THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Benchmarking Density Functionals, Basis Sets, and Solvent Models in Predicting Thermodynamic Hydricities of Organic Hydrides

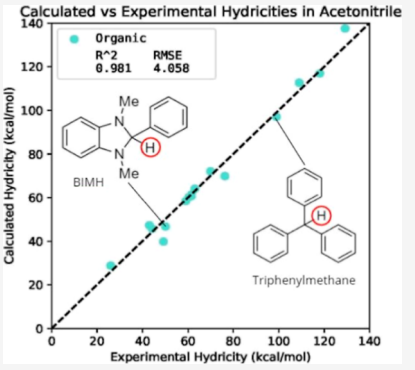
Christina Yeo, Minh Nguyen, and Lee-Ping Wang*

Cite This: *J. Phys. Chem. A* 2022, 126, 7566−7577

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Many renewable energy technologies, such as hydrogen gas synthesis and carbon dioxide reduction, rely on chemical reactions involving hydride anions (H^-). When selecting molecules to be used in such applications, an important quantity to consider is the thermodynamic hydricity, which is the free energy required for a species to donate a hydride anion. Theoretical calculations of thermodynamic hydricity depend on several parameters, mainly the density functional, basis set, and solvent model. In order to assess the effects of the above three parameters, we carry out hydricity calculations with different combinations of density functionals, basis sets, and solvent models for a set of organic molecules with known experimental hydricity values. The data are analyzed by comparing the R^2 and root-mean-squared error (RMSE) of linear fits with a fixed slope of 1 and using the Akaike Information Criterion to determine statistical significance of the RMSE rank ordering. Based on these results, we quantified the accuracy of theoretical predictions of hydricity and found that the best compromise between accuracy and computational cost was obtained by using the B3LYP-D3 density functional for the geometry optimization and free-energy corrections, either ωB97X-D3 or M06-2X-D3 for single-point energy corrections, combined with a basis set no larger than def-TZVP and the C-PCM ISWIG solvation model. At this level of theory, the RMSEs of hydricity calculations for organic molecules in acetonitrile and dimethyl sulfoxide were found to be <4 and <10 kcal/mol, respectively, for an experimental data set with a dynamic range of 20−150 kcal/mol.



1. INTRODUCTION

The thermodynamic hydricity of a molecule measures the free energy of a heterolytic M−H (metallic hydrides) or C−H (organic hydrides) bond dissociation.[1−4] Reactions involving hydride anions play a crucial role in various renewable energy technologies, such as the electrochemical reduction of CO_2 into carbon-based fuels[5−10] or H_2 synthesis.[8,11] The hydricity is therefore a key property in an important class of renewable energy reactions as the free energy of hydride transfer from a donor molecule to an acceptor is given by the difference in their hydricities.

Traditionally, transition-metal hydrides have been the most popular candidates for such applications,[1] but many of these metals are expensive, unsustainable, and toxic.[12] Organic hydrides such as dihydropyridine[5,6] and benzimidazoles[7] are promising metal-free, renewable alternatives to their costly counterparts. Therefore, studying the hydricities of organic compounds in the hopes of selecting more metal-free catalysts for CO_2 reduction would be a valuable effort toward closing the carbon cycle.

However, the hydricity is expensive and laborious to measure experimentally; it involves summing equilibrium constants, acid dissociation constants, and free energies over several thermochemical reactions,[9] not to mention having to synthesize the molecule of interest in the first place. There have been several works that use Kohn−Sham density functional theory (DFT) to theoretically predict hydricities. Reference 12 calculated the hydricities of several metal-free hydrides via two different approaches specified in refs 13 and 14. Reference 15 calculated the hydricities of various p- and o-quinones in dimethyl sulfoxide (DMSO) with geometry optimizations done using B3LYP/6-31+G*, single-point calculations using B3LYP/6-311++G and MP2/6-311++G**, and correction terms calculated using B3LYP/6-31+G*, with solvent model IEFPCM for all steps of the calculation. Reference 16 calculated the hydricities of 6d transition-metal hydrides using B3LYP as the density functional, LACVP** and LACV3P++** as the basis sets for the geometry optimization/frequency analysis and single-point calculations, respectively, and the Poisson−Boltzmann solvent model.

These works all use two selected methods for the entire set of tested molecules, rather than testing several different methods and observing the effects. A benchmark of DFT methods for calculating hydricities did not exist until July

Received: May 4, 2022
Revised: September 25, 2022
Published: October 17, 2022

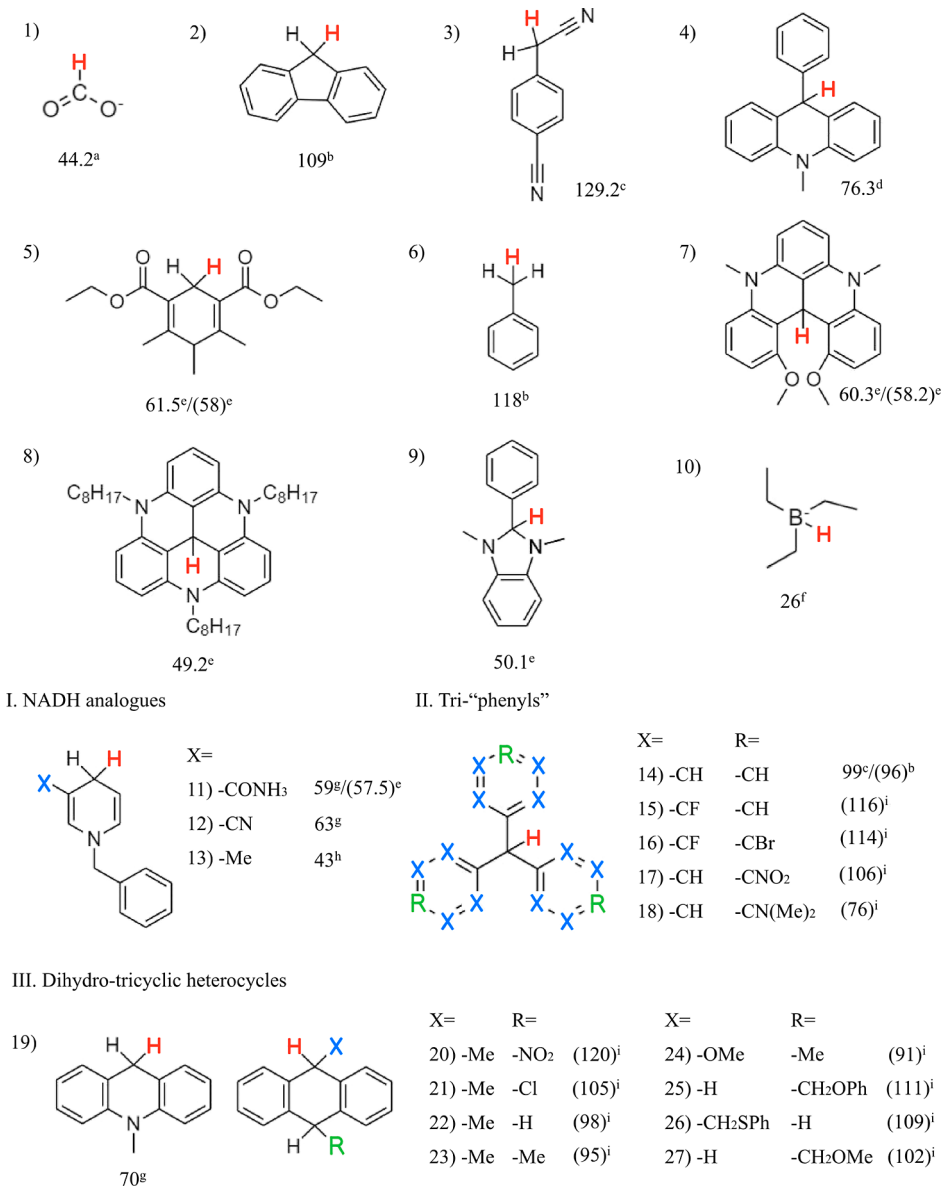


Figure 1. Figure of all molecules tested. Hydridic hydrogens are colored red. The listed numbers are experimentally determined hydricities, with the superscripts indicating reference. Numbers with no parentheses were obtained in acetonitrile, while those with parentheses were obtained in DMSO. a: ref 20, b: ref 21, c: ref 22, d: ref 10, e: ref 14, f: ref 23, g: ref 24, h: ref 25, and i: ref 12.

2021,[17] and this benchmark was exclusively for 3d transition-metal complexes, while we provide a benchmark for organic hydrides.

When using density functional theory to calculate hydricity, the three most important parameters to consider are the density functional, basis set, and solvent model. In this paper, we test different combinations of these three parameters, all stemming from the "base" level of theory, which uses B3LYP-D3 as the functional, TZVP as the basis set, and PCM as the solvent model. We will first summarize the systems we studied and the methods we used to calculate their hydricities. Then we will introduce the statistical measures we utilized to compare performance across different models. From these results, we were able to formulate a set of guidelines for carrying out theoretical calculations of thermodynamic hydricities for organic hydrides. We end by discussing concluding thoughts and future directions.

2. METHODS

Figure 1 lists the donor structures for all the molecules we tested, along with their experimentally measured hydricity. Starting structures were obtained from WebCSD, an online crystal structure database provided by the Cambridge Crystallographic Data Centre and Leibniz Institute for Information Infrastructure.[18] Any structures not available on WebCSD were built by hand on GaussView, a graphical interface used for preparing input files for quantum chemistry calculations.[19] In such cases, we took special care to make the initial structures as close to the expected final structures as possible by manually adjusting bond angles.

For molecules numbered 20−27, two acceptor structures were modeled, where the hydridic hydrogen, that is, the H that is removed after the C−H bond dissociation, was taken from the side of the middle six-membered ring corresponding to substituent X or R, and the structure with the lower optimized energy was chosen for the final calculation. For molecules 1

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

and 10, the donor and acceptor structures have charge −1/0, respectively. For the others, the donor and acceptor structures have charge 0/+1, respectively. The experimental hydricity of all molecules taken from the literature were measured in either acetonitrile or dimethyl sulfoxide (DMSO) solvent.

Table 1 gives the combinations of density functionals, basis sets, and solvent models used for our calculations. For every

Table 1. Combinations of Parameters Used[a]

model ID	single-point energy	optimization and frequency
D01		B3LYP-D3/6-31G*/PCM
D02		B3LYP-D3/TZVP/PCM
D03		ωB97X-D3/TZVP/PCM
D04		BP86-D3/TZVP/PCM
D05		B3LYP-D3/TZVP/Gas
D06		B3LYP-D3/6-31G*/Gas
D07		B3LYP*-D3/TZVP/PCM
D08	B3LYP-D3/TZVP/PCM	B3LYP-D3/6-31G*/PCM
D09	B3LYP-D3/TZVP+/PCM	B3LYP-D3/TZVP/PCM
D10	B3LYP-D3/6-31G*/PCM	B3LYP-D3/6-31G*/Gas
D11	B3LYP-D3/TZVP/PCM	B3LYP-D3/TZVP/Gas
D12	ωB97X-D3/TZVP+/PCM	ωB97X-D3/TZVP/PCM
D13		ωB97X-D3/6-31G*/PCM
D14	B3LYP-D3/TZVP/SMD	B3LYP-D3/TZVP/Gas
D15	B3LYP-D3/TZVP/SMD	B3LYP-D3/TZVP/PCM
D16	ωB97X-D3/TZVP/PCM	B3LYP-D3/TZVP/PCM
D17	M06-2X-D3/6-31G*/PCM	B3LYP-D3/TZVP/PCM
D18	M06-2X-D3/TZVP/PCM	B3LYP-D3/TZVP/PCM
D19	M06-2X-D3/TZVP+/PCM	B3LYP-D3/TZVP/PCM

[a]Each level of theory is specified in the following format: density functional/basis set/solvent model. The "single-point energy" column is filled only if different from the level of theory for optimization and frequency.

calculation, the D3(BJ) empirical dispersion correction[26] was included by default; we discuss the effects of this in Section 3. BP86, which has the lowest computational cost of the density functionals we used, is a generalized gradient approximation (GGA) functional, meaning that it only uses the local electron density and gradient. B3LYP is the most widely used hybrid functional; that is, it mixes the DFT exchange−correlation energy and Hartree-Fock (HF) exchange energy with a fixed ratio.[27−29] B3LYP has 20% HF exchange, while B3LYP* has 15%.[30] ωB97X is a range-separated hybrid functional, meaning that the mixing ratios of the DFT and HF contributions vary depending on the distance between electrons.[27,31] M06-2X is a hybrid meta-GGA with the highest amount of HF exchange (54%) of the functionals we selected. B3LYP has three empirical parameters fitted to experiments, while ωB97X has 17 and M06-2X has 29.[28,29,32,33]

For the basis sets, we selected 6-31G*, TZVP, and TZVP+ (short for ma-def2-TZVP(-f)-LTZ+), which have 14, 19, and 28 basis functions per carbon atom, respectively.[34−37] For any atoms beyond potassium (K), 6-31G* is replaced by the LANL2DZ (LDZ) basis set and an effective core potential (ECP), while TZVP and TZVP+ are replaced by the LANL2TZ (LTZ) basis set and an ECP.[38−40] TZVP+ not only is larger than TZVP but also has diffuse functions added to non-hydrogen atoms.[34] Basis functions with f or higher angular momentum are absent in the two smaller basis sets and removed from TZVP+.

Last, we used two continuum solvent models for our calculations: C-PCM ISWIG (PCM for short) and SMD. C-PCM ISWIG is a conductor-like polarizable continuum solvent model (C-PCM)[41,42] with a "smooth discretization" via the Improved Switching/Gaussian (ISWIG) method. Polarizable continuum models represent the solvent by placing the solute in a cavity with an apparent charge distribution over the surface of that cavity. Boundary-element methods are used to discretize the solute/continuum interface, but this often leads to a discontinuous potential energy surface for the solute, leading to singularities. The ISWIG method is a discretization scheme that overcomes such limitations.[43,44] SMD is another type of polarizable continuum solvent model that includes an additional parameterized term for short-range solvent−solute interactions such as dispersion and solvent structural effects (e.g., hydrogen bonding or exchange repulsion).[45]

For models that have two columns, the first column indicates the level of theory used for the single-point calculations, while the second column indicates the level of theory used for the geometry optimization and frequency analysis calculations. This composite approach allows for using more expensive levels of theory while keeping the computational cost low as geometry optimization and frequency analysis calculations involve taking many gradients, while single-point calculations do not. If a model has a single column, the same level of theory was used for all calculations. Levels of theory using the M06-2X functional or SMD solvent model were calculated using Q-Chem,[46] while others were calculated using TeraChem.[47]

The free energies of the acceptor and donor can easily be estimated on TeraChem and Q-Chem via the harmonic approximation. On the other hand, a hydride anion has complicated interactions with the solvent that a continuum solvent model cannot account for. Even if we were to use an explicit solvent model in an attempt to calculate the free energy of the hydride, the hydride may quickly react with the solvent molecules, making it difficult to obtain a reasonable estimate of its free energy in the solvent. There are several ways to circumvent this problem. Reference 13 calculated the free energy of the hydricity half reaction

$$\Delta G_{\mathrm{HHR}} = G(\mathrm{acceptor}) - G(\mathrm{donor}) \quad (1)$$

and then used a reference reaction to evaluate $G(H^-)$ and construct the thermodynamic hydricity via an isodesmic reaction scheme

$$AH^- + B \rightarrow A + BH^-, \quad (2)$$

where AH^- and BH^- are the donor structures of species A and B, respectively. Meanwhile, ref 17 used a thermochemical cycle given by

$$MH + H^+ \rightleftharpoons M^+ + H_2 \quad (3)$$

$$H_2 \rightleftharpoons H^+ + H^- \quad (4)$$

and modeled the protons as a complex with discrete solvent molecules.

In this paper, we chose to calculate the free energy of the hydricity half reaction and treat the free energy of the solvated hydride as a fitting parameter. More specifically, we compute the ΔG_{HHR} for all molecules using one model, create a linear fit with the slope fixed at 1, and apply an overall vertical shift to all the data points so that the linear fit goes through the origin. This vertical shift corresponds to the free energy of the

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

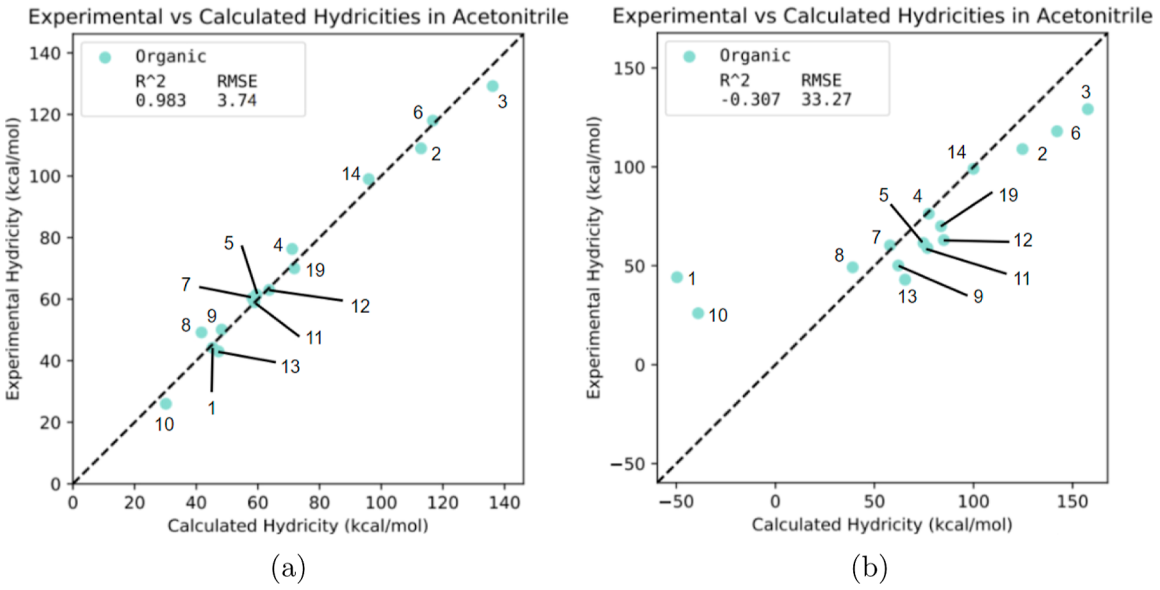


Figure 2. Two example calculated vs experimental hydricity plots for molecules in acetonitrile. Each data point is labeled by the corresponding molecule number given in Figure 1. The calculated values were obtained using (a) D18 (M06-2X-D3/TZVP/PCM//B3LYP-D3/TZVP/PCM), one of our most accurate models, and (b) D06 (B3LYP-D3/6-31G*/Gas), one of our least accurate models.

hydride, and the final linear fit for all our data sets is $y = x$. Then we calculate the R^2, root-mean-squared error (RMSE), and Akaike information criterion (AIC) for each data set to quantify the accuracy of the given model.

For this benchmark, we used the AIC as a tool to verify that each model's performance is statistically significant. For instance, when model A gives a better R^2 and RMSE compared to model B, we want to confirm that this result is due to model A truly being more accurate than model B and not because of random statistical fluctuations. That is, if we sample some noise from a Gaussian distribution and apply it to the data set produced by model B, the new linear fit better not give a better R^2 and RMSE value compared to that of model A. Bootstrapping could be another alternative, but we would have to perform many iterations to get the desired degree of confidence. This is why we chose the AIC, whose difference across different models serves the same purpose as P-values,[48] as our measure of statistical significance.

For our data, we used the second-order AIC for small sample sizes

$$\mathrm{AIC} = -2\ln\mathcal{L} + 2K + \frac{2K(K+1)}{n-K-1} \tag{5}$$

which is applicable for $n/K < 40$, where n is the sample size and K is the number of model parameters. Since our model is a linear fit with data points that do not lie exactly on the line, $K = 3$. $-2\ln\mathcal{L}$ is given by

$$-2\ln\mathcal{L} = \sum_i \left(\frac{(a_i - x_i)^2}{\sigma^2} + 2\ln\sqrt{2\pi}\sigma \right) \tag{6}$$

where a_i are the values that the model predicts, x_i are the actual values, and σ is the uncertainty. Although we can use the reported experimental uncertainties, the uncertainties from different experiments vary across 1 order of magnitude,[14,22,24,25] which would lead to some data points contributing significantly more to the AIC than others. To avoid this, we used an uncertainty of 2 kcal/mol for all data points when calculating the AIC, corresponding to the value reported in refs 14 and 24.

To compare the AIC across the different models, we calculate the exponential of the AIC differences or the relative likelihood

$$e^{(\mathrm{AIC}_{\min} - \mathrm{AIC}_i)/2} \tag{7}$$

which is proportional to the ith model's probability for minimizing information loss. We have disregarded the normalization factor that would give us the AIC weights for convenience. The model with the lowest AIC ($\mathrm{AIC}_{\min}$) has a relative likelihood of exactly 1. Since the n and K of all our models are the same, the $2K + 2K(K+1)/(n-K-1)$ terms cancel when computing the relative likelihood, leaving only the log-likelihood terms in the exponent.[49]

3. RESULTS AND DISCUSSION

Figure 2 shows two example hydricity plots, followed by R^2, RMSE, AIC, and relative likelihood values for molecules in acetonitrile and DMSO listed in Tables 2 and 3. When discussing levels of theory, we use the format "level of theory used for energy calculations//level of theory used for geometry optimization and free energy corrections".

Figure 3 compares the accuracy across different density functionals for a fixed choice of basis set (TZVP) and solvent model (PCM). We can see that ωB97X-D3 and M06-2X-D3 give the lowest RMSE for acetonitrile. Although ωB97X-D3 performs slightly better, the difference in accuracy between the two is minimal, and their relative likelihoods (Figures 9 and 10) confirm that this difference is not statistically significant. It is also apparent that the effect of using B3LYP*-D3 instead B3LYP-D3 is minimal; this is because B3LYP*-D3 was originally developed to more accurately predict low-spin/high-spin energy splitting in molecules with multiple spin configurations, that is, transition-metal complexes.[30] Since we are only working with organic hydrides, changing the percentage of HF exchange is expected to have a smaller impact on the hydricity.

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

Table 2. R^2, RMSE (kcal/mol), AIC, and Relative Likelihood for Molecules in Acetonitrile

	acetonitrile			
model ID	R^2	RMSE (kcal/mol)	AIC	$e^{(AIC_{min}-AIC_i)/2}$
D01	0.967	5.31	1.62×10^2	2.75×10^{-12}
D02	0.968	5.17	1.57×10^2	4.12×10^{-11}
D03	0.981	4.04	1.18×10^2	1.24×10^{-2}
D04	0.965	5.45	1.68×10^2	1.65×10^{-13}
D05	−0.171	31.49	3.77×10^3	0
D06	−0.307	33.27	4.21×10^3	0
D07	0.967	5.27	1.61×10^2	6.14×10^{-12}
D08	0.970	5.05	1.52×10^2	3.94×10^{-10}
D09	0.963	5.59	1.74×10^2	8.05×10^{-15}
D10	0.967	5.32	1.63×10^2	2.10×10^{-12}
D11	0.968	5.18	1.57×10^2	3.59×10^{-11}
D12	0.979	4.25	1.24×10^2	5.06×10^{-4}
D13	0.976	4.50	1.32×10^2	8.33×10^{-6}
D14	0.971	4.99	1.50×10^2	1.27×10^{-9}
D15	0.970	5.01	1.51×10^2	9.54×10^{-10}
D16	0.983	3.81	1.11×10^2	3.84×10^{-1}
D17	0.976	4.54	1.34×10^2	4.25×10^{-6}
D18	0.983	3.74	1.09×10^2	1
D19	0.981	4.04	1.18×10^2	1.33×10^{-2}

Table 3. R^2, RMSE (kcal/mol), AIC, and Relative Likelihood for Molecules in DMSO

	DMSO			
model ID	R^2	RMSE (kcal/mol)	AIC	$e^{(AIC_{min}-AIC_i)/2}$
D01	0.796	9.21	3.99 × 102	2.06×10^{-3}
D02	0.767	9.85	4.48 × 102	5.73×10^{-14}
D03	0.737	10.46	4.97 × 102	1.19×10^{-24}
D04	0.785	9.47	4.18 × 102	1.44×10^{-7}
D05	0.641	12.23	6.58 × 102	1.08×10^{-59}
D06	0.702	11.14	5.56 × 102	2.08×10^{-37}
D07	0.774	9.71	4.37 × 102	1.47×10^{-11}
D08	0.767	9.85	4.48 × 102	5.42×10^{-14}
D09	0.779	9.60	4.28 × 102	9.84×10^{-10}
D10	0.803	9.05	3.87 × 102	6.79×10^{-1}
D11	0.762	9.95	4.56 × 102	9.30×10^{-16}
D12	0.756	10.08	4.66 × 102	7.46×10^{-18}
D13	0.792	9.32	4.07 × 102	4.34×10^{-5}
D14	0.768	9.84	4.46 × 102	1.06×10^{-13}
D15	0.775	9.67	4.34 × 102	6.83×10^{-11}
D16	0.761	9.97	4.57 × 102	5.28×10^{-16}
D17	0.804	9.04	3.87 × 102	1
D18	0.764	9.91	4.52 × 102	6.13×10^{-15}
D19	0.781	9.51	4.25 × 102	5.09×10^{-9}

Molecules in DMSO show poor accuracy overall compared to those in acetonitrile, and the fact that more expensive functionals are performing worse is also alarming. To explain this unexpected result, we first noted that about half of our experimental data in DMSO solvent came from eight substituted dihydro-tricyclic heterocycles (molecules 20−27). If these data points are removed from the data set, the RMSE decreases significantly (e.g., from 9.85 to 6.87 kcal/mol for model D02), but the data set becomes too small to reliably rank the computational methods. We suspect that this is because these molecules possess substantial conformational flexibility, including the puckering of the heterocycle and rotatable bonds of the substituents, making it difficult to find the true global minimum when optimizing the geometry and to approximate the potential energy surface as a harmonic oscillator. We compared three of our calculated hydricities in DMSO that also had computational results reported in ref 14 as a sanity check and saw good agreement with an RMSE of <1 kcal/mol, giving us confidence that the unsatisfactory results for molecules in DMSO are due to the conformational flexibility of the dihydro-tricyclic heterocycles rather than an error in our calculation setup.

Comparing B3LYP-D3/TZVP/PCM, ωB97X-D3/TZVP/PCM, and ωB97X-D3/TZVP/PCM//B3LYP-D3/TZVP/PCM, we can conclude that the improvement from using a more expensive functional is primarily due to the different energies rather than the different structures or thermodynamic corrections. It may be surprising that using the lower level of theory for the geometry optimization/frequency analysis and the higher level of theory for the single point gave better accuracy compared to using the higher level of theory for all steps of the hydricity calculation. This could be attributed to the fact that when using ωB97X-D3 or M06-2X-D3, the published vibrational frequency scale factors, which we did not use in our free-energy corrections, deviate more from 1 compared to B3LYP-D3 for most basis sets.[50,51] ωB97X-D3 and M06-2X-D3 also show larger errors under bond stretching compared to B3LYP-D3 for H_2, N_2, and F_2 dissociation,[52] an observation that is consistent with the greater deviation from 1 of their vibrational frequency scale factors.

In order to observe the effects of the dispersion correction on the accuracy, we removed the dispersion correction from the single-point energies, which corresponds to B3LYP/TZVP/PCM//B3LYP-D3/TZVP/PCM. This had a minimal effect on the calculated hydricity save for one outlier, molecule 26. Upon removing the dispersion correction, all other molecules in DMSO saw a change in free energy of −0.77 ± 0.39 kcal/mol, whereas molecule 26 saw a change of 7.135 kcal/mol. This is due to the CH_2SPh ligand being significantly closer to the tricyclic portion in the acceptor structure relative to the donor. This makes the dispersion correction contribution to the acceptor energy more negative, which means that adding the dispersion decreases the free energy of the hydricity half reaction (eq 1). To make sure that this was not simply a feature of the acceptor structure we chose, where the hydridic hydrogen is attached to the same carbon as the CH_2SPh ligand, we tried calculating the hydricity of the acceptor structure that exactly matches that of ref 12, where the hydridic hydrogen is attached to the same carbon as the "R" ligand (see Figure 1, molecule 26). This structure showed the same behavior, so we kept our original structure as it had a lower energy after optimization. This large change in conformation upon the removal of a hydride anion was only observed for molecule 26, but there is a possibility that it could happen with other molecules not included in our study. This stresses the need for future studies to carry out a greater number of minimizations from different initial starting conformations or to perform conformational sampling using molecular dynamics or other methods.

Looking at Figures 4−6, it is clear that using a bigger basis set does not necessarily give better results. With B3LYP-D3 as the density functional, using 6-31G* for the geometry optimization/frequency analysis and TZVP gave the best accuracy for acetonitrile, while 6-31G* gave the best accuracy for DMSO. With ωB97X-D3 or M06-2X-D3 as the density

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

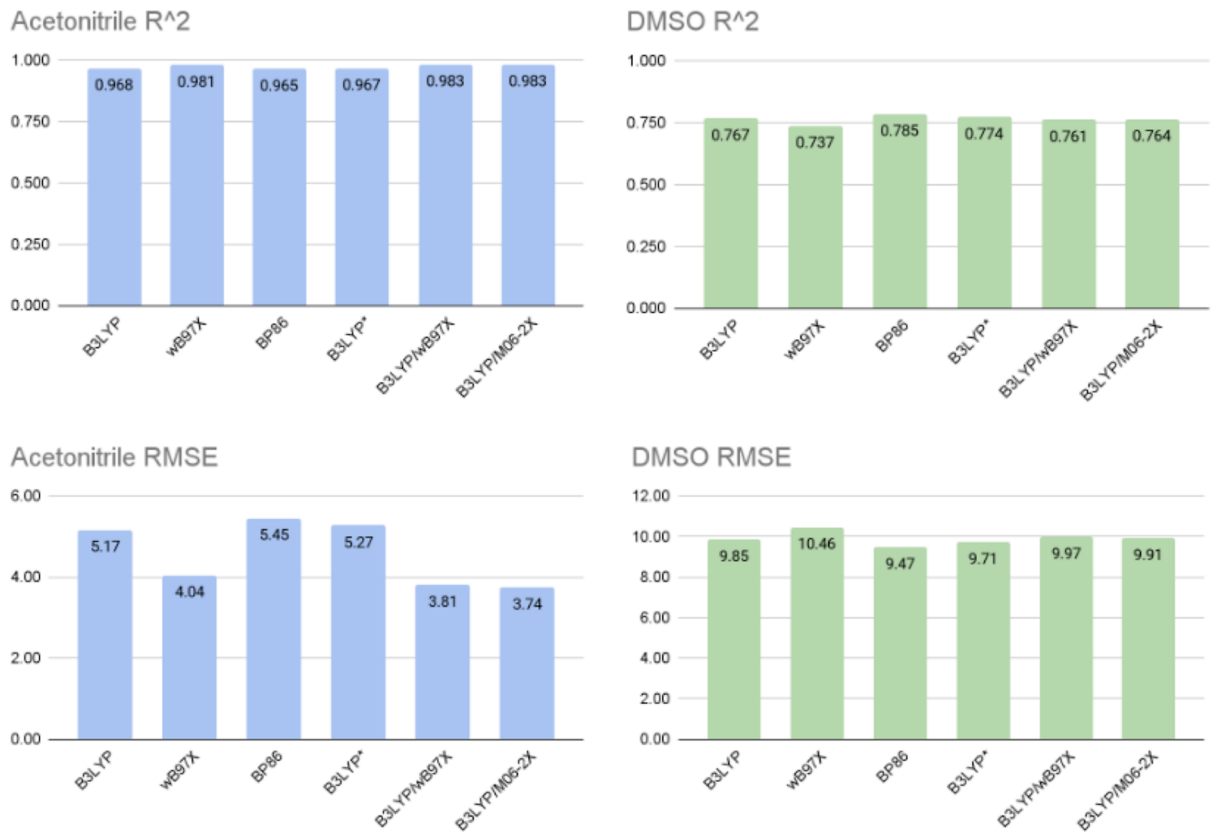


Figure 3. Comparison of R^2 and RMSE values across different density functionals, with TZVP as the basis set and PCM as the solvent model. All functionals in this figure include the dispersion correction, but the suffix -D3 was omitted from the axis labels to save room.

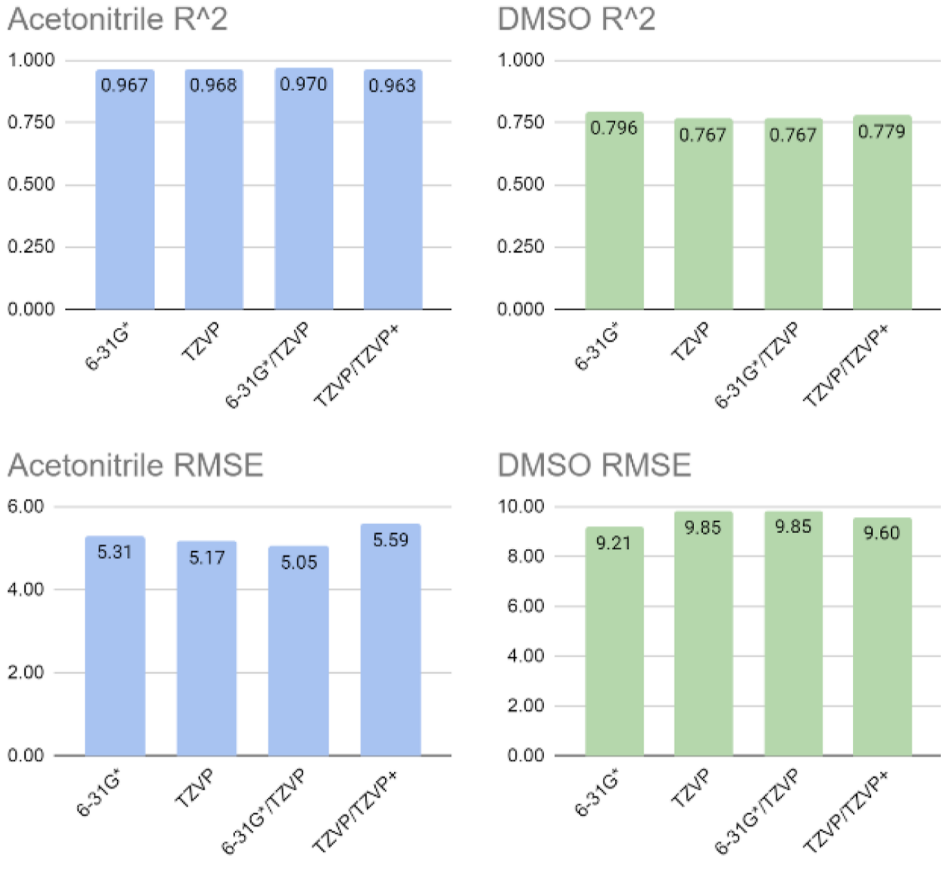


Figure 4. Comparison of R^2 and RMSE values across different basis sets, with B3LYP-D3 as the density functional and PCM as the solvent model.

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

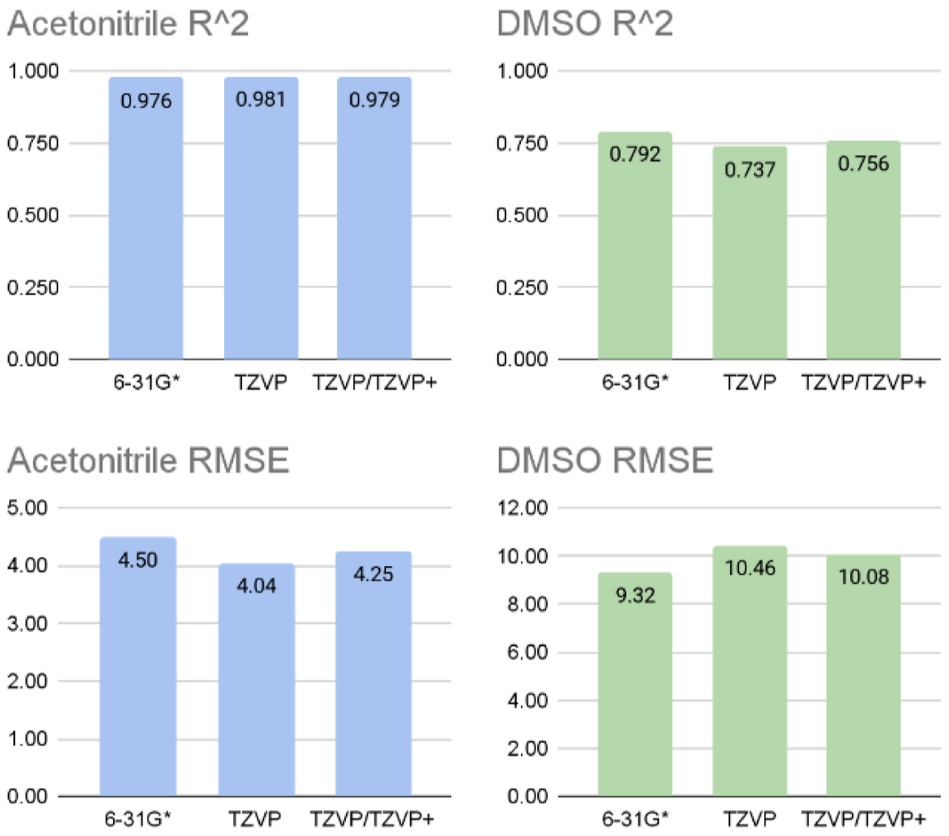


Figure 5. Comparison of R^2 and RMSE values across different basis sets, with ωB97X-D3 as the density functional and PCM as the solvent model.

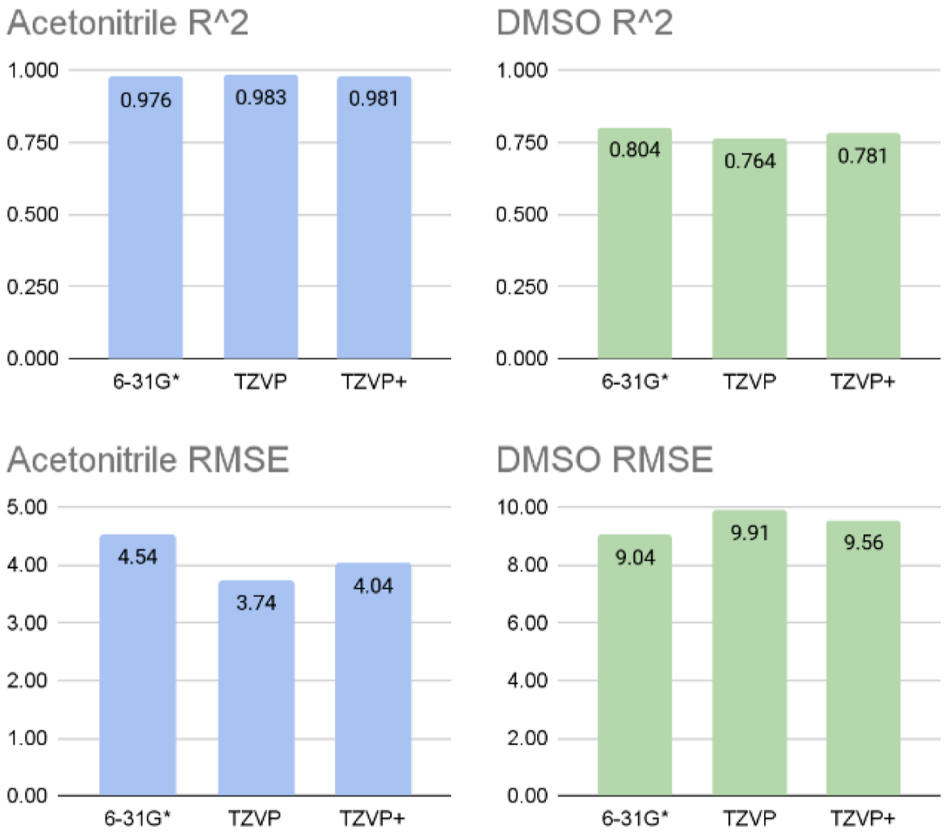


Figure 6. Comparison of R^2 and RMSE values across different basis sets, with geometry optimization and frequency analysis done with B3LYP-D3/TZVP/PCM and single-point calculations done with M06-2X-D3 as the density functional and PCM as the solvent model.

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

Figure 7. Comparison of R^2 and RMSE values across different solvent models, with B3LYP-D3 as the density functional and 6-31G* as the basis set.

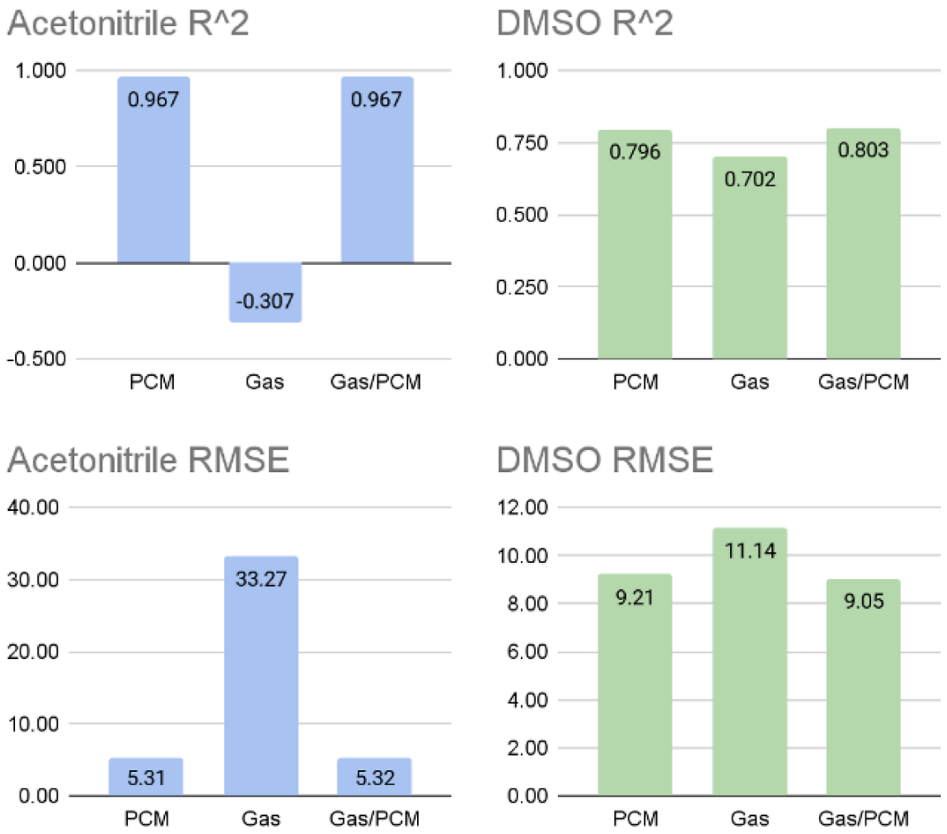


Figure 8. Comparison of R^2 and RMSE values across different solvent models, with B3LYP-D3 as the density functional and TZVP as the basis set.

functional, TZVP performed best for molecules in acetonitrile, while 6-31G*, the smallest basis set, gave the best accuracy. This is not surprising given that our test molecules were all organic hydrides, meaning that we rarely deal with elements heavier than K. For users looking to calculate the hydricity of an organic molecule, we would recommend using a basis set no larger than TZVP, considering the trade-off between computational cost and accuracy.

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

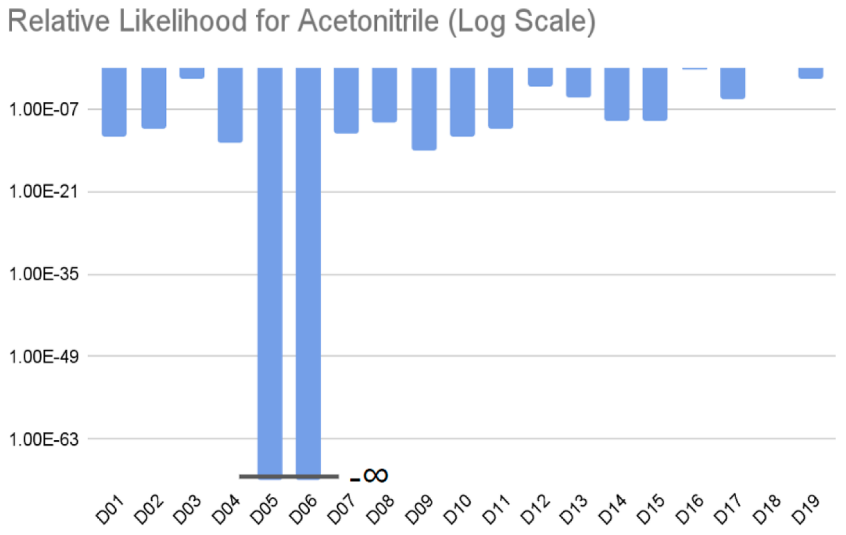


Figure 9. Log-scale plot of relative likelihood values for molecules in acetonitrile.

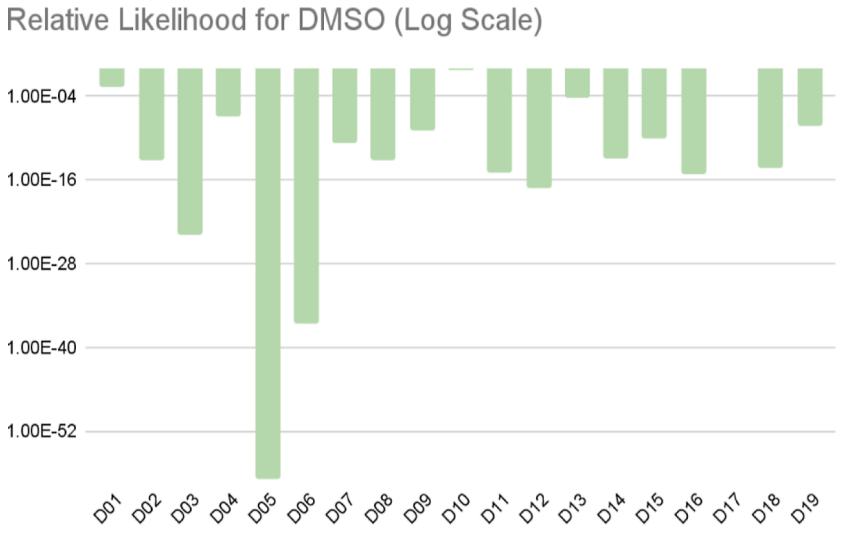


Figure 10. Log-scale plot of relative likelihood values for molecules in DMSO.

From Figures 7 and 8, we can conclude that the solvent model can be turned off for geometry optimization and frequency analysis calculations. We can also see that turning off the solvent model for the single-point calculation has a minimal impact on accuracy for molecules in DMSO compared to those in acetonitrile. This is because the effect of using a continuum solvent model is largely influenced by the net charge of the solute, and we apply the same vertical shift to all data points in a single data set. If all molecules in a data set have the same charge, such as those in DMSO, the near-identical effect of turning off the solvent model is effectively "canceled out" by the vertical shift. However, if any molecules have a different charge compared to the rest of the data set, like molecules 1 and 10 in acetonitrile, their calculated hydricities differ drastically from the other molecules, leading to the result shown in Figure 2b.

Comparing the effects of using PCM versus SMD, SMD does perform slightly better than PCM, but the difference is minimal. When comparing models D02 and D15, which differed by the use of PCM versus SMD in the single-point energies, we calculated a likelihood ratio of 0.04, which is barely sufficient to claim a statistically more accurate result if a P-value threshold of 0.05 is used. However, SMD is a more heavily parameterized model with a greater computational cost, and our SMD calculations often required the use of custom solver parameters to avoid convergence errors, which further increased the time needed to reach the solution. This leads us to recommend using PCM when calculating the hydricity of an organic hydride.

Last, we examine the relative likelihoods of each method, shown in Figures 9 and 10, to test the statistical significance of our results. We will mainly focus on the relative likelihoods for acetonitrile as half of the molecules in DMSO were dihydro-tricyclic heterocycles with large RMSEs. Most methods give a log-scale relative likelihood on the order of 10^{-12}. Looking at models that used ωB97X-D3 (D03, D12, D13, D16) and M06-2X-D3 (D17, D18, D19), we can see that all such models gave a relative likelihood noticeably closer to 1 (closer to 0 in the log scale) compared to other methods, giving us confidence that the improved accuracy coming from using ωB97X-D3 or M06-2X-D3 over other density functionals is statistically significant. The difference in accuracy between the two models (D16, D18) is not statistically significant, so users should

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

expect equivalent accuracy between the ωB97X-D3 or M06-2X-D3 functionals for single-point energies.

The improvement in accuracy upon using B3LYP-D3 to optimize the structures and calculate the free-energy corrections (D16, D18) instead of performing all steps with the more expensive level of theory (D03) does not seem very statistically significant, but since the former method is more computationally efficient, we would recommend using ωB97X-D3//B3LYP-D3 or M06-2X-D3//B3LYP-D3.

The effect of changing the basis set generally is not statistically significant, comparing across D01, D02, D08, and D09 (B3LYP-D3/*X*/PCM). Comparing across D03, D12, and D13 (ωB97X-D3/*X*/PCM), TZVP outperforms 6-31G* and TZVP+ for molecules in acetonitrile, while 6-31G* gave the lowest RMSE for molecules in DMSO. This confirms our statement above that using the largest basis set is not recommended, although we suspect that conformational flexibility is a larger source of error for the DMSO data set.

The relative likelihoods of methods using no solvent model at all (D05, D06) show that the improvements coming from using a solvent model during the single-point calculation are also statistically significant. Meanwhile, the effect of using SMD (D14, D15) over PCM is not statistically significant, which again leads us to recommend using PCM, the less computationally expensive solvent model.

4. CONCLUSIONS

When calculating the hydricity of organic hydrides, we recommend using B3LYP-D3 to optimize the molecular geometry and obtain the free-energy corrections and ωB97X-D3 or M06-2X-D3 to compute the single-point energy, a basis set no larger than TZVP, and C-PCM ISWIG as the solvent model. We also generally advise that PCM may be turned off for geometry optimization and frequency analysis calculations, but it needs to be kept on for the single-point calculations. If the molecule of interest has a net neutral charge, or if one is interested only in the relative hydricities across different molecules with the same charge, then it is safe to turn off PCM for the single-point calculations as well. When calculating the hydricity of molecules with large conformational flexibility, we advise optimizing a greater number of initial starting geometries or performing conformational sampling to ensure that the molecular structures are true global minima.

A future direction of this project is running ab initio molecular dynamics simulations to study the hydration of solutes in water. Continuum solvent models such as PCM or SMD fail to model strong solvent−solvent and solvent−solute interactions such as hydrogen bonding, so for hydrides in water, a simple DFT calculation with just a continuum solvent will not suffice. A diagnostic algorithm that can determine the importance of certain solvent−solute interactions and give the user a recommendation as to which solvent model to use is also a future direction that would be valuable to the theoretical chemistry literature. Such an algorithm could be developed using machine learning approaches.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c03072.

Quantum chemistry software input files; tables of calculated free energies of hydricity half reaction and calculated hydricity; charge and spin multiplicities of donor and acceptor structures; effects of dispersion correction on accuracy; conformational change of molecule 26 after heterolytic C−H bond dissociation; and comparison between hydricities of three molecules obtained in this work versus those obtained in Ilic et al.[14] (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Lee-Ping Wang − *Department of Chemistry, University of California, Davis, Davis, California 95616, United States;* orcid.org/0000-0003-3072-9946; Email: leeping@ucdavis.edu

Authors

Christina Yeo − *Department of Physics and Astronomy, University of California, Davis, Davis, California 95616, United States;* Present Address: Department of Physics and Astronomy, University of California, Los Angeles. 405 Hilgard Ave, Los Angeles, CA 90095.

Minh Nguyen − *Department of Chemistry, University of California, Davis, Davis, California 95616, United States;* Present Address: Department of Chemistry and Biochemistry, University of California, Los Angeles. 405 Hilgard Ave, Los Angeles, CA 90095

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c03072

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Louise Berben for inspiring this project. We would also like to recognize former and current group members Yudong Qiu for developing a large fraction of our group's cluster tools that we still use and benefit from to this day, Hyesu Jang for her helpful modifications to the TeraChem output file format, and Nathan Yoshino for helping us figure out the difference between B3LYP and B3LYP*. We would also like to acknowledge Maria Fernanda Guizar for answering our questions about statistical hypothesis testing. C.Y. acknowledges funding support from a UC Regents scholarship. M.N. acknowledges funding support from a McNair scholarship. This research was further supported by ARO award number W911NF-17-1-0434.

■ REFERENCES

(1) Wiedner, E. S.; Chambers, M. B.; Pitman, C. L.; Bullock, R. M.; Miller, A. J. M.; Appel, A. M. Thermodynamic Hydricity of Transition Metal Hydrides. *Chem. Rev.* **2016**, *116*, 8655−8692.

(2) Dedieu, A. *Transition Metal Hydrides*; Wiley-VCH, 1992.

(3) Berning, D. E.; Noll, B. C.; DuBois, D. L. Relative Hydride, Proton, and Hydrogen Atom Transfer Abilities of [HM(diphosphine)2]PF6 Complexes (M = Pt, Ni). *J. Am. Chem. Soc.* **1999**, *121*, 11432−11447.

(4) Curtis, C. J.; Miedaner, A.; Ellis, W. W.; DuBois, D. L. Measurement of the Hydride Donor Abilities of [HM(diphosphine)2]+ Complexes (M = Ni, Pt) by Heterolytic Activation of Hydrogen. *J. Am. Chem. Soc.* **2002**, *124*, 1918−1925.

(5) Lim, C.-H.; Holder, A. M.; Hynes, J. T.; Musgrave, C. B. Reduction of CO2 to Methanol Catalyzed by a Biomimetic Organo-Hydride Produced from Pyridine. *J. Am. Chem. Soc.* **2014**, *136*, 16081−16095.

(6) Lim, C.-H.; Holder, A. M.; Hynes, J. T.; Musgrave, C. B. Catalytic Reduction of CO2 by Renewable Organohydrides. *J. Phys. Chem. Lett.* **2015**, *6*, 5078−5092.
(7) Lim, C.-H.; Ilic, S.; Alherz, A.; Worrell, B. T.; Bacon, S. S.; Hynes, J. T.; Glusac, K. D.; Musgrave, C. B. Benzimidazoles as Metal-Free and Recyclable Hydrides for CO2 Reduction to Formate. *J. Am. Chem. Soc.* **2019**, *141*, 272−280.
(8) Rakowski Dubois, M.; Dubois, D. L. Development of Molecular Electrocatalysts for CO2 Reduction and H2 Production/Oxidation. *Acc. Chem. Res.* **2009**, *42*, 1974−1982.
(9) Loewen, N. D.; Neelakantan, T. V.; Berben, L. A. Renewable Formate from C−H Bond Formation with CO2: Using Iron Carbonyl Clusters as Electrocatalysts. *Acc. Chem. Res.* **2017**, *50*, 2362−2370.
(10) Yang, X.; Walpita, J.; Zhou, D.; Luk, H. L.; Vyas, S.; Khnayzer, R. S.; Tiwari, S. C.; Diri, K.; Hadad, C. M.; Castellano, F. N.; Krylov, A. I.; Glusac, K. D. Toward Organic Photohydrides: Excited-State Behavior of 10-Methyl-9-phenyl-9,10-dihydroacridine. *J. Phys. Chem. B* **2013**, *117*, 15290−15296.
(11) Bullock, R. M.; Appel, A. M.; Helm, M. L. Production of Hydrogen by Electrocatalysis: Making the H−H Bond by Combining Protons and Hydrides. *Chem. Commun.* **2014**, *50*, 3125−3143.
(12) Ilic, S.; Alherz, A.; Musgrave, C. B.; Glusac, K. D. Thermodynamic and Kinetic Hydricities of Metal-Free Hydrides. *Chem. Soc. Rev.* **2018**, *47*, 2809−2836.
(13) Muckerman, J. T.; Achord, P.; Creutz, C.; Polyansky, D. E.; Fujita, E. Calculation of Thermodynamic Hydricities and the Design of Hydride Donors for CO2 Reduction. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15657−15662.
(14) Ilic, S.; Pandey Kadel, U.; Basdogan, Y.; Keith, J. A.; Glusac, K. D. Thermodynamic Hydricities of Biomimetic Organic Hydride Donors. *J. Am. Chem. Soc.* **2018**, *140*, 4569−4579.
(15) Zhu, X.-Q.; Wang, C.-H.; Liang, H.; Cheng, J.-P. Theoretical Prediction of the Hydride Affinities of Various p- and o-Quinones in DMSO. *J. Org. Chem.* **2007**, *72*, 945−956.
(16) Kang, S.-B.; Cho, Y.; Hwang, S. Density Functional Theoretical Study on the Hydricities of Transition Metal Hydride Complexes in Water. *Bull. Korean Chem. Soc.* **2009**, *30*, 2927−2929.
(17) Goodfellow, A. S.; Bühl, M. Hydricity of 3d Transition Metal Complexes from Density Functional Theory: A Benchmarking Study. *Molecules* **2021**, *26*, 4072.
(18) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171−179.
(19) Dennington, R.; Keith, T. A.; Millam, J. M. *GaussView* version 6; Semichem Inc.: Shawnee Mission KS, 2019.
(20) Miller, A. J. M.; Labinger, J. A.; Bercaw, J. E. Trialkylborane-Assisted CO2 Reduction by Late Transition Metal Hydrides. *Organometallics* **2011**, *30*, 4308−4314.
(21) Cheng, J.; Handoo, K. L.; Parker, V. D. Hydride Affinities of Carbenium Ions in Acetonitrile and Dimethyl Sulfoxide Solution. *J. Am. Chem. Soc.* **1993**, *115*, 2655−2660.
(22) Zhang, X.-M.; Bruno, J. W.; Enyinnaya, E. Hydride Affinities of Arylcarbenium Ions and Iminium Ions in Dimethyl Sulfoxide and Acetonitrile. *J. Org. Chem.* **1998**, *63*, 4671−4678.
(23) DuBois, D. L.; Blake, D. M.; Miedaner, A.; Curtis, C. J.; DuBois, M. R.; Franz, J. A.; Linehan, J. C. Hydride Transfer from Rhodium Complexes to Triethylborane. *Organometallics* **2006**, *25*, 4414−4419.
(24) Ellis, W. W.; Raebiger, J. W.; Curtis, C. J.; Bruno, J. W.; DuBois, D. L. Hydricities of BzNADH, C5H5Mo(PMe3)(CO)2H, and C5Me5Mo(PMe3)(CO)2H in Acetonitrile. *J. Am. Chem. Soc.* **2004**, *126*, 2738−2743.
(25) Matsubara, Y.; Fujita, E.; Doherty, M. D.; Muckerman, J. T.; Creutz, C. Thermodynamic and Kinetic Hydricity of Ruthenium(II) Hydride Complexes. *J. Am. Chem. Soc.* **2012**, *134*, 15743−15757.
(26) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(27) Sholl, D.; Steckel, J. A. *Density Functional Theory: A Practical Introduction*; Somerset: Wiley, 2011.
(28) Becke, A. D. A New Mixing of Hartree−Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(29) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(30) Reiher, M.; Salomon, O.; Artur Hess, B. Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48−55.
(31) Seth, M.; Ziegler, T. Range-Separated Exchange Functionals with Slater-Type Functions. *J. Chem. Theory Comput.* **2012**, *8*, 901−907.
(32) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom−Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(33) Zhao, Y.; Truhlar, D. G. 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.
(34) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814−4820.
(35) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(36) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(37) Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.
(38) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for K to Au Including the Outermost Core Orbitals. *J. Chem. Phys.* **1985**, *82*, 299−310.
(39) Roy, L. E.; Hay, P. J.; Martin, R. L. Revised Basis Sets for the LANL Effective Core Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1029−1031.
(40) Xu, X.; Truhlar, D. G. Accuracy of Effective Core Potentials and Basis Sets for Density Functional Calculations, Including Relativistic Effects, As Illustrated by Calculations on Arsenic Compounds. *J. Chem. Theory Comput.* **2011**, *7*, 2766−2779.
(41) 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**, *2*, 799−805.
(42) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.
(43) Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J. Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015**, *11*, 3131−3144.
(44) Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010**, *133*, 244111.
(45) 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. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(46) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard,

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577

R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184−215.

(47) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.

(48) Murtaugh, P. A. In Defense of P Values. *Ecology* **2014**, *95*, 611−617.

(49) Burnham, K. P.; Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer-Verlag, 2002.

(50) Laury, M. L.; Carlson, M. J.; Wilson, A. K. Vibrational Frequency Scale Factors for Density Functional Theory and the Polarization Consistent Basis Sets. *J. Comput. Chem.* **2012**, *33*, 2380−2387.

(51) Kesharwani, M. K.; Brauer, B.; Martin, J. M. L. Frequency and Zero-Point Vibrational Energy Scale Factors for Double-Hybrid Density Functionals (and Other Selected Methods): Can Anharmonic Force Fields Be Avoided? *J. Phys. Chem. A* **2015**, *119*, 1701−1714.

(52) Liu, F.; Proynov, E.; Yu, J.-G.; Furlani, T. R.; Kong, J. Comparison of the Performance of Exact-Exchange-Based Density Functional Methods. *J. Chem. Phys.* **2012**, *137*, 114104.

https://doi.org/10.1021/acs.jpca.2c03072
J. Phys. Chem. A 2022, 126, 7566−7577