in Fig. 2 are confirmed by MUE statistics. The MUE values (in kcal/mol) over all molecules can be grouped into three classes: M06-2X has mean, median and maximal MUE of 0.26, 0.20 and 0.66 that are significantly lower than in the second group of methods consisting of MP2 (mean 0.48) and B2PLYP (mean 0.48) and finally the other six functionals with mean MUE between 0.65 and 0.80 and max MUE between 1.44 and 1.72. The same overall picture is seen for gas-phase, but with slightly higher values for M06-2X (mean 0.39), MP2 (mean 0.61), B2PLYP (mean 0.52) and another increase in the maximum MUE of 1.63 to 2.30 for the weaker performing DFT methods.

In case of continuum solvation M06-2X is the best performing method in 7 cases, and not far off for the other molecules. MP2 is top performing two times, and B2PLYP three times. For gas-phase M06-2X is ranked best only three times, but not far off otherwise. Overall, M06-2X is highly consistent across compounds as reflected in the low standard deviations of 0.18 kcal/mol in both environments and therefore the preferred method of choice.

Low-cost Methods

Let's now look into more compute cost-effective methods. Pure Hartree-Fock HF/def2-SVP was considered as a baseline here due to its missing electron correlation and limited basis set size. HF-3c with corrections for basis set incompleteness, electron correlation and dispersion was expected to outperform pure HF. With AM1 and PM3 two NDDO semiempirical methods with a long history were included, alongside the recently introduced GFN2-xTB that is now widely used in many areas of research. We also apply three cost-optimized density functional approaches from the Grimme group, namely PBEh-3c, B97-3c and r2SCAN-3c, and finally the general all-atom force-field OPLS4.

Though it is kind of "text-book knowledge" that force-field energies are themselves meaningless when comparing compounds with different configurations, i.e. topologies (what tautomers are), This method-inherent fact is overseen quite often,⁵ and it was eye-opening to find tautomer energy differences to be far off for almost any state of any molecule. Far off here means even of up to hundreds of kcal/mol for some states, as can be seen from the truncated curves in Fig. 3 (similar data were obtained for MMFF test calculations). Therefore, the OPLS4 values are not considered in the following.

As expected, the deviations from the CCSD(T) reference energies are larger for cheaper methods, and less consistent between the methods, as can be seen from Fig. 3 and from the complete sets of state energy plots in water and gas-phase in Supporting Information Figures S5 and S6.

Of the solvated molecules, only **1**, **10**, **12** and **2** (except for almost degenerate states) are ranked correctly by all methods. In gas-phase, this is true only for compound **1**.

Comparison of pure HF with HF-3c reveals that the HF energies in both water and gas-phase are significantly closer to the reference than the HF-3c energies, though the former does not include electron correlation. HF therefore performs well but due to error-compensation. The three semiempirical methods all over- and undershoot for some molecules but are close for others, with errors as high as 10 kcal/mol (**4**, water) or even the AM1 value of -114 kcal/mol for the thiol tautomer of thio-imidazole **12** in gas-phase (outlier value was re-confirmed by AM1 calculations with orca).

For both media the cost-optimized DFT methods perform very well as shown in Fig. 4 (complete sets of plots for water and gas-phase are provided in Supporting Information S7 and S8), especially for the low-energy states, with PBEh-3c slightly inferior to the others. The notable exception again is compound **8**, for which PBEh-3c is much closer to CCSD(T) than B97-3c and r2SCAN-3c.

As before, method performance is quantified by the mean unsigned error MUE between the state energies of the respective method and the reference CCSD(T) for each molecule, as given in Table 3.

For the cost-optimized methods, there is only a slight difference between water and gas-phase performance. Here, we identify three groups of methods. The first group consists of r2SCAN-3c and B97-3c. By far best-performing is r2SCAN-3c with mean, median and maximum MUE of 0.73 (0.72), 0.72 (0.63) and 1.72 (2.09) kcal/mol for water and in brackets gas-phase, respectively. It is followed by B97-3c with mean MUE of slightly more than 1 kcal/mol, but still separated from the other methods by lower maximum errors. The second group consists of PBEh-3c, GFN2-xTB and HF with mean errors below 2 kcal/mol, and finally the NDDO semiempirical methods AM1 and PM3 but also HF-3c, which was Grimme's first attempt for a parametrized low-cost QM method. For those three, the mean and especially the maximum MUE are much too high for practical usage. Even if the AM1 failure for thioimidazole **12** is excluded from the calculation, the mean, median and max gas-phase MUE for AM1 are still 3.91, 3.44 and 11.91 kcal/mol.

The obvious choice thus is r2SCAN-3c, the "swiss-army knife" as it was called by the authors.³⁰ It is the top-performing method for 10 out of 13 compounds in case of water solvation, and for 9 out of 13 in case of gas-phase. Again, this is reflected by the standard deviations of 0.50 and 0.57 kcal/mol in water and gas-phase. It does not reach the quality of M06-2X or MP2 with quadrupole-zeta basis set. Nevertheless, in case compute time is limited, it is the obvious choice, like for larger molecules that require the calculation of multiple conformations and for each conformation all tautomer states. For the small molecules here, r2SCAN-3c takes around 5 to 10 s per single-point, whereas M06-2X is in the range of 2 min.

Table 3

Mean unsigned errors MUE in kcal/mol for the individual molecules and mean, median and maximum MUE values and standard deviations in kcal/mol calculated as scaled absolute deviations between CCSD(T) and respective method. Top half of the table reports results in water and bottom half in in gas-phase. Best performing method is in bold and italic, MUE equal or below 0.5 kcal/mol in bold.

	AM1	PM3	GFN2-xTB	HF	HF-3c	PBEh-3c	B97-3c	r2SCAN-3c
Continuum solvation								
1	0.35	1.21	0.35	1.54	4.43	0.37	0.13	0.13
2	1.61	2.51	2.92	2.36	6.12	2.35	0.41	0.36
3	2.04	2.19	2.43	2.16	3.76	1.39	1.43	0.95
4	5.12	3.59	0.48	0.57	1.47	0.35	0.00	0.18
5	5.33	3.46	0.01	0.69	2.10	0.56	0.13	0.03
6	2.50	1.87	2.10	3.07	3.57	3.02	1.59	0.74
7	2.01	1.47	2.67	2.72	2.46	3.21	2.17	1.72
8	0.38	0.44	1.17	0.79	3.77	0.85	1.87	1.28
9	3.03	1.46	3.25	3.87	4.00	2.09	2.07	1.04
10	4.47	1.23	6.30	4.48	7.81	1.88	1.50	0.31
11	1.61	1.57	1.45	1.71	7.00	1.85	0.65	0.63
12	2.75	3.92	0.28	1.16	8.92	1.50	1.38	1.36
13	1.32	2.74	1.39	0.39	0.35	0.79	0.65	0.72
Mean	2.50	2.13	1.91	1.96	4.29	1.55	1.08	0.73
Median	2.04	1.87	1.45	1.71	3.77	1.50	1.38	0.72
Max	5.33	3.92	6.30	4.48	8.92	3.21	2.17	1.72
Stddev	1.56	1.01	1.63	1.25	2.44	0.92	0.75	0.50
Gas-phase								
1	1.14	0.08	0.11	1.94	4.64	0.56	0.13	0.10
2	2.79	3.51	2.72	3.93	5.74	2.89	0.53	0.43
3	3.83	2.84	1.13	1.62	4.34	1.41	1.24	0.60
4	5.02	4.60	0.35	0.87	1.61	0.39	0.09	0.15
5	5.60	4.98	1.11	1.04	2.60	0.57	0.17	0.01
6	3.13	3.54	1.93	3.09	3.15	3.25	1.99	0.79
7	3.33	1.49	1.83	1.37	2.49	3.28	2.73	2.09
8	0.85	1.21	0.62	0.25	2.74	0.81	1.80	1.10
9	3.54	3.93	3.13	3.12	4.64	2.25	1.92	0.88
10	4.48	4.05	4.12	4.14	7.16	2.40	1.92	0.63
11	1.35	2.13	2.63	2.30	6.47	2.07	0.27	0.39

	AM1	PM3	GFN2-xTB	HF	HF-3c	PBEh-3c	B97-3c	r2SCAN-3c
12	60.92	6.24	0.64	1.61	9.48	1.99	1.68	1.61
13	11.91	2.57	2.53	0.20	0.16	0.89	0.42	0.66
Mean	8.30	3.17	1.76	1.96	4.25	1.75	1.14	0.72
Median	3.54	3.51	1.83	1.62	4.34	1.99	1.24	0.63
Max	60.92	6.24	4.12	4.14	9.48	3.28	2.73	2.09
Stddev	15.43	1.62	1.17	1.24	2.42	1.01	0.87	0.57

Overall, the results with continuum solvation and in gas-phase are quite congruent for each methods and molecules, except for a few notable exceptions, especially molecule **8**.

Basis set Dependence

In the previous section we had compared various density functionals and MP2 with our gold-standard CCSD(T). Since coupled-cluster requires a large basis set to achieve meaningful results, we did apply the same quadruple-zeta quality basis set for DFT and MP2, to be consistent and not mix method and basis set effects. In this section we now explore the actual basis set requirements in two directions, namely extrapolation to complete basis set limit and the minimal basis set requirements.

Table 4
Absolute CCSD(T) energies for 2-pyridone **8.1**, and 2-hydroxypyridine **8.2** and relative tautomer energies for three correlation consistent basis sets and two extrapolation schemes applying either double- and triple-zeta or triple- and quadruple zeta bases.

	<i>E</i> CCSD(T) [hartrees]		<i>dE</i> [kcal/mol]
	8.1	8.2	
cc-pVDZ	-322.6501452	-322.6538478	2.32
cc-pVTZ	-322.9676178	-322.971337	2.33
cc-pVQZ	-323.0639754	-323.0672737	2.07
extrapol. 2/3	-323.1306625	-323.1346141	2.48
extrapol. 3/4	-323.1235078	-323.1266242	1.96

The second test performed is for the minimum basis set requirements for DFT methods, i.e. for the winning DFT method M06-2X. One can expect that the findings for this DF will apply for any other functional as well, as there is additional evidence from the Grimme benchmark paper. The data in Table 5 are for calculations performed in continuum solvation.

Table 5

Mean unsigned errors MUE in kcal/mol for the 13 compounds by the density functional method M06-2X to CCSD(T) applying four different basis sets of double-zeta to quadruple-zeta quality. Summary statistics of mean, median, maximal MUE and standard deviation are provided for the set of molecules.

	def2-SVP	def2-TZVP	def2-QZVPP	6-31G**
1	0.48	0.31	0.31	0.45
2	1.23	0.16	0.12	1.43
3	0.95	0.44	0.29	0.54
4	0.54	0.04	0.02	0.23
5	0.70	0.20	0.20	0.43
6	1.04	0.38	0.18	0.84
7	0.78	0.84	0.66	0.66
8	0.24	0.20	0.10	0.71
9	0.85	0.18	0.30	0.62
10	1.31	0.12	0.51	0.44
11	0.83	0.03	0.05	1.15
12	0.17	0.44	0.44	2.01
13	0.08	0.23	0.25	0.22
Mean	0.71	0.27	0.26	0.75
Median	0.78	0.20	0.25	0.62
STDDEV	0.38	0.21	0.18	0.49
Max	1.31	0.84	0.66	2.01

On the example of the most challenging molecule in the set, compound **8**, the schemes for complete basis set (CBS) extrapolation by Zhong et al. for SCF^{40,41} part and by Helgaker et al. for the correlation energy^{42,41} part are applied to calculations in gas-phase. Results are given in Table 4.

Using three correlation-consistent bases, two extrapolation schemes to complete basis set limit are possible with either double- and triple-zeta or triple- and quadruple-zeta bases. Extrapolation from the lower-level scheme results in a CBS estimate for the tautomer energy difference that is even higher than the respective bases with 2.48 kcal/mol, whereas the higher-level scheme yields a CBS estimate of 1.96 kcal/mol, suggesting that the quadruple-zeta basis is almost converged with a deviation to CBS extrapolation of only 0.11 kcal/mol. The energy difference for the def2-QZVPP basis set is 2.1 kcal/mol (cf. Table 1), i.e. almost identical to the cc-pVDZ basis.

Obviously, there is a clear separation between the two double-zeta bases and the larger bases. Both def2-SVP and the very popular Pople basis 6-31G** show statistical values comparable to or only slightly better than r2SCAN-3c (mean: 0.73; median: 0.72; max: 1.72; Stddev: 0.50), which uses an optimized double-zeta quality basis. On the other hand, the slight added quality of the def2-QZVPP values does not justify the significantly larger compute cost.

Thermochemical Corrections

Energy differences in the previous sections always refer to electronic energy differences, not to differences in free energies. The justification to do so which simplifies the calculation workflow significantly is provided here for the example of 2-pyridone **8**. Frequency calculations with PBE0-D3(BJ) and the def2-TZVP basis set, the level used for geometry optimization, were carried out to yield free energies and their energy components.

Table 6

Free energy correction terms G-E_{el} in kcal/mol for the two tautomers of compound **8** at temperatures of 0 and 298 K, and all energy components thereof in units of hartrees as there are free energy G, entropy S, enthalpy H, zero-point energy ZPE, and thermal and electronic energies E_{thermal} and E_{el}.

	T / K	E _{el} / h	ZPE / h	E _{thermal} / h	H / h	S / h	G / h	G-E _{el} / kcal/mol
8.1	0	-323.274288	0.094205	-323.174845	-323.173901	0.035083	-323.173901	40.98
	298	-323.274288	0.094205	-323.174850	-323.173906	0.035060	-323.208966	40.99
8.2	0	-323.273771	0.094018	-323.174568	-323.173624	0.034906	-323.208530	40.94
	298	-323.273771	0.094018	-323.174573	-323.173629	0.034883	-323.208512	40.95

The overall error introduced (see Table 6) in neglecting the free energy contribution to the tautomer energy difference is 0.04 kcal/mol for both 0 and 298 K, i.e. about one third of the difference between the quadruple-zeta CCSD(T) energy delta and the one from the complete basis set extrapolation with the 3/4 scheme or the 0.56 kcal/mol between the two CBS schemes. This delta will also be significantly lower as the errors introduced by conformer sampling for molecules relevant in a drug discovery context, the errors due to continuum solvation models, and the errors from the experimental determination of tautomer equilibria.

Application to experimental data

Primary scope of this publication is to identify a reliable and cost-effective computational approach for the calculation of tautomer equilibria. Nevertheless, calculations not reflecting experiment are not of practical interest.

The SAMPL2 challenge performed in 2010 contained a tautomer prediction task, that was recently taken up again by Wieder et al.¹³ Though the ultimate goal in their work was to combine allchemical and machine learning potentials, they reported also on density functional theory calculations with B3LYP/aug-cc-pVTZ/B3LYP/6-31G(d)/SMD. They chose to select a subset of 14 challenging equilibria from the original paper (shown in Scheme 2) to compare with the results of the four best-performing submissions by Klamt, Ribeiro, Kast and Soteras, all applying quantum mechanics and continuum solvation models. The original summary publication, additionally to experimental data collected from literature, provided estimates for the experimental errors, which are 0.2 to 0.4 kcal/mol for all pairs except **12D_12C** (naming from original publication) and **7A_7B** with 0.7 and 1.5 kcal/mol.

In the following, I will compare the best approaches identified earlier, namely M06-2X/def2-QZVPP-(SMD)//PBE0/def2-TZVP and M06-2X/def2-TZVP-(SMD)//PBE0/def2-TZVP (both with D3 dispersion correction), r2SCAN-3c and DLPNO-CCSD(T)/def2-QZVPP-(SMD)//PBE0/def2-TZVP with the results from literature. The data is presented in Table 7. Three experimental values for pairs **4A_4B**, **6A_6B** and **7A_7B** that are wrongly reported in the Wieder paper¹³ were corrected to the original values.⁵⁹

Table 7

Performance of four methods from this benchmark and 5 published approaches on a subset of 14 tautomer pairs from the SAMPL2 challenge in 2010. Experimental data as from the summary publication with corrected values for three pairs. M06-Q and M06-T are calculations with the M06-2X density functional and quadruple- respectively triple-zeta basis sets. Regarding the other methods please refer to text.

pair	set	dE kcal/mol	M06- Q	M06- T	r2SCAN- 3c	CCSD(T)	Wieder [13]	Klamt [60]	Ribeiro [61]	Kast [62]	Soteras [63]
1A_1B	1	-4.8	-3.40	-3.60	-5.76	-3.21	-4.7	-4	-3	-7.7	-4.6
2A_2B	1	-6.1	-6.38	-6.55	-7.96	-5.92	-6.8	-5.7	-5.7	-9.7	-6.3
3A_3B	1	-7.2	-7.63	-7.93	-9.34	-7.27	-8.4	-7.7	-6.7	-11.2	-7.7
4A_4B	1	-2.3	0.82	0.63	-2.29	0.99	-0.4	0.5	0.8	-4.6	0.6
5A_5B	1	-4.8	-3.56	-3.77	-5.54	-3.28	-4.7	-3.9	-4.4	-6.2	-5.6
6A_6B	1	-9.2	-10.81	-11.16	-12.28	-10.34	-11.4	-7.6	-9.7	-11.2	-10
7A_7B	2	7	5.16	5.74	6.36	6.23	4.9	5.3	6.5	5.1	5.5
10B_10C	2	-2.9	-1.14	-1.22	-3.85	-0.83	-5.3	1.7	0	-2.8	2.2
10D_10C	2	-1.2	-1.14	-1.22	-3.85	-0.83	-1.7	3.8	2.6	-0.6	5
12D_12C	2	-1.8	-0.10	-0.12	-2.71	0.08	-2.1	3.3	3.1	-0.8	3
14D_14C	2	0.3	1.20	1.00	-0.10	2.65	-1.6	1.9	0.8	0.2	4
15A_15B	2	0.9	2.88	3.24	5.56	4.07	6.1	-3	3.6	0	0.9
15A_15C	2	-1.2	2.87	3.31	3.82	5.65	0.7	-0.6	2.3	-1.9	1.4
15B_15C	2	-2.2	-0.02	0.07	-1.74	1.59	-5	1.8	-1.2	-1.9	0.5
RMSE set1			1.6	1.6	1.8	1.7	1.3	1.4	1.5	2.8	1.3
RMSE set2			2.1	2.2	2.7	3.3	2.6	3.7	2.9	0.9	3.8
RMSE all			1.9	2.0	2.3	2.7	2.1	2.9	2.4	2.0	3.0

The providers of the SAMPL2 challenge had divided the dataset into three subsets. Pairs **1** to **6** belong to the so-called obscure subset and **7** to **15** to the explanatory subset (no pairs from the investigatory subset were used here). The QM approaches of the SAMPL2 submission of Klamt et al. are MP2/QZVPP//BP86/TZVP with COSMO-BP86/TZVP solvation energies and thermal corrections, for Ribeiro et al. M06-2X/MG3S//M06-2X/MG3S with SM8AD solvation contributions by M06-2X/6-31G(d). Kast et al. did B3LYP/6-311++G(d,p) gas-phase optimizations with EC-RISM-MP2/aug-cc-pVDZ energies. Soteras et al. finally used the IEF-MST solvation model on gas-phase energies by MP2 basis set extrapolation and correlation from the CCSD-MP2/6-31+G(d) energy difference.

Looking at the root mean square errors in Table 7 we see that all except one approach yield better results for the obscure than the explanatory subset with RMSE of 1.3 to 1.8 kcal/mol for the former and 2.1 to 3.8 kcal/mol for the latter. The one exception is the submission by Kast et al. with RMSE of 2.8 and 0.9 kcal/mol for the two sets. The second observation that is unexpected is that the most elaborate methods, i.e. CCSD(T), the Klamt MP2 calculations with quadruple-zeta basis set and the Soteras CCSD correlation energy corrections for the explanatory dataset deviate significantly more from experiment than the cheaper methods.

Contrary to the polarizable continuum solvation models used in the other approaches, Kast et al. apply the methodologically different EC-RISM approach. Different solvation treatment could therefore be the root cause of the differences seen. Since the

respective data are not available in the publications, the analysis here is restricted to the methods used in this paper (cf. Supporting Information Tables T1 and T2 for water and gas-phase values and energy deltas to experiment). Mean absolute errors for M06-2X with def2-QZVPP, def2-TZVP and r2SCAN-3c in gas phase are 0.8, 0.6 and 2.3 kcal/mol and comparable to the ones in water solvation (0.8; 0.8; 2.3). As expected, r2SCAN-3c is not in the same ball-park as the other functionals. Nevertheless, all correlation coefficients between methods are extremely high as seen in Fig. 5a, with the lowest one between M06-2X/def2-QZVPP and r2SCAN-3c still being 0.944. Molecules **14** and **15** are found to be very challenging and contribute to the overall error with up to 3.11 kcal/mol for the pair **15A_15C**.

Not having the data at hand, we can only speculate that the correlations of the gas-phase energy differences of all other QM methods will be similarly high. Looking at the solvation phase correlations in Fig. 5b, it becomes even more obvious that the solvation models used throughout have a strong impact on the correlations observed. The lowest correlation coefficient between the methods from this work is 0.960 (always the same SMD solvation model), followed by the data from Ribeiro using SM8AD, another derivative of the SMx solvation model family. Again less correlated is the Sotera approach with IEF-MST solvation and the much weaker correlated COSMO-RS with correlation coefficients between 0.644 and 0.794. Remarkably, the data from the Kast group show higher correlations to the other methods than the COSMO-RS data, though EC-RISM is methodologically a completely different approach.

If we now take a closer look on specific pairs as listed in Table 7, we find one strong outlier in set 1 with six-membered ring tautomers and multiple tautomer pairs with differences between the methods in set 2 for five-membered ring tautomers. Otherwise the only obvious pattern is that Kast et al. consistently provide more negative energy deltas than the others.

Set one consists only of examples for derivatives of pyridine/pyridone type tautomerism. The one outlier is **4A_4B** which predominantly exists in the lactam as does compound **1A_1B** (= compound **8** from dataset 1). A review of the primary experimental publication reveals that the predominant forms in water, CCl₄ and ethanol were determined and confirmed by comparison with the ultraviolet spectra of methylated derivatives with frozen lactim or lactam structures.

Compounds **1** to **4** from the SAMPL2 set are pyridine/pyridone isomers, and thus shed some insight into the problem. The CCSD(T) gas-phase energy differences are 2.1 kcal/mol for **1A_1B**, -2.14 for **2A_2B**, -3.92 for **3A_3B** and 8.94 for **4A_4B**. The positions of the carbonyl group and the nitrogen thus have a strong influence on the relative gas-phase energies, preferring the lactim form in case of pure pyridone **1** and compound **4** without heteroatom in alpha-position to the phenyl ring. Experimentally, all four examples predominantly exist in lactam form in water. The respective energy differences with CCSD(T) in SMD water are -3.21 kcal/mol for **1A_1B**, -5.92 for **2A_2B**, -7.27 for **3A_3B** and 0.99 for **4A_4B**, resulting in the incorrect lactim predicted for example 4. The deltas $\Delta E(\text{gas-solv})$ for the four examples of -5.11, -3.78, -3.35, -7.94 kcal/mol show that the SMD continuum model accounts for the differences in charge distribution due to regioisomerism, but not to the extent needed to reflect experiment. There are two submissions that were able to identify the correct lactam form of example 4. Wieder et al. report a slightly negative value of only -0.4 kcal/mol and Kast et al. one of -4.6 kcal/mol. The latter submission, in Table 5 of the original publication, provides ΔE_{solv} of 5.93, 9.00, 6.57 and 17.56 kcal/mol for the four examples. The significantly higher solvation free energy value for example **4** thus is responsible for the correct assignment of the lactam form.

The deviations for the second set are not consistent between the methods and without detailed data available, any root cause analysis would be speculative. Nevertheless, the high correlation ($r^2 = 0.85$, $y = 1.15x + 4.598$) between gas-phase and water energy deltas for the CCSD(T) calculations combined with the observations for the pyridone derivatives provide strong evidence that the major error source is the energy contributions from the continuum solvation model. Polarizable continuum models are known to struggle in case of strongly localized polar functional groups that lead to tightly associated and ordered solvent molecules. There is active research but no generally applicable solution on the so-called microsolvation approach but testing those concepts is beyond the scope of this work.

Conclusions

In this work, results on relative tautomer energies are provided by thoroughly benchmarking various density functionals of different sophistication, two Hartree-Fock approaches, second-order Moeller-Plesset theory, three semiempirical methods and one current force-field against the accepted gold-standard DLPNO-CCSD(T) coupled-cluster theory. The dataset comprises of tautomeric moieties found in the cores and substituents of common drug molecules.

Since it is not possible to do a comprehensive benchmark on the plethora of other functionals and quantum chemical and semiempirical methods, a subset was picked to reflect the band widths of common choices in literature and methodological complexity. Similarly, this work is restricted to water as the most important solvent and to the application of the method or software default solvation models. Finally, to avoid the additional dependences on conformation and charge states, datasets of rigid neutral chemical moieties nevertheless relevant for drug-like molecules was selected.

In the group of standard density functionals, M06-2X clearly outperforms the other functionals including the double-hybrid B2PLYP and MP2. In the group of compute-cost optimized methods r2SCAN-3c outperforms the other approaches and can compete with the other standard functionals like PBE0 or the widely used B3LYP. The performance of the methods is comparable between gas-phase and continuum solvation in water.

Applying the superior density functional on a challenging subset of compounds with experimental relative tautomer energies, we find that the deviations to experiment are significant in some cases, with up to 7 kcal/mol. Similar deviations are obtained by the participants of the original SAMPL2 challenge in 2010. Since the molecules are about the same size and do not show conformational flexibility, the deviations must be rooted in the calculation if one assumes the experimental values to be correct.

The results from this publication and from literature are highly correlated. All approaches are based on quantum chemical methods accounting for electron correlation to some extent, mostly with appropriate basis sets and PCM solvation models, except for one case using EC-RISM. The gas-phase calculations are even higher correlated.

The four lactam/lactim tautomer pairs in the dataset can shed light into the root cause of the large deviations for some example molecules. Chemically highly similar, the four molecules predominantly exist as lactam form in water, whereas 1 and 4 are predominantly lactims in gas-phase, based on CCSD(T) calculations. The deviation therefore has to come from the solvation free energy, which indeed is very comparable for the four examples, applying the SMD solvation model. Contrary, EC-RISM provides a solvation free energy contribution for example 4 that is significantly higher and reverses the predominant form.

The results presented here, and counting on the often profen and generally accepted gold-standard coupled-cluster, provide strong evidence that the biggest obstacle for reliable calculation of tautomer equilibria is the solvation free energy contribution.

Declarations

Data Availability

The datasets supporting the conclusions of this article are included within the article and its additional files.

References

1. Sayle, R. A. So You Think You Understand Tautomerism? *J Comput Aided Mol Des* **2010**, *24*, 485–496. <https://doi.org/0.1007/s10822-010-9329-5>.
2. Taylor, P.J.; van der Zwan, G.; Antonov, L. Tautomerism: Introduction, History, and Recent Developments in Experimental and Theoretical Methods, in *Tautomersim*, ed. Antonov, L., Wiley VCH, Weinheim, Germany, 2014.

3. Martin, Y. Tautomerism, Hammett σ , and QSAR, *Journal of Computer-Aided Molecular Design*, Springer Netherlands, **2010**, *24*, 613-616.
4. Martin, Y. Let's Not Forget Tautomers. *J Comput Aided Mol Des* **2009**, *23* (10), 693–704. <https://doi.org/10.1007/s10822-009-9303-2>.
5. Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. Tautomerism in Computer Aided Drug Design. *Journal of Receptors and Signal Transduction* **2003**, *23* (4), 361–371. <https://doi.org/10.1081/RRS-120026975>.
6. Clark, T. Tautomers and Reference 3D-Structures: The Orphans of in Silico Drug Design. *J Comput Aided Mol Des* **2010**, *24* (6), 605–611. <https://doi.org/10.1007/s10822-010-9342-8>.
7. Cavasin, A. T.; Hillisch, A.; Uellendahl, F.; Schneckener, S.; Göller, A. H. Reliable and Performant Identification of Low-Energy Conformers in the Gas Phase and Water. *J Chem Inf Model* **2018**, *58* (5), 1005–1020. <https://doi.org/10.1021/acs.jcim.8b00151>.
8. Seep, L.; Bonin, A.; Meier, K.; Diedam, H.; Göller, A. H. Ensemble Completeness in Conformer Sampling: The Case of Small Macrocycles. *J Cheminform* **2021**, *13*, 55. <https://doi.org/10.1186/s13321-021-00524-0>.
9. Fraczekiewicz, R.; Lobell, M.; Göller, A. H.; Krenz, U.; Schoenneis, R.; Clark, R. D.; Hillisch, A. Best of Both Worlds: Combining Pharma Data and State of the Art Modeling Technology to Improve in Silico pKa Prediction. *J Chem Inf Model* **2015**, *55* (2), 389–397. <https://doi.org/10.1021/ci500585w>.
10. Sitzmann, M.; Ihlenfeldt, W.-D.; Nicklaus, M. C. Tautomerism in Large Databases. *J Comp-Aided Mol. Des.* **2010**, *24*, 521–551. <https://doi.org/10.1007/s10822-010-9346-4>.
11. Martin, Y. Let's Not Forget Tautomers. *J Comput Aided Mol Des* **2009**, *23* (10), 693–704. <https://doi.org/10.1007/s10822-009-9303-2>.
12. Tautomersim, ed. Antonov, L., Wiley VCH, Weinheim, Germany, 2014.
13. Wieder, M.; Fass, J.; Chodera, J. D. Fitting Quantum Machine Learning Potentials to Experimental Free Energy Data: Predicting Tautomer Ratios in Solution. *Chem. Sci.* **2021**, *12* (34), 11364–11381. <https://doi.org/10.1039/D1SC01185E>.
14. Warr, W. A. Tautomerism in Chemical Information Management Systems. *J Comput Aided Mol Des* **2010**, *24*, 497–520. <https://doi.org/10.1007/s10822-010-9338-4>.
15. Sayle, R.; Delany, J. J. Canonicalization and Enumeration of Tautomers. 1999.
16. RDKit: Open-source cheminformatics. <http://www.rdkit.org>.
17. Kochev, N. T.; Paskaleva, V. H.; Jeliaskova, N. Ambit-Tautomer: An Open Source Tool for Tautomer Generation. *Molecular Informatics* **2013**, *32* (5–6), 481–504. <https://doi.org/10.1002/minf.201200133>.
18. Guasch, L.; Peach, M. L.; Nicklaus, M. C. Tautomerism of Warfarin: Combined Chemoinformatics, Quantum Chemical, and NMR Investigation. *The Journal of Organic Chemistry* **2015**, *80* (20), 9900–9909. <https://doi.org/10.1021/acs.joc.5b01370>.
19. Oziminski, W.P., Wiśniewski, I. Quantum-chemical study on the relative stability of sildenafil tautomers. *Struct Chem* **32**, 1733–1743 (2021). <https://doi.org/10.1007/s11224-021-01818-7>
20. Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals-Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2011**, *7* (2), 291–309. <https://doi.org/10.1021/ct100466k>.
21. Geballe, M. T.; Skillman, A. G.; Nicholls, A.; PeterGuthrie, J.; Taylor, P. J. The SAMPL2 Blind Predictionchallenge: Introduction and Overview. *J. Comput.-Aided Mol.Des.*, **2010**, *24*, 259–279. <https://doi.org/10.1007/s10822-010-9350-8>.
22. Fogarasi, G. Studies on Tautomerism: Benchmark Quantum Chemical Calculations on Formamide and Formamidine. *J Mol Struct* **2010**, *978*, 257–262. <https://doi.org/10.1016/j.molstruc.2010.02.065>.
23. Milletti, F.; Vulpetti, A. Tautomer Preference in PDB Complexes and Its Impact on Structure-Based Drug Discovery. *J Chem Inf Model* **2010**, *50* (6), 1062–1074. <https://doi.org/10.1021/ci900501c>.

24. Milletti, F.; Storchi, L.; Sforza, G.; Cross, S.; Cruciani, G. Tautomer Enumeration and Stability Prediction for Virtual Screening on Large Chemical Databases. *J Chem Inf Model* **2009**, *49* (1), 68–75. <https://doi.org/10.1021/ci800340j>.
25. Geballe, M. T.; Skillman, A. G.; Nicholls, A.; PeterGuthrie, J.; Taylor, P. J. The SAMPL2 Blind Prediction Challenge: Introduction and Overview. *J. Comput.-Aided Mol.Des.*, 2010, *24*, 259–279. <https://doi.org/10.1007/s10822-010-9350-8>.
26. Pipeline Pilot, client version 21.2.0.2574, server version 21.2.0.2575, Dassault Systemes Biovia Corp, 2020.
27. Settings: Enumerate all Tautomers; maximally 1000; defaults.
28. (a) J. Sadowski, J. Gasteiger, G. Klebe, *J. Chem. Inf. Comput. Sci.* 1994, *34*, 1000-1008. (b) CORINA, Molecular Networks GmbH, Erlangen, Germany (<http://www.molecular-networks.com>).
29. Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB: An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **2019**, *15* (3), 1652–1671. <https://doi.org/10.1021/acs.jctc.8b01176>.
30. Riplinger, C.; Neese, F. An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *The Journal of Chemical Physics* **2013**, *138* (3), 34106. <https://doi.org/10.1063/1.4773581>.
31. P. PIECUCH, S.A. KUCHARSKI, K. KOWALSKI, AND M. MUSIAL, *COMP. PHYS. COMMUN.* 149, 71-96 (2002).
32. Neese, F. The ORCA Program System. *WIREs Computational Molecular Science*. 2012, pp 73–78. <https://doi.org/https://doi.org/10.1002/wcms.81>.
33. Dewar, M. J. S.; Zois, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *Journal of the American Chemical Society* **1985**, *107* (13), 3902–3909. <https://doi.org/10.1021/ja00299a024>.
34. Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *Journal of Computational Chemistry* **1989**, *10* (2), 209–220. <https://doi.org/https://doi.org/10.1002/jcc.540100208>.
35. Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *Journal of Chemical Theory and Computation* **2021**, *17* (7), 4291–4300. <https://doi.org/10.1021/acs.jctc.1c00302>.
36. Maestro Version 13.0.135, MMshare Version 5.6.135, Platform Linux-x86_64, Small-Molecule Drug Discovery Suite 2021–4, Schrodinger, LLC, New York, NY; 2021.
37. Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew Burke Ernzerhof Exchange-Correlation Functional. *The Journal of Chemical Physics* 1999, *110* (11), 5029–5036. <https://doi.org/10.1063/1.478401>.
38. Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *The Journal of Chemical Physics* 1999, *110* (13), 6158–6170. <https://doi.org/10.1063/1.478522>.
39. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *The Journal of Chemical Physics* **2010**, *132* (15), 154104. <https://doi.org/10.1063/1.3382344>
40. Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *The Journal of Chemical Physics* **2005**, *123* (15), 154101. <https://doi.org/10.1063/1.2065267>.
41. Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305. <https://doi.org/10.1039/B508541A>.
42. Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8* (9), 1057–1065. <https://doi.org/10.1039/B515623H>.
43. Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8* (9), 1057–1065. <https://doi.org/10.1039/B515623H>.
44. Hellweg, A.; Hättig, C.; Höfener, S.; Klopper, W. Optimized Accurate Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations for the Atoms Rb to Rn. *Theoretical Chemistry Accounts volume* **2007**, *117*, 587–597. <https://doi.org/10.1007/s00214-007-0250-5>.

45. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*(18), 3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865>.
46. Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91* (14), 146401. <https://doi.org/10.1103/PhysRevLett.91.146401>.
47. Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation. *The Journal of Physical Chemistry Letters* **2020**, *11* (19), 8208–8215. <https://doi.org/10.1021/acs.jpcllett.0c02405>.
48. Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615–6620. <https://doi.org/10.1039/B810189B>.
49. Zhao, Y.; D. G. Truhlar. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.*, **2008**, *120*, 215–241. <https://doi.org/10.1007/s00214-007-0310-x>.
50. Becke, A. D. A New Mixing of Hartree-Fock and Local Density-functional Theories. *The Journal of Chemical Physics* **1993**, *98* (2), 1372–1377. <https://doi.org/10.1063/1.464304>.
51. Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *The Journal of Chemical Physics* **2006**, *124* (3), 34108. <https://doi.org/10.1063/1.2148954>.
52. Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J Comput Chem* **2013**, *34* (19), 1672–1685. <https://doi.org/10.1002/jcc.23317>.
53. Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *The Journal of Chemical Physics* **2015**, *143* (5), -. <https://doi.org/http://dx.doi.org/10.1063/1.4927476>.
54. Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A Revised Low-Cost Variant of the B97-D Density Functional Method. *The Journal of Chemical Physics* **2018**, *148* (6), 64104. <https://doi.org/10.1063/1.5012601>.
55. Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A “Swiss Army Knife” Composite Electronic-Structure Method. *The Journal of Chemical Physics* **2021**, *154* (6), 64103. <https://doi.org/10.1063/5.0040021>.
56. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B* **2009**, *113* (18), 6378–6396. <https://doi.org/10.1021/jp810292n>.
57. Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, No. 5, 799–805. <https://doi.org/10.1039/P29930000799>.
58. Zhong, S.; Barnes, E. C.; Petersson, G. A. Uniformly Convergent N-Tuple-Zeta-Augmented Polarized (nZaP) Basis Sets for Complete Basis Set Extrapolations. I. Self-Consistent Field Energies. *The Journal of Chemical Physics* **2008**, *129*(18), 184116. <https://doi.org/10.1063/1.3009651>.
59. Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated Ab Initio Methods? *Journal of Chemical Theory and Computation* **2011**, *7* (1), 33–43. <https://doi.org/10.1021/ct100396y>.
60. Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *The Journal of Chemical Physics* **1997**, *106* (23), 9639–9646. <https://doi.org/10.1063/1.473863>.
61. Peterson, K. A.; Dunning, T. H. Accurate Correlation Consistent Basis Sets for Molecular Core-Valence Correlation Effects: The Second Row Atoms Al–Ar, and the First Row Atoms B–Ne Revisited. *The Journal of Chemical Physics* **2002**, *117* (23), 10548–10560. <https://doi.org/10.1063/1.1520138>.
62. Jensen, F. Introduction to Computational Chemistry, Wiley VCH, Weinheim, 1999, p. 29.

63. Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys Chem Chem Phys* **2011**, *13*, 6670–6688. <https://doi.org/10.1039/c0cp02984j>.
64. Klamt, A.; Diederhufen, M. Some Conclusions Regarding the Predictions of Tautomericequilibria in Solution Based on the SAMPL2 Challenge. *J. Comput. Aided Mol. Des.* **2010**, *24*, 621–625.
65. Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Prediction of SAMPL2 Aqueous Solvation Free Energies and Tautomeric Ratios Using the SM8, SM8AD, and SMD Solvation Models(4),. , *J. Comput.-Aided Mol. Des.* **2010**, *24*, 317–333.
66. Kast, S.; Heil, J.; GÃ. Prediction of tautomer ratios by embedded-cluster integral equation theory. *J Comput Aided Mol Des* **2010**, *24* (4), 343–353. <https://doi.org/10.1007/s10822-010-9340-x>.
67. Soteras, I.; Orozco, M.; Luque, F. J. Performance of the IEF-MST Solvation Continuum Modelin the SAMPL2 Blind Test Prediction of Hydrationand Tautomerization Free Energies. *J Comput Aided Mol Des* 2010, *24*, 281–291. <https://doi.org/10.1007/s10822-010-9331-y>.
68. Evans, D. A.; Smith, G. F.; Wahid, M. A. The Tautomerism of 3-Hydroxyisoquinolines. *J. Chem. Soc. B* **1967**, No. 0, 590–595. <https://doi.org/10.1039/J29670000590>.
69. Sure, R.; el Mahdali, M.; Plajer, A.; Deglmann, P. Towards a Converged Strategy for Including Microsolvation in Reaction Mechanism Calculations. *J. Comp-Aided Mol. Des.* **2021**, No. 35, pages473–492. <https://doi.org/10.1007/s10822-020-00366-2>.

Scheme

Scheme 1 and 2 are available in supplementary section.

Figures

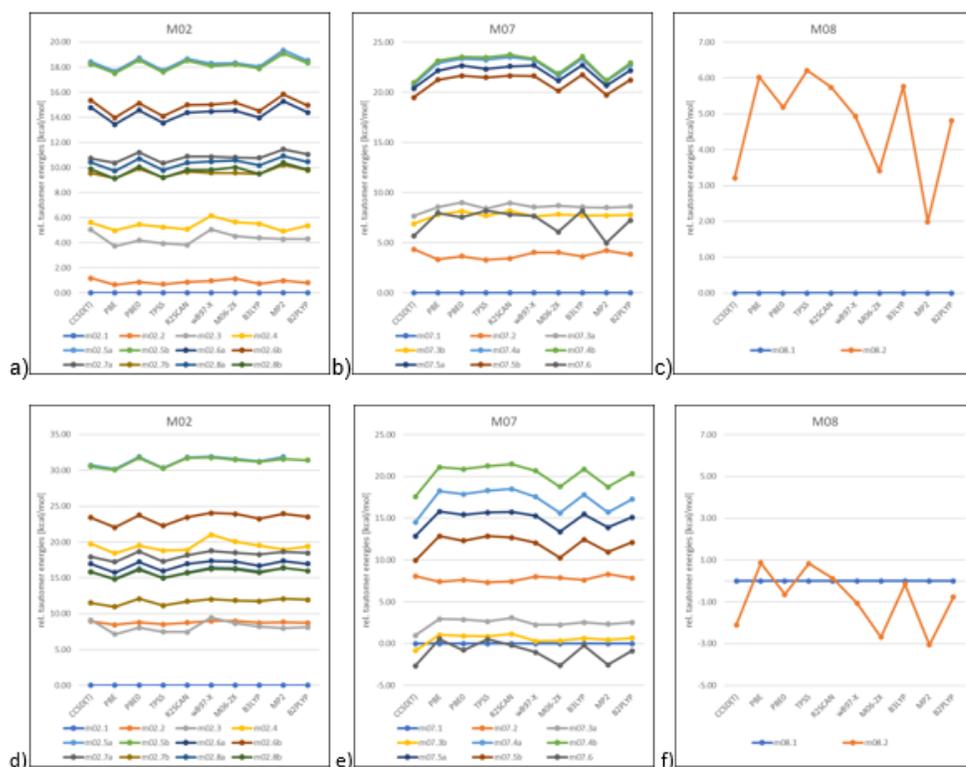


Figure 1

Comparison of tautomer state energies calculated with CCSD(T), MP2 and various DFT methods and def2-QZVPP basis set for three example molecules **2**, **7** and **8**. Plots a) to c) show results with SMD continuum solvation model and plots d) to f) results in gas-phase.



Figure 2

Tautomer state energy differences in kcal/mol relative to CCSD(T) energy differences for each state for each density functional method and MP2 for calculations with SMD water solvation.

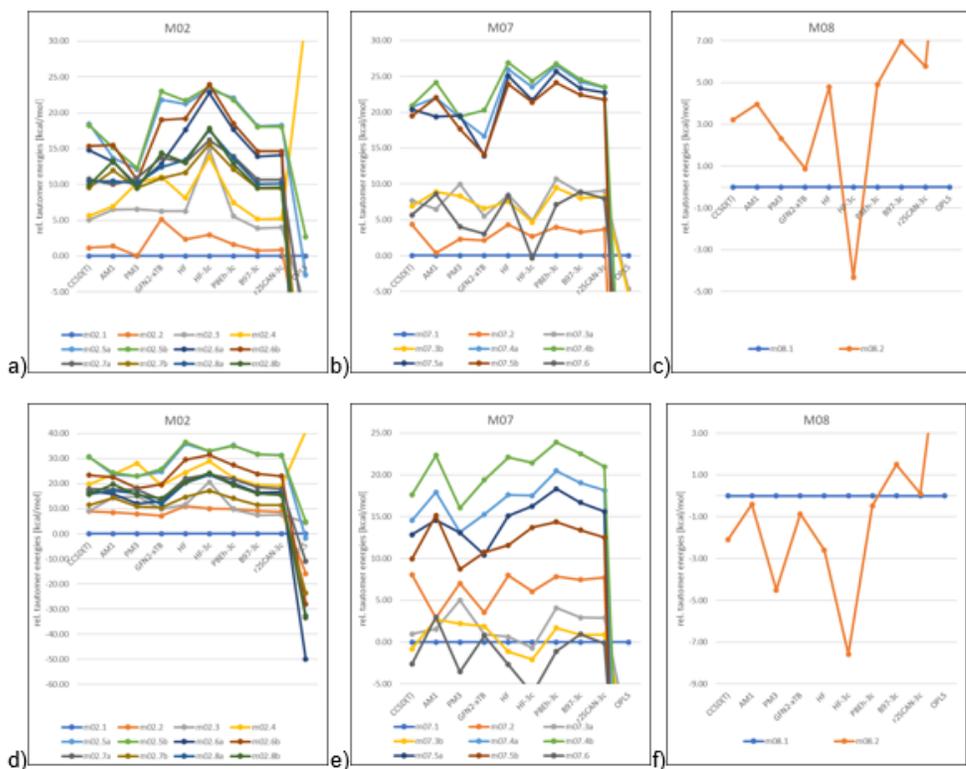


Figure 3

Comparison of tautomer state energies in kcal/mol calculated with CCSD(T), HF, semiempirical, cost-optimized DFT and all-atom force-field OPLS4 for three example molecules **2**, **7** and **8**. Plots a) to c) show results in continuum solvation using different solvent models (see Methods for details) and plots d) to f) results in gas-phase.



Figure 4

Tautomer state energy differences in kcal/mol relative to CCSD(T) energy differences for each state for various low-cost quantum-mechanics, semiempirical and DFT methods for calculations in water solvation. Note the different y-axis scales.

	M06-Q	M06-T	r2SCAN-3	CCSD(T)
M06-Q	1.000	0.998	0.944	0.985
M06-T		1.000	0.963	0.992
r2SCAN-3c			1.000	0.973
CCSD(T)				1.000

	M06-Q	M06-T	r2SCAN-3	CCSD(T)	Wieder	Klamt	Ribeiro	Kast	Soteras
M06-Q	1.000	0.999	0.960	0.991	0.915	0.794	0.952	0.924	0.885
M06-T		1.000	0.967	0.992	0.919	0.786	0.952	0.924	0.878
r2SCAN-3c			1.000	0.972	0.948	0.644	0.898	0.883	0.779
CCSD(T)				1.000	0.897	0.756	0.921	0.907	0.858
Wieder					1.000	0.594	0.905	0.835	0.747
Klamt						1.000	0.857	0.894	0.950
Ribeiro							1.000	0.946	0.941
Kast								1.000	0.941
Soteras									1.000

Figure 5

Matrix of the correlations of the four methods of this work and the five published methods by Wieder, Klamt, Ribeiro, Kast and Soteras. M06-Q denotes M06-2X/def2-QZVPP//PBE0/def2-TZVP, M06-T denotes M06-2X/def2-TZVP//PBE0/def2-TZVP, and CCSD(T) denotes DLPNO-CCSD(T)/def2-QZVPP//PBE0/def2-TZVP. The plot is color-coded by the Pearson correlation coefficient. a) provides gas-phase results from this work, b) provides correlations from all approaches in continuum solvation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ReliableBenchmarksSI.21.docx](#)
- [dataset1.sdf](#)
- [dataset2.sdf](#)
- [Scheme01.png](#)
- [Scheme02.png](#)