THE JOURNAL OF PHYSICAL CHEMISTRY A

pubs.acs.org/JPCA Article

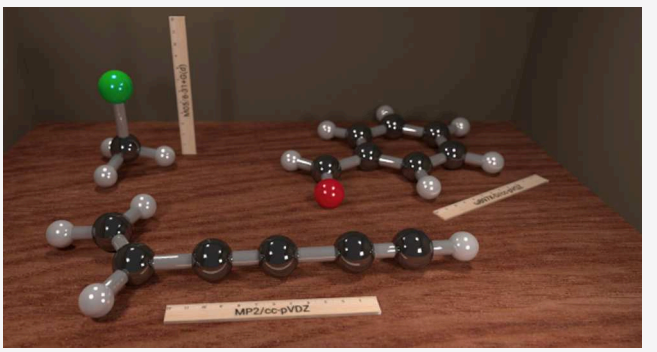
Bayesian Analysis of Theoretical Rotational Constants from Low-Cost Electronic Structure Methods

Kin Long Kelvin Lee* and Michael McCarthy

Cite This: https://dx.doi.org/10.1021/acs.jpca.9b09982 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: With an ever-increasing usage of electronic structure programs by the microwave spectroscopy community, there is a growing need to assess the performance of commonly used, low-cost quantum chemical methods, particularly with respect to rotational constants because these quantities are central in guiding experiments. Here, we systematically benchmark the predictive power afforded by several low-level ab initio and density functionals combined with a variety of basis sets that are commonly employed in the rotational spectroscopy literature. The data set in our analysis consists of 6916 optimized geometries of 76 representative species where high-resolution experimental gas-phase rotational constants are available. We adopted a Bayesian approach for analyzing the performance of each method and basis set combination, employing Hamiltonian Monte Carlo sampling to determine the uncertainty in theoretical predictions of rotational constants and dipole moments. Our analysis establishes a hierarchy of accuracy and uncertainty, with commonly used methods in the rotational spectroscopy literature such as B3LYP and MP2 yielding lower accuracy and higher uncertainty than newer-generation functionals such as those from the Minnesota family, and ωB97X-D, which, when paired with a modestly sized 6-31+G(d) basis, provides optimal performance with respect to computational cost. Additionally, we provide statistical scaling factors that can be used to empirically correct for vibration−rotation effects, as a means to further improve the accuracy of rotational constants predicted from these relatively low-cost theoretical methods. As part of this, we demonstrate that the uncertainties can be used in simulations of rotational spectra to cross-correlate with broadband spectra, a methodology that could be used to quickly and efficiently survey experimental spectra for new molecules.



INTRODUCTION

Rotational spectroscopy is a powerful analytical tool whose usage has grown in recent years; it has been applied in many fields, ranging from the analysis of complex mixtures,[1] measurement of chirality,[2,3] molecular structure determination,[4,5] in addition to its central role in molecular astrophysics and astrochemistry. The allure of rotational spectroscopy lies in its highly selective nature; the observed patterns in spectra depend sensitively on the three rotational constants, which, in turn, are inversely related to the principal moments of inertia of a molecule. Furthermore, spectrometers, either in a laboratory or on a telescope, routinely provide a high spectral resolution on the order of parts per million, meaning that the assignment of a molecule based on a rotational spectrum is normally unambiguous.

Despite being a long-standing measurement technique, the discovery space for rotational spectroscopy remains extraordinarily large. A recent example from our group is the analysis of an electrical discharge of benzene, which resulted in the detection and spectroscopic characterization of nine new hydrocarbon species,[1] a number that has since grown to a dozen.[6] While these discoveries were facilitated, in part, by an efficient workflow, this work highlights the relative ease with which new species can be detected in a broadband spectrum, provided the tools are in place to analyze complex chemical mixtures in an unbiased and exhaustive fashion. Using automated experimental techniques[7,8] in tandem with automated fitting approaches,[9,10] in principle, and increasingly in practice, multicomponent mixtures can be rapidly decomposed and rotational constants determined for unknown molecules.

Determining the elemental composition and molecular structure from a set of experimental rotational constants, however, remains challenging. The most common procedure is to perform electronic structure calculations to optimize chemically intuitive structures and to include corrections that progressively improve the accuracy of the theoretical structure. Nevertheless, the high accuracy of the rotational constants

Received: October 23, 2019
Revised: January 2, 2020
Published: January 7, 2020

necessitates extremely accurate electronic structure methods[11−13] to meaningfully constrain the number of possible molecular carriers. One such approach is composite extrapolation schemes by Barone et al.,[13−15] which rely on geometry optimizations performed with coupled-cluster methods combined with extrapolations to the complete basis set limit of the correlation contribution to the molecular geometry. Vibration−rotation coupling is usually calculated with perturbation theory (e.g., second-order vibrational perturbation theory[16]) or with variational solutions (e.g., vibrational self-consistent theory[17,18]) to the nuclear Schrödinger equation. Although this approach typically yields highly accurate rotational constants (to 0.1% or better), it comes at a considerable cost, i.e., well in excess of the time needed to perform the laboratory experiment for a single molecule. Moreover, this level of sophistication is intractable for all but the smallest molecules due to the need for higher-order coupled-cluster terms [e.g., CCSDT(Q)] and large basis sets that not only include many shells but also specifically tuned to incorporate scalar relativistic contributions,[11] core−valence interactions,[19,20] and so on to the molecular structure.

For these reasons, it is highly desirable to accurately predict rotational constants with a reduced computational cost to better match the cadence of spectral acquisition and analysis. Determination of empirical scaling factors, such as those widely adopted to correct for anharmonicity in harmonic frequency analysis,[21,22] would appear to be an attractive approach, provided the electronic structure of a molecule is treated equally well for a given theoretical method and basis set and that discrepancies between experiment and theory are captured within some statistical model.

To assess the usefulness of this approach, it is essential to benchmark equilibrium rotational constants of various electronic structure methods relative to experimental data. While there are many examples of systematic benchmarking of structures in the literature (e.g., Martin et al.,[23] Spackman et al.,[24] Bak et al.,[25] Coriani et al.,[26] and Jensen[27]), the motivation is usually to demonstrate "state-of-the-art": the focus is toward small and relatively light (2−4 nonhydrogen atoms) species; comparisons are based on bond lengths, and high-accuracy and computationally expensive ab initio methods such as coupled-cluster theory are employed. A somewhat different comparison is the ROT25 benchmark of Grimme and Steinmetz,[28] where experimental rotational constants of nine (bio)organic molecules were compared to dispersion-corrected density functional theory (DFT) calculations. This study, however, specifically targeted fairly large, asymmetric molecules, with all but one being close to the prolate limit ($A \gg B \approx C$). Furthermore, given the small size of this data set, it is difficult to derive scaling factors with statistical confidence.

To robustly establish the performance of low-cost methods, the ideal sample set should contain molecules whose structures range from the prolate to the oblate limit ($-1 \leq \kappa \geq 1$, where κ is Ray's asymmetry parameter)[29] and include many calculations to derive statistically meaningful results. With this in mind, the aim of the present study is to assess the performance of commonly used, low-cost electronic structure methods and basis sets, specifically with respect to their ability to predict rotational constants. The intent is to provide a guide for experimentalists in choosing a cost-effective method and basis set combinations with minimal compromise in accuracy and precision. By employing a Bayesian approach to our analysis, we have also sought to quantify statistical uncertainties and, in turn, to determine the most reliable methodology. Compared to conventional frequentist statistics, Bayesian modeling provides statistical estimates by integrating over possible parameter space, as opposed to simple least-squares fits over a limited set of observations. The result is a robust model of method uncertainty, in other words, how likely the predicted equilibrium value of a given method and basis set deviates from the experimental vibrationally averaged one. Based on these models, scaling factors can then be derived that effectively correct for vibration−rotation interaction under the assumption that it is the dominant source of error, without the need to explicitly compute vibration−rotation interaction terms. As a demonstration, we detail a workflow of how these uncertainties can be used to predict rotational spectra, either to constrain the frequency range of a targeted experimental search for a new species or to search through broadband spectra in a high-throughput fashion for evidence of undiscovered molecules.

THEORETICAL METHODS

Electronic structure calculations were performed using the Gaussian '16 suite of programs.[30] The methods we considered are commonly used in the rotational spectroscopy literature, including Møller−Plesset second-order perturbation theory[31] (MP2) and the Becke-3−Lee−Yang−Parr (B3LYP) hybrid density functional.[32] As a point of comparison, we also considered a range of functionals, including the local generalized gradient approximation (GGA) functional with empirical dispersion corrections B97-D3,[33] the long-range corrected hybrid GGA ωB97X-D functional from Head-Gordon and co-workers,[34] and the hybrid meta-GGA Minnesota functionals from the Truhlar group,[35,36] which have been engineered to produce reliable structures, energetics, and in some of the cases include dispersion or long-range corrections that are non-negligible for a balanced treatment of many kinds of systems.[37,38] The basis sets considered here include split-valence basis sets of Pople[39,40] (i.e., X−YZG, where X = 6, Y = (3,2), and Z = 1) and the correlation-consistent basis sets of Dunning (cc-pVXZ, where X = D, T, Q).[41,42] For the Pople basis sets, we also considered a minimum amount of additional polarization and diffuse functions. Table 1 summarizes the methods and basis sets used in our calculations.

Table 1. Electronic Structure Methods and Basis Sets Considered in This Work[a]

method		basis	
ab initio	density functional	Pople	Dunning
HF[43,44]	B3LYP[32]	3-21G	
MP2[31]	BLYP[45]	6-31G	cc-pVDZ
	M05-2X[36]	6-31G(d)	cc-pVTZ
	M06-2X[46]	6-31+G(d)	cc-pVQZ
	M05[36]		
	M06[46]		
	ωB97X-D[34]		
	PW6B95-D3[35]		
	PBEPBE[47]		
	B97-D3[33]		
	O3LYP[48]		

[a]References for the density functionals employing Grimme's D3 dispersion correction; refs 49 and 50 refer to the original publications.

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Calculations of the molecular structure were automated using a set of Python scripts in which molecules are first specified using the SMILES representation, and OpenBabel[51] is then used to generate the corresponding three-dimensional Cartesian coordinates, followed by rudimentary optimization with universal force field (UFF) molecular mechanics.[52] The coordinates are used as an initial guess for a chain of geometry optimizations, starting with the least-sophisticated method, Hartree−Fock (HF), and 3-21G basis. Subsequent calculations use the previous geometry as the starting point, which is typically the same electronic structure method with a smaller basis set. Geometries are considered to be converged when the root-mean-squared value for the change in gradient is less than 10^{-6} Hartrees/Bohr. The integration grid used for all calculations is the standard "ultrafine" setting implemented in Gaussian. No symmetry constraints were artificially imposed in any of the calculations.

We have attempted to select an unbiased group of molecules for this study; it consists of 76 molecules with a variety of structures (i.e., linear, symmetric, and asymmetric tops; Figure 2) and a large distribution of masses (from roughly 20 to 120 amu; Figure 1), for which experimentally derived rotational

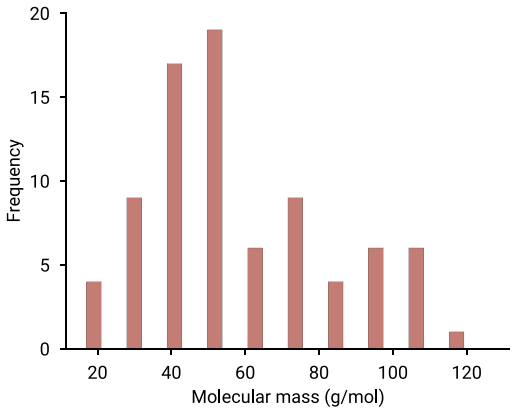


Figure 1. Distribution of molecular masses considered in this work.

constants have been reported in the literature. In the vast majority of cases, constants were determined by pure rotational spectroscopy, although some species (e.g., benzene) are highly symmetrical and therefore possess neither a permanent dipole moment nor a rotational spectrum. In these instances, rotational constants were determined by other high-resolution techniques, typically at infrared wavelengths. For this same reason, Figure 1 appears slightly skewed toward smaller molecules (~50 amu), which simply reflects that the rotational spectroscopy is far more common for relatively light species. A complete list of molecules, rotational constants, dipole moments, and references is provided in the Supporting Information (Table S1).

In selecting the data set, however, we have chosen to exclude classes of molecules with complicated electronic structure and nuclear motion. Examples include weakly bound or nonrigid molecules where vibrational coupling can be nonperturbative (e.g., van der Waals complexes, large amplitude motion) and species that have significant competition between dynamic and static correlation effects (e.g., heavy transition metals, radicals). For these systems, it is very likely that the low-cost methods investigated here would fail,[25,53,54] and more sophisticated calculations would be required. Because our approach requires on statistical inference, where these complicated species are likely to be outliers, our modeling focuses on well-behaved, stable species that are typical to those observed in broadband assays of electrical discharge mixtures.[1]

We estimated the performance of each method and basis set combination in predicting rotational constants as a signed percentage difference $\delta A(BC)$

$$\delta = 100 \times \frac{A(BC)_{0,\mathrm{exp}} - A(BC)_{e,\mathrm{theory}}}{A(BC)_{0,\mathrm{exp}}} \quad (1)$$

where $A(BC)_{e,\mathrm{theory}}$ denotes theoretical values, and $A(BC)_{0,\mathrm{exp}}$ are the experimental ones. For a given method/basis, the probability mass function, $\Pr(\delta)$, is computed by histogram-binning the values of δ across all of the molecules studied. Treating the three rotational constants as independent observations equates to a total of 228 observations (with 76 molecules) for each method-basis set combination.

Comparisons between theoretical equilibrium dipole moments and experimental values were made by absolute differences (Δ_μ). While dipole precision is rarely a requirement in laboratory searches, the same is not true in predicting intrinsic line strengths ($S_{ij}\mu^2$) since this quantity is important in radiative transfer modeling and for deriving column density estimates from astronomical observations. Since line strengths depend on the square of the dipole moment, uncertainties in units of Debye, rather than percentage differences, are more relevant. We note that the dipole moment data is limited somewhat by experiment: among the 76 molecules, experimental dipole moments along the *a*-axis have been reported for 42, 15 along the *b*-axis and 3 along the *c*-axis. Among the 16 remaining species, 3 do not possess permanent dipole moments and the remaining 13 have not yet been measured.

In this work, we apply a Bayesian approach to model the distributions $\Pr(\delta)$ and $\Pr(\Delta_\mu)$ by approximating the observations as Student's *t*-distributions, $f(\nu, \mu, \sigma)$

$$f(\nu, \mu, s) = \frac{G\left(\frac{\nu+1}{2}\right)}{G\left(\frac{\nu}{2}\right)\sqrt{\mathrm{p}\nu s^2}}\left(1 + \frac{1}{\nu}\frac{(x-\mu)^2}{s}\right)^{-\nu+1/2} \quad (2)$$

where Γ is the γ function, μ and σ have the conventional definitions of mean and standard deviation, and ν determines the decay rate of the distribution tail. The parameter ν allows for better flexibility in the modeling of data and thus not constrained as normally distributed observations; we believe that this is a reasonable approach since not every molecule can be equally well described by a given method and basis set.

Equation 3 expresses our model in terms of Bayes' theorem, comprising the observed data $\Pr(y)$ (which is substituted for δ or Δ_μ) and the prior $\Pr(\nu, \mu, \sigma)$, and the likelihood $\Pr(y|\nu, \mu, \sigma)$. The joint prior distribution $\Pr(\nu, \mu, \sigma)$ is estimated by sampling from typical probability distributions: for μ and σ, we chose to use normal distributions while ν is taken from a uniform distribution; tests with different distributions showed that the posteriors are insensitive to the choice of priors. The objective is to determine the posterior distribution, $\Pr(\nu, \mu, \sigma|y)$, by integration over all possible model values of ν, μ, and σ. Samples can then be drawn from the posterior distribution to predict the likelihoods of interest: for example, the probability of a given method/basis is offset from the experimental value by δ.

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

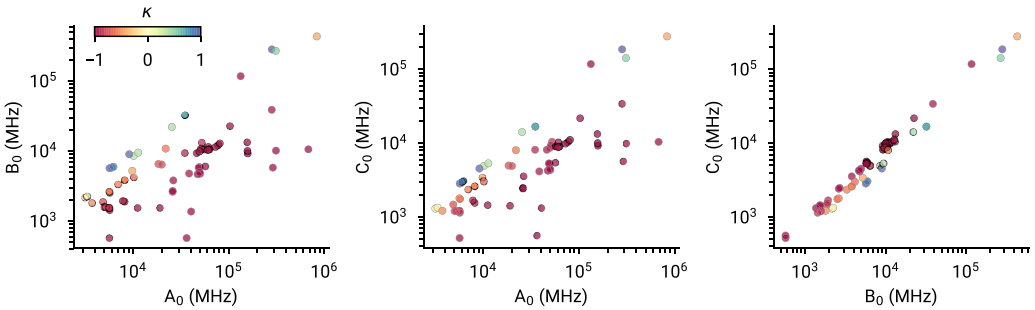


Figure 2. Distribution of rotational constants for the 76 nonlinear molecules considered in this benchmark study. Each data point corresponds to an experimentally determined combination of rotational constants, and its color represents Ray's asymmetry parameter (κ), which varies from the prolate ($\kappa = -1$) to the oblate ($\kappa = 1$) limit.

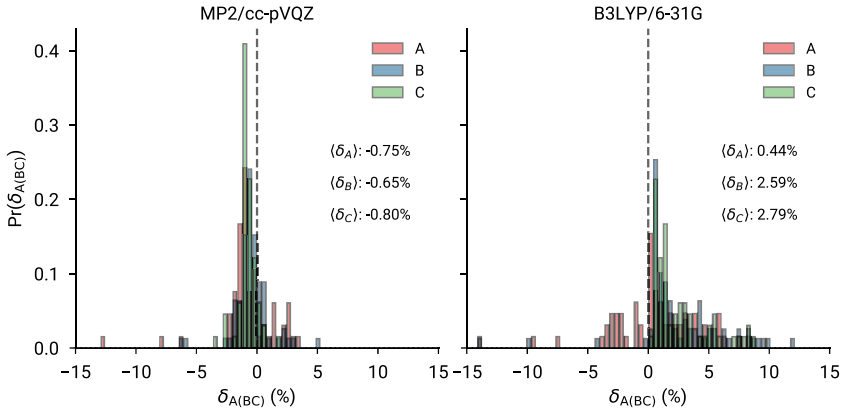


Figure 3. Deviations (δ, in percentage) of equilibrium rotational constants computed at the MP2/cc-pVQZ (left) and B3LYP/6-31G (right) levels of theory relative to the experimental constants. In both plots, each rotational constant is indicated by a different color. The dashed line indicates zero error; the bin spacing is roughly 0.4%.

$$\Pr(\nu, \mu, s|y) = \frac{\Pr(y|\nu, \mu, s)\Pr(\nu, \mu, s)}{\Pr(y)} \quad (3)$$

To perform the integration, we used the Markov chain Monte Carlo (MCMC) methods implemented in the PyMC3 package.[55] The MCMC sampling was carried out with the Hamiltonian-based no-U turn sampling (NUTS) algorithm, which has been shown to improve the rate of convergence over sampling methods such as the canonical Metropolis−Hastings algorithm.[56] The sampling was distributed among four parallelized Markov chains, where the first 2000 simulations ("burn-in") are discarded to ensure equilibration, with 5000 samples taken thereafter from each chain, and the results are combined to give a total of 20 000 samples per run.

RESULTS AND DISCUSSION

Electronic Structure Calculations. In general, the electronic structure results for rotational constants and dipole moments between combinations of methods and basis sets are all in qualitative agreement with one another. There are, however, some fringe cases that warrant extra discussion, where the combined method and basis give qualitatively incorrect results. Perhaps most serious is when the M05 and M06 methods are used with a 6-31G basis set, which predicts a planar equilibrium structure for NH_3, a finding that was confirmed by harmonic frequency analysis (i.e., no imaginary frequencies). The addition of diffuse and polarization functions in the split-valence basis set alleviates this issue, and there are no other issues based on the current data set with the M05 and M06 functionals. This occurrence does indicate that significant deviations from statistical behavior are possible for low-cost theoretical calculations even for simple molecules like NH_3, although such cases appear to be fairly isolated and infrequent, as highlighted in this case, it appears best to avoid basis sets without diffuse and polarization functions for all calculations Figure 2.

Another isolated nonstatistical result was found for cyclopropenylidene (c−C_3H_2): the three methods BLYP, B97-D3, and MP2 combined with the minimal split-valence basis set 3-21G predict the permanent electric dipole moment lies along the *a*-axis, whereas other method-basis set combinations predict the dipole moment along the *b*-axis, in agreement with the experiment. In this instance, the dipole axes are switched due to small differences in the mass distribution: distances within the 3-membered carbon ring change with the degrees of dynamic electron correlation, resulting in exchange of the *a* and *b* principal inertial axes. Since the effect of correlation here is subtle, it highlights the need for sophisticated treatments of electron correlation, particularly with reactive species including carbenes such as c−C_3H_2.

To illustrate typical results, Figure 3 shows three histograms of δ, the percentage difference between the theoretical and experimental rotational constants, for the MP2/cc-pVQZ and B3LYP/6-31G calculations. In this example, MP2/cc-pVQZ represents an ab initio method with the largest correlation-consistent basis set used in this study, whereas the B3LYP/6-31G result is indicative of a commonly used functional, paired with a modest split-valence basis set without diffuse or polarization functions. The former demonstrates an ideal statistical case: each rotational constant is predicted with

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

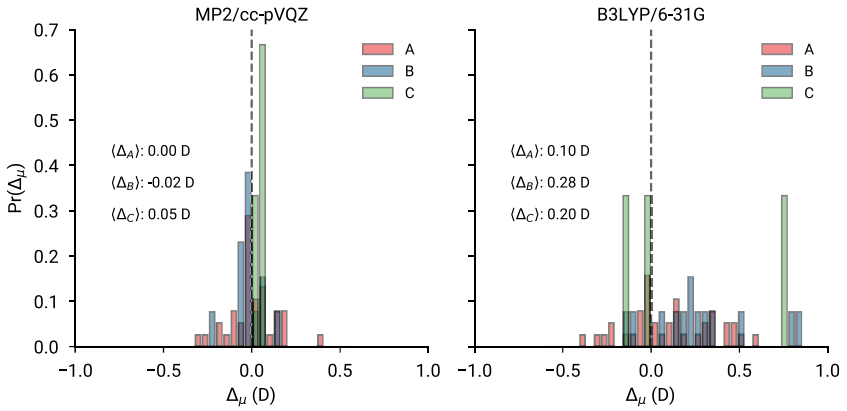


Figure 4. Absolute deviations (in Debye) of equilibrium dipole moments computed at the MP2/cc-pVQZ (left) and B3LYP/6-31G (right) levels of theory relative to the experimental values. The three projections of the dipole moment are binned separately and indicated in different colors. The dashed line indicates zero error.

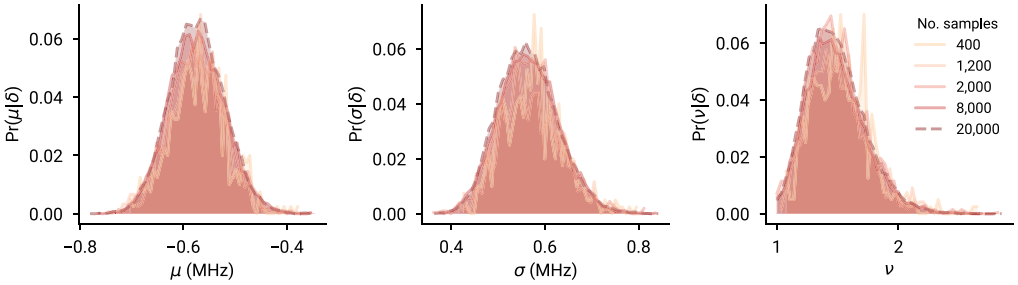


Figure 5. Posterior distributions of the model variables μ, σ, and ν for δ as a function of the number of samples, on the MP2/cc-pVQZ data set for a fixed burn-in of 2000 samples. Values in the legend correspond to the total number of samples, summed over four Markov chains. The dashed curve indicates the distribution corresponding to 20 000 samples, which is the amount used for inference.

Table 2. Representative Parameter Statistics Obtained from the Posterior Distributions Shown in Figure 5[a]

parameter	mean	sampling error	% error	95% HPD[b]	$\hat{R} - 1$
μ_δ	−0.5693	0.0005	0.0816	−0.6741:−0.4650	1.2×10^{-4}
σ_δ	0.5691	0.0006	0.1058	0.4496:0.6990	1.7×10^{-5}
ν_δ	1.5059	0.0024	0.1621	1.0468:1.9577	9.2×10^{-5}

[a]The sampling error and percentage error correspond to the uncertainty associated with the Monte Carlo sampling, the latter as a percentage of the mean. [b]Highest posterior density; 95% credible interval.

similar accuracy and precision, and the distribution peak overestimates by 0.7%, with a spread of ∼3% and a distribution that appears Gaussian-like. The latter case is qualitatively different; the *A* rotational constant is predicted with greater accuracy than either *B* and *C*, both of which are significantly less precise than the MP2/cc-pVQZ results. The distributions of $\Pr(\delta_B)$ and $\Pr(\delta_C)$ appear to be shifted toward larger values of δ, whereas the spread in $\Pr(\delta_A)$ is roughly centered near zero. In contrast to the MP2/cc-pVQZ results, the B3LYP/6-31G histograms do not appear to conform to a normal distribution.

Figure 4 makes a similar comparison to Figure 3 but, instead, shows the absolute differences in theoretical equilibrium dipole moments relative to the experimental values. Notwithstanding limited experimental data, as with the comparison of rotational constants, the MP2/cc-pVQZ predictions are both more accurate and precise than the B3LYP/6-31G counterparts. In the former case, <Δ> is close to zero with a spread of ±0.2 D, whereas the latter <Δ> is roughly 10 times larger with a much larger spread. Once again, the MP2/cc-pVQZ results appear to have a well-defined statistical distribution, while the B3LYP/6-31G histograms lack a simple obvious pattern.

Bayesian Analysis. Probability distributions of rotational constants and dipole moments allow one to perform MCMC simulations to determine the posterior distribution $\Pr(\nu, \mu, \sigma | y)$. Since the model integration is performed with a sampling (as opposed to variational) method, initial testing was done to ensure convergence of the model parameters with respect to the number of samples—for this, we use the MP2/cc-pVQZ δ results as a representative example of the full data set to demonstrate convergence.

Figure 5 shows the posterior distributions for the three parameters used in Student's *t*-distribution, as a function of the number of MCMC samples. As the number of samples increases, the three distributions become progressively smoother but no qualitative shifts occur; the expectation values and the central position of the peak do not change significantly beyond 8000 samples, and the overall shape closely resembles that of a normal distribution. Additionally, errors arising from the Monte Carlo sampling (Table 2) are ∼0.1% of the mean values and well within the 95% highest posterior density (HPD).

In addition to a qualitative inspection of convergence, we computed two commonly used quantitative indicators: the

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Geweke z-score diagnostic[57] and the Rubin−Gelman $\hat{R}$ statistic.[58] The former measures convergence as the difference in z-scores of the Markov chains at early and late intervals of sampling, with small differences indicating convergence. The latter $\hat{R}$ statistic uses the ratio of intra- and interchain sampling variances to assess convergence—a converged simulation should yield similar values of both types of variance (i.e., a ratio close to one). A plot of the Geweke z-score diagnostics is given in the Supporting Information, and the Rubin−Gelman statistics for each variable are given in Table 2, with both metrics indicating convergence. Full unabridged tables of the parameter statistics can be found in Tables S1 and S2.

Having established MCMC sampling of the parameters converged, samples from the posterior distribution $\Pr(\nu, \mu, \sigma|\delta)$ can now be used to make predictive (model) distributions. As seen in Figure 6, the model successfully

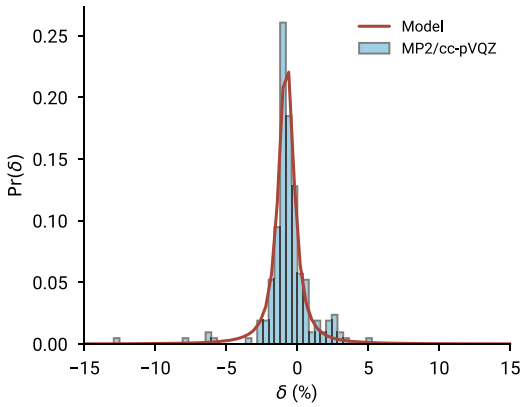


Figure 6. Comparison of the probability density function $\Pr(\delta)$ from the theoretical MP2/cc-pVQZ data set (blue bars) and the posterior predictive distribution following 20 000 samples. Both observed and model data are histogram-binned into ∼12 MHz bins. The vertical lines represent expectation values calculated with the model (dashed) and observations (solid).

captures both the peak and the shape of the observed $\Pr(\delta)$ distribution, and the expectation value of δ obtained from the model (−0.77%) is also in quantitative agreement with the observed value (−0.73%). The unique aspect of adopting a Bayesian approach to the analysis is the ability to derive meaningful uncertainties given the observed data: for MP2/cc-pVQZ, we obtain a 95% highest posterior density (HPD) range of −4.1−2.9%, meaning that our model statistics state that 95% of rotational constants predicted at this combination of theory and basis set should be captured within this δ range, similar to a frequentist confidence interval, with the most probable result, $<\delta>$, being −0.77% of the experimentally determined value. There is a nuanced difference between a frequentist confidence interval and a Bayesian posterior density that should be emphasized: the former is usually determined by fitting observations to scalar values of parameters, assuming that the central limit theorem is valid; as we see in Figures 3 and 4 an assumption that is not necessarily valid. The latter is determined by sampling and therefore integrating over parameter space (i.e., the posterior distribution, $\Pr(\nu, \mu, \sigma|y)$), with the only assumption being that the model posterior faithfully reproduces the observations.

Moving on from the MP2/cc-pVQZ results, we now seek to make comparisons of modeled precision (uncertainty as HPD range) and accuracy ($<\delta>$ proximity to zero) between different method/basis combinations. For brevity, Figure 7 provides a qualitative overview of the model and observed δ results for every method and basis considered; an unabridged version of the modeling results can be found in Table S2. The first point of comparison is uncertainties, which are represented by the top bars in Figure 7. For the sake of simplicity, comparisons of basis sets are averaged across all methods.

Within the split-valence basis sets, there are two main observations. First, a minimal basis (3-21G) and to a lesser extent double-ζ quality basis (6-31G) yield extremely large uncertainties. Second, increasing basis set size decreases uncertainty for all methods except HF. The most substantial decrease is gained by adding polarization functions [6-31G → 6-31G(d)]; on average, the addition of polarization functions halves the 95% HPD range (from 18.5 to 8.5%) for a modest increase in cost. The addition of diffuse functions [6-31G(d) → 6-31+G(d)] decreases the average uncertainty only slightly, by 0.2%. For correlation-consistent basis sets, the smallest basis (cc-pVDZ) yields a similar average 95% HPD range as the 6-31G(d) results (9.3 and 8.5%, respectively). Increasing the basis set up to triple-ζ (cc-pVTZ) and quadruple-ζ (cc-pVQZ) quality basis results gradually decreases in the average uncertainty range, from 9.3 to 8.5 to 8.2%, while the expectation value shifts toward negative values of δ. This is attributed to the systematic improvement of the equilibrium structure predicted by a given correlation method: the equilibrium rotational constants converge toward the complete basis set limit, which acts to lower their uncertainty and "overestimate" the experimental values due to the lack of vibration−rotation interactions. Interestingly, from a statistical perspective, there does not appear to be a clear precision advantage using either basis set families: the 6-31G(d) (8.5%) results are of comparable uncertainty to those obtained with the cc-pVQZ (8.2%) basis, despite the latter being significantly more computationally expensive.

There are a few notable trends that can be gleaned from the methods sampled here. The first concerns overall uncertainty—the low-level ab initio methods (HF and MP2) generally do not perform well, even when paired with a large basis set, e.g., cc-pVQZ. In this context, it is noteworthy that for the HF method, distributions do not change significantly when comparing the cc-pVXZ (X = D, T, Q) results. The implication is that the so-called "HF limit" is reached at least in a qualitative sense even without quadruple-ζ quality basis sets. A similar observation was made by Bak et al.[25] in their study of bond length convergence; they found that HF/cc-p(C)VTZ bond lengths do not significantly differ from the quadruple-ζ counterparts. This finding suggests that composite methods requiring basis set extrapolation of the HF contribution to structures could be truncated, thereby avoiding the need for large basis sets. For instance, in the composite scheme presented by Puzzarini,[15] the HF/CBS contribution is obtained by performing a geometry optimization at the HF/cc-pV6Z level: this work, and that of Bak et al.,[25] suggests that the HF structure and therefore rotational constants converge quickly.

Our results also suggest that some of the more widely adopted methods in the rotational spectroscopy literature such as B3LYP, BLYP, and MP2 are neither necessarily the most precise nor accurate electronic structure methods for determining rotational constants. Table 3 shows the 10 method and basis set combinations that result in the lowest uncertainty in the theoretical rotational constants. The meta-

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

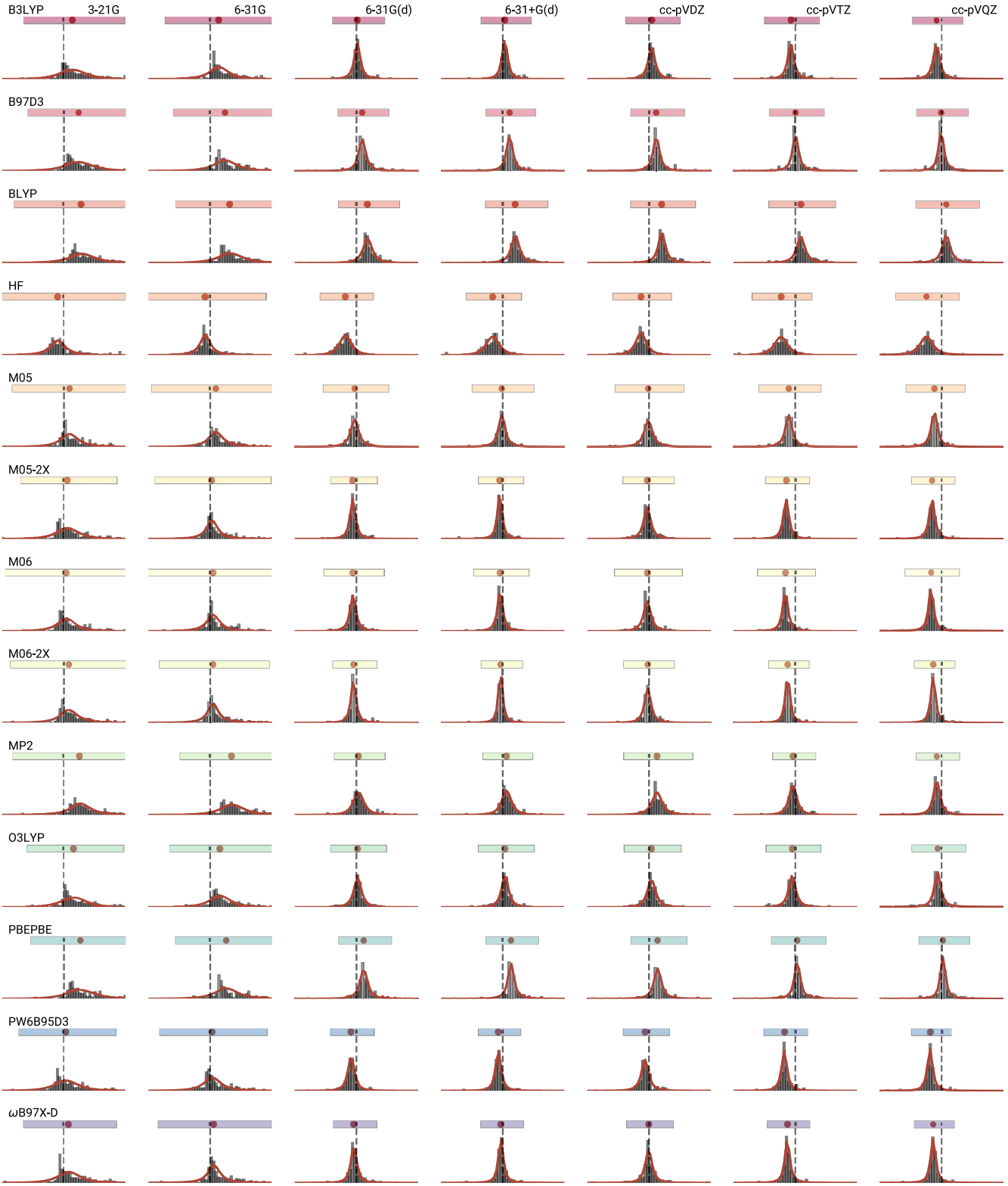


Figure 7. Visualization of the δ posterior predictive distributions (red curves) and histogram-binned observations (bars) for each method (rows) and basis set (columns). Basis sets increase in size from left to right but no particular order has been used for the various methods. The scatter points above distributions indicate the expectation value, while bars represent the 95% HPD interval. The full scale of δ ranges from −10 to 10%, and the dashed line indicates zero.

GGA functionals M05-2X and M06-2X and the hybrid functional ωB97X-D are consistently the most precise, with PW6B95-D3/cc-pVQZ only marginally less uncertain than the ωB97X-D/cc-pVQZ results. These observations are in agreement with the recent benchmarks of other properties such as thermochemistry.[37,38] Within the Minnesota functionals, M06-2X provides a convenient albeit confusing hierarchy of performance and computational cost: cc-pVTZ contains

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 3. Performance of 10 Method and Basis Set Combinations with the Lowest Modeled Uncertainty (HPD Range), Sorted in Ascending Order of HPD Range[a,b]

method	basis set	HPD range	<δ>	lower bound	upper bound
PW6B95-D3	cc-pVQZ	6.48	98.16	95.11	101.59
ωB97X-D	cc-pVQZ	6.49	98.66	95.60	102.08
M06-2X	cc-pVTZ	6.57	98.78	95.70	102.26
	cc-pVQZ	6.61	98.67	95.58	102.19
	6-31+G(d)	6.73	99.65	96.50	103.23
ωB97X-D	cc-pVTZ	6.91	98.72	95.46	102.38
PW6B95-D3	6-31+G(d)	6.93	99.29	96.03	102.96
ωB97X-D	6-31+G(d)	6.98	99.71	96.42	103.40
M05-2X	cc-pVQZ	7.01	98.48	95.17	102.18
MP2	cc-pVTZ	7.01	99.61	96.31	103.32

[a]Values are given as percentages. [b]Upper and lower bounds correspond to the 2.5 and 97.5% HPD boundaries, respectively.

more basis functions than 6-31+G(d) and, for this reason, perhaps not surprisingly yields a smaller uncertainty range. It is unclear as to why the M06-2X/cc-pVQZ results are slightly more uncertain and less accurate than when the cc-pVTZ basis set is used instead. The difference, however, is not large enough to suggest where the fault may lie, either in the statistical analysis or in the electronic structure.

From Table 3, it is possible to put forth a semiquantitative hierarchy of sophistication, which compromises between computational cost and performance, the overall theme of this paper. While PW6B95-D3/cc-pVQZ and ωB97X-D/cc-pVQZ provide the lowest uncertainty, they are also among the most costly, due to the large basis sets employed. M06-2X/6-31+G(d) and ωB97X-D/6-31+G(d), however, employ significantly smaller basis sets while yielding only a slight increase in the average uncertainty; compared to using cc-pVQZ basis sets, the HPD range and average δ increase by less than 0.5% and roughly 1%, respectively, while requiring approximately four times fewer basis functions. Comparing the two functionals, while M06-2X is more precise, ωB97X-D is more accurate, although in both cases, the expected deviation (<δ>) from the experiment is only ∼0.3% and thus are the most accurate methods considered in this work.

More generally, the precision and accuracy obtained here are comparable to those reported using state-of-the-art methods despite requiring a fraction of the computational cost—qualitatively, a difference between minutes and days or even weeks. For example, Puzzarini[59] reports the mean frequentist errors on the order of 0.2% while including complete basis set extrapolation, core−valence interaction, vibrational corrections, and perturbative quadruple excitations in the coupled-cluster expansion, albeit with much smaller uncertainties. The comparable average accuracy from these low-cost method/basis set combinations would therefore appear to be attractive and statistically reliable ways of predicting rotational constants.

Dipole Moments. A similar analysis has also been performed for the theoretical dipole moments. Figure 8 provides a qualitative overview of Δ_μ for each method and basis set; the full set of modeling results can be found in Table S3. As with the rotational constants, the smallest split-valence (3-21G and 6-31G) basis sets result in the largest uncertainty averaged across the methods, with the average 95% HPD range spanning over 1 D. The addition of polarization and diffuse functions decrease the uncertainty substantially, from an average of 1.3 D (6-31G), to 0.9 D (6-31G(d)), to 0.7 D (6-31+G(d)). The $<\Delta_\mu>$ calculated using the 6-31+G(d) basis set consistently overestimates dipole moments regardless of the method used.

Correlation-consistent basis sets generally outperform the split-valence basis sets, in terms of both precision and accuracy. The average uncertainty ranges from cc-pVDZ to cc-pVQZ are 0.8, 0.6, and 0.5 D, respectively, with cc-pVDZ performing marginally better than the 6-31G(d) basis set. Analogous to that done for the rotational constants, Table 4 summarizes the 10 method and basis set combinations with the lowest predicted uncertainties in dipole moment. In contrast to the results from the previous section, no split-valence basis sets appear on this list, which is instead dominated by those that are at least triple- or quadruple-ζ quality. We thus conclude that no obvious compromise in basis set is possible to accurately and precisely predict dipole moments (approximately ±0.2 D), although it does appear that different electronic structure methods are relatively invariant with respect to these two factors.

Predicting Rotational Spectra. By deriving uncertainties for low-cost electronic structure methods and basis sets, one can begin to make informed predictions for rotational spectra based on the equilibrium theoretical rotational constants. The probability distributions in Figure 7, or at a reduced level, the 95% HPD, can be used to determine the upper and lower limits for every transition frequency predicted, which will provide experimentalists with statistically meaningful constraints of the required search ranges. In the former case, the distributions in Figure 7 represent $\Pr(\nu, \mu, \sigma|\delta)$: the posterior probability that any given value of $A(BC)_e$ is offset from the experimental value by δ (eq 1). Here, δ effectively acts as a statistical correction for vibration−rotation interaction, with the assumption that the method and basis set adequately describe a given molecule. If $\bar{A}(BC)_0$ represents the statistical "experimental" rotational constants of a new molecule, which we wish to identify, then eq 1 can be rearranged to scale the theoretical values

$$\bar{A}(BC)_0 \approx \frac{100 \times A(BC)_{e,\text{theory}}}{100 - d} \tag{4}$$

For simple use, δ in eq 4 can be substituted by the expectation value and the lower and upper bounds in Table 3 to obtain estimates. A more powerful way to utilize the statistics is to perform predictive sampling of δ and obtain distributions of $\Pr(\bar{A}(BC)_0)$ from scalar theoretical predictions of $A(BC)_e$. The combinations of $\bar{A}(BC)_0$ can subsequently be used to simulate rotational spectra using a standard Hamiltonian model, e.g., with SPCAT, PGopher, or analytic expressions, and the corresponding transition frequencies histogram-binned to yield probability distributions for every transition. This approach differs from the conventional practice of predicting rotational spectra based on theoretical values, which assumes scalar values for $A(BC)_{e,\text{theory}}$, perhaps including some treatment of zero-point vibrational effects, and predicts singular values for every frequency, resulting in a stick spectrum with transitions represented as delta functions. With the method described here, we obtain a spectrum where the simulated line intensity can be convolved with the probability distribution, thereby obtaining a line width based on the statistical uncertainty of a given transition.

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

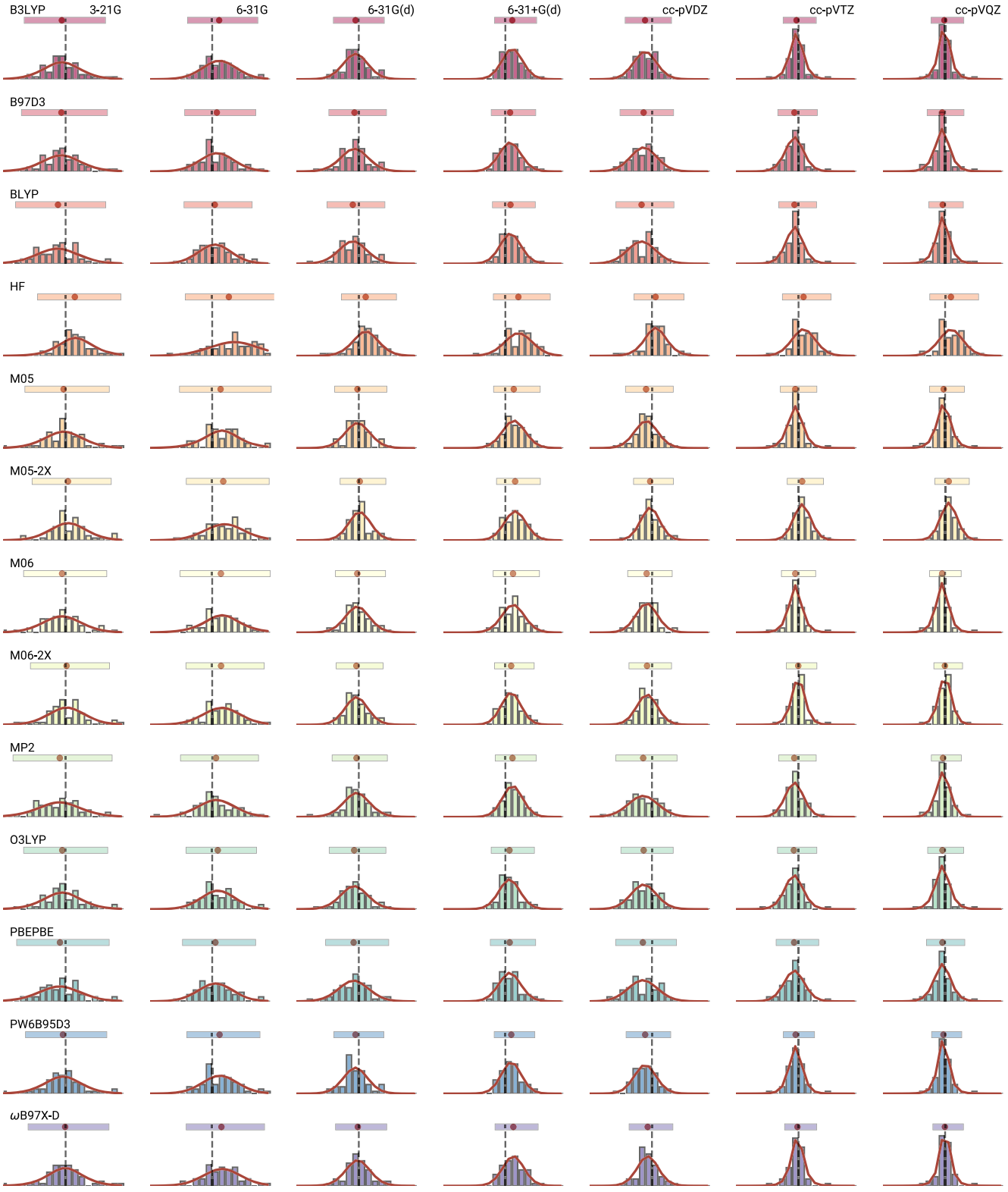


Figure 8. Visualization of the Δ_μ posterior distributions and histogram-binned observations for each method (rows) and basis set (columns). The scatter points above distributions indicate the expectation value, while bars represent the 95% highest posterior density. The full scale of Δ_μ ranges from −1 to 1 D, and the dashed line indicates zero.

As a demonstration, we can produce an experimental simulation by convolving a Gaussian line shape (with widths of 50 kHz) with experimentally determined frequencies, in which white noise has been added to the resulting spectrum. A "prediction" using equilibrium rotational constants can be created by assuming the molecule as a rigid rotor: for a prolate top with a nonzero dipole moment along the A inertial axis, the fundamental a-type transition frequencies are approximated by $\nu(J') = (B + C)J'$. We take the ωB97X-D/cc-pVQZ predictions for $A(BC)_{e,\text{theory}}$, which are shifted by values of δ

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 4. Performance of 10 Method and Basis Set Combinations with the Lowest Modeled Uncertainty (HPD Range) in Dipole Moment Deviation (Δ_μ), Sorted in Ascending Order of HPD Range[a,b]

method	basis set	$\langle\Delta_\mu\rangle \times 10^3$	HPD range	lower bound	upper bound
M06-2X	cc-pVQZ	−3.51	0.46	−0.18	0.28
MP2	cc-pVQZ	−33.89	0.48	−0.22	0.26
PW6B95-D3	cc-pVQZ	−27.07	0.49	−0.22	0.27
ωB97X-D	cc-pVQZ	−3.73	0.49	−0.20	0.30
M06-2X	cc-pVTZ	−4.72	0.49	−0.20	0.30
PW6B95-D3	cc-pVTZ	−47.23	0.50	−0.25	0.26
M06	cc-pVQZ	−47.77	0.51	−0.25	0.26
ωB97X-D	cc-pVTZ	−17.69	0.52	−0.22	0.29
B3LYP	cc-pVQZ	−12.91	0.53	−0.23	0.30
	cc-pVTZ	−32.89	0.53	−0.25	0.29

[a]Values are given in units of Debye. [b]Upper and lower bounds correspond to the 2.5 and 97.5% HPD boundaries, respectively.

based on 10 000 random samples drawn from the posterior distribution. The result is shown in Figure 9 with the predicted spectrum (in blue) overlaid with the experimental simulation (in red) up to $J' = 10$. The highest probabilities in the prediction are coincident with the experimental lines at low-J. Comparing the $J = 1 \rightarrow 0$ transition frequencies, the experimental value is 2734.6970 MHz compared to the predicted expectation value $\langle\nu(J' = 1)\rangle$ of 2721.4033 MHz, an offset of −13 MHz. For reference, if the equilibrium values were used without scaling with δ, the transition frequency is overestimated by 28 MHz. Beyond the $J = 3 \rightarrow 2$ transition, the prediction becomes worse relative to the experiment and is shifted to higher frequency owing to the neglect of centrifugal distortion.

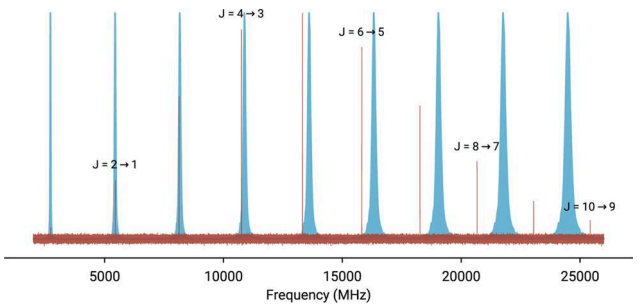


Figure 9. Simulated experimental spectrum (red trace) and predicted (blue regions) a-type rotational spectra for phenylacetylene over the 2−26 GHz range. The predicted blue regions represent 95% HPD for the prediction.

Since the predicted spectrum has a finite width, which is related to its uncertainty, in principle, it can be cross-correlated with an experimental spectrum to systematically and automatically target a molecule, which has yet to be detected—effectively, the matched filter approach used frequently in digital signal processing. The left-most panel in Figure 10 shows the cross-correlation of the two spectra in Figure 9, in which the predicted spectrum has been arbitrarily scaled by an exponential factor that deweights the predicted probability toward higher frequencies; in this instance, we scaled the predicted spectrum by $e^{-\nu k}$, where $k = 3 \times 10^{-4}$, a value tuned empirically as to minimize the contribution from high J transitions ($J' = 4$). The peak of the cross-correlation spectrum determines the optimal overlap between the prediction and the experiment, and as the line widths are relatively narrow, they should contain few distinct maxima. In our example, we obtain a peak shift value of 6.9 MHz for the entire spectrum (left panel, Figure 10), which corresponds quite closely to the −13 MHz offset between the experiment and theory for the $J = 1 \rightarrow 0$ transition. Because the cross-correlation includes contributions from higher-frequency transitions, however, there is a small, but non-negligible difference between the peak shift and one-half the frequency offset (6.5 MHz).

Although the example illustrated in Figure 10 appears somewhat trivial, the primary use of this approach would be instances where the rotational spectrum of multiple species is overlapped (such as the case in astronomical or laboratory spectra). The shift determined by the cross-correlation approach could be used to assess whether the target molecule is present in a mixture and whether the spectroscopic prediction needs to be revisited perhaps with higher-level ab initio methods. With the low-cost methods, we have surveyed

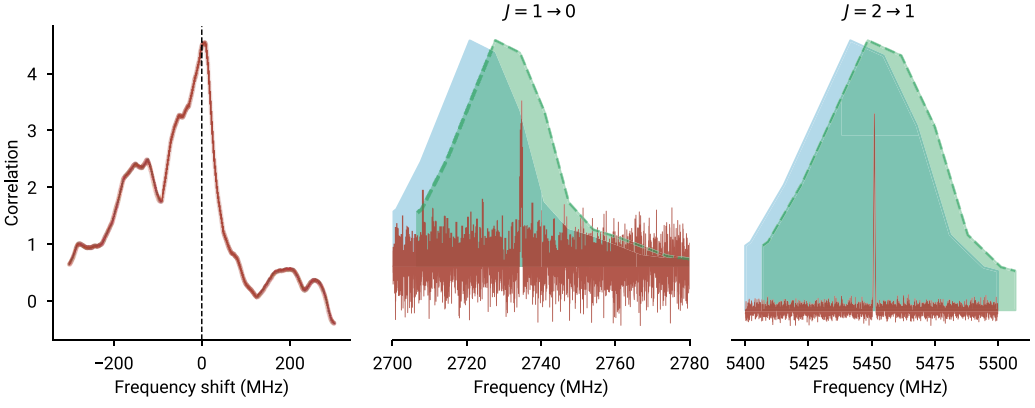


Figure 10. Left panel: cross-correlation spectrum of the predicted and experimental spectra, with a lag step size of 300 kHz. The dotted vertical line marks zero. Center and right panel: the unshifted predicted spectrum (blue region) compared with the same spectrum shifted by the optimal lag (green region, dotted trace) overlaid on top of the experimental spectrum (red) for $J = 1{-}0$ (center) and $J = 2{-}1$ (right).

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

in this paper, geometry optimizations of relatively large species (more than 10 heavy atoms) could be performed quickly and used to rapidly assay broadband spectra. Alternatively, the predicted spectrum could be used to set statistically meaningful constraints on search ranges for deep surveys with cavity-based methods, where frequency coverage and integration dwell time represent competing constraints (owing to discharge stability, sample consumption, etc.). The latter case is particularly relevant for targeted searches of transient species since the laboratory experiment is often a multivariate optimization problem (more often than not solved with stochastic sampling), the approach we have outlined here ensures time-efficient spectral line surveys, allowing more time to be spent on assaying experimental conditions.

CONCLUSIONS

In this work, we have quantified the statistical accuracy and uncertainty of 13 common, low-cost electronic structure methods combined with 7 basis sets, using a representative sample of 76 molecules for which accurately determined gas-phase spectroscopic data is readily available. The results show that commonly used methods in rotational spectroscopy literature, such as MP2 and B3LYP, have higher uncertainties and lower accuracies than density functionals such as PW6B95-D3, ωB97X-D, and M06-2X. For accurate predictions, these functionals should be paired with a quadruple-ζ quality (cc-pVQZ) basis set, which produces the most statistically reliable rotational constants. We also recommend the same functionals, paired with the 6-31+G(d) basis set, which provides comparable performance at a substantial reduction (approx. four times fewer basis functions) in basis set cost, an important factor when considering large molecules with little to no symmetry. In general, basis functions without diffuse/polarization functions should be avoided (e.g., 3-21G and 6-31G), as the performance of these bases is highly nonstatistical, regardless of the method used.

For dipole moments, we find that basis set choice is the largest contributing factor; dipole moments calculated with split-valence basis sets have very high uncertainty; so to achieve satisfactory accuracy, the cc-pVQZ basis is required. With the exception of HF, dipole moment calculations are relatively invariant to the electronic structure method used. Thus, for astronomical modeling studies where the absolute value of the dipole moment is critical, we recommend calculations with correlation-consistent basis sets. For laboratory studies, where only the relative values of dipole moments are often important, the most cost-efficient approach is to employ PW6B95-D3, ωB97X-D, or M06-2X functional with a 6-31+G(d) basis set.

Finally, we have also demonstrated a method of using the Bayesian uncertainties for predicting rotational spectra, which act as a statistical approach to correcting vibration−rotation interaction, as well as providing informed constraints on the frequency ranges needed to identify transitions. By cross-correlating predicted spectra with broadband experimental data, it may be possible to perform computational assays to infer the presence of entirely new molecules. When combined with low-cost quantum chemistry, laboratory and even astronomical rotational spectra might then be rapidly and exhaustively screened for large numbers of calculated molecules in this manner.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.9b09982.

Rotational constants (in megahertz) and dipole moments; summary of prior statistics from the MCMC modeling; full tables of the Bayesian analysis results (PDF)

Theoretical data set used to generate the analysis (CSV)

AUTHOR INFORMATION

Corresponding Author

Kin Long Kelvin Lee − *Harvard & Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0002-1903-9242; Email: kinlee@cfa.harvard.edu

Author

Michael McCarthy − *Harvard & Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0001-9142-0008

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.9b09982

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge financial support from NSF grants AST-1615847 and AST-1908576, NASA grants NNX13AE59G and 80NSSC18K0396, and computing resources from the Smithsonian Institution High Performance Cluster (SI/HPC, "Hydra"). The authors thank Dr. John F. Stanton and Dr. Brooks Pate for helpful suggestions at ISMS.

REFERENCES

(1) Lee, K. L. K.; McCarthy, M. Study of Benzene Fragmentation, Isomerization, and Growth Using Microwave Spectroscopy. *J. Phys. Chem. Lett.* **2019**, *10*, 2408−2413.

(2) Domingos, S. R.; Pérez, C.; Schnell, M. Sensing Chirality with Rotational Spectroscopy. *Annu. Rev. Phys. Chem.* **2018**, *69*, 499−519.

(3) Patterson, D.; Schnell, M.; Doyle, J. M. Enantiomer-specific detection of chiral molecules via microwave spectroscopy. *Nature* **2013**, *497*, 475−477.

(4) Bohn, R. K.; Montgomery, J. A.; Michels, H. H.; Fournier, J. A. Second moments and rotational spectroscopy. *J. Mol. Spectrosc.* **2016**, *325*, 42−49.

(5) Demaison, J.; Boggs, J. E.; Császár, A. G., Eds., *Equilibrium Molecular Structures: From Spectroscopy to Quantum Chemistry*; CRC Press: Boca Raton, 2011.

(6) Lee, K.; McCarthy, M. C. In *Benzene's Inferno, Part II: Automated Analysis and Identification*, 74th International Symposium on Molecular Spectroscopy, June 17−21; International Symposium on Molecular Spectroscopy, 2019.

(7) Crabtree, K. N.; Martin-Drumel, M.-A.; Brown, G. G.; Gaster, S. A.; Hall, T. M.; McCarthy, M. C. Microwave spectral taxonomy: A semi-automated combination of chirped-pulse and cavity Fourier-transform microwave spectroscopy. *J. Chem. Phys.* **2016**, *144*, No. 124201.

(8) Martin-Drumel, M.-A.; McCarthy, M. C.; Patterson, D.; McGuire, B. A.; Crabtree, K. N. Automated microwave double resonance spectroscopy: A tool to identify and characterize chemical compounds. *J. Chem. Phys.* **2016**, *144*, No. 124202.

(9) Zaleski, D. P.; Prozument, K. Automated assignment of rotational spectra using artificial neural networks. *J. Chem. Phys.* **2018**, *149*, No. 104106.
(10) Seifert, N. A.; Finneran, I. A.; Perez, C.; Zaleski, D. P.; Neill, J. L.; Steber, A. L.; Suenram, R. D.; Lesarri, A.; Shipman, S. T.; Pate, B. H. AUTOFIT, an automated fitting tool for broadband rotational spectra, and applications to 1-hexanal. *J. Mol. Spectrosc.* **2015**, *312*, 13−21.
(11) Puzzarini, C.; Stanton, J. F.; Gauss, J. Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy. *Int. Rev. Phys. Chem.* **2010**, *29*, 273−367.
(12) Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theor. Chem. Acc.* **2012**, *131*, No. 1079.
(13) Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. Accurate structure, thermodynamic and spectroscopic parameters from CC and CC/DFT schemes: the challenge of the conformational equilibrium in glycine. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10094−10111.
(14) Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory. *J. Chem. Phys.* **2006**, *125*, No. 044108.
(15) Puzzarini, C. Extrapolation to the Complete Basis Set Limit of Structural Parameters: Comparison of Different Approaches. *J. Phys. Chem. A* **2009**, *113*, 14530−14535.
(16) Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.* **2004**, *122*, No. 014108.
(17) Roy, T. K.; Gerber, R. B. Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications. *Phys. Chem. Chem. Phys.* **2013**, *15*, 9468−9492.
(18) Barone, V.; Biczysko, M.; Puzzarini, C. Quantum Chemistry Meets Spectroscopy for Astrochemistry: Increasing Complexity toward Prebiotic Molecules. *Acc. Chem. Res.* **2015**, *48*, 1413−1422.
(19) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.* **1995**, *103*, 4572−4585.
(20) Pawlowski, F.; Jørgensen, P.; Olsen, J.; Hegelund, F.; Helgaker, T.; Gauss, J.; Bak, K. L.; Stanton, J. F. Molecular equilibrium structures from experimental rotational constants and calculated vibration-rotation interaction constants. *J. Chem. Phys.* **2002**, *116*, 6482−6496.
(21) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.
(22) Pople, J. A.; Scott, A. P.; Wong, M. W.; Radom, L. Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies from HF/6-31G* and MP2/6-31G* Harmonic Frequencies. *Isr. J. Chem.* **1993**, *33*, 345−350.
(23) Martin, J. M. L.; El-Yazal, J.; François, J.-P. Basis set convergence and performance of density functional theory including exact exchange contributions for geometries and harmonic frequencies. *Mol. Phys.* **1995**, *86*, 1437−1450.
(24) Spackman, P. R.; Jayatilaka, D.; Karton, A. Basis set convergence of CCSD(T) equilibrium geometries using a large and diverse set of molecular structures. *J. Chem. Phys.* **2016**, *145*, No. 104101.
(25) Bak, K. L.; Gauss, J.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Stanton, J. F. The accurate determination of molecular equilibrium structures. *J. Chem. Phys.* **2001**, *114*, 6548−6556.
(26) Coriani, S.; Marchesan, D.; Gauss, J.; Hättig, C.; Helgaker, T.; Jørgensen, P. The accuracy of ab initio molecular geometries for systems containing second-row atoms. *J. Chem. Phys.* **2005**, *123*, No. 184107.
(27) Jensen, F. Polarization consistent basis sets. IV. The basis set convergence of equilibrium geometries, harmonic vibrational frequencies, and intensities. *J. Chem. Phys.* **2003**, *118*, No. 2459.
(28) Grimme, S.; Steinmetz, M. Effects of London dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031−16042.
(29) Ray, B. S. Über die Eigenwerte des asymmetrischen Kreisels. *Z. Phys.* **1932**, *78*, 74−91.
(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.. et al.*Gaussian 16*, *revision A.01*, 2016.
(31) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622.
(32) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(33) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554−8560.
(34) 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.
(35) Zhao, Y.; Truhlar, D. G. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.
(36) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.
(37) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.
(38) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.
(39) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* basis set for atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223−1229.
(40) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265−3269.
(41) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(42) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358−1371.
(43) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(44) Hall, G. G. Lennard-Jones John Edward, The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials. *Proc. R. Soc. Lond. A* **1951**, *205*, 541−552.
(45) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(46) 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.
(47) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(48) Handy, N. C.; Cohen, A. J. Left-right correlation energy. *Mol. Phys.* **2001**, *99*, 403−412.
(49) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX

(50) 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. *J. Chem. Phys.* **2010**, *132*, No. 154104.
(51) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, No. 33.
(52) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(53) Gräfenstein, J.; Cremer, D. Can density functional theory describe multi-reference systems? Investigation of carbenes and organic biradicals. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2091−2103.
(54) 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*, No. 114104.
(55) Salvatier, J.; Wiecki, T. V.; Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2016**, *2*, No. e55.
(56) Homan, M. D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593−1623.
(57) Geweke, J. *Bayesian Statistics*; Oxford University Press, 1992; pp 169−193.
(58) Gelman, A.; Rubin, D. B. *Bayesian Statistics*; Oxford University Press, 1992; pp 625−631.
(59) Puzzarini, C. Rotational spectroscopy meets theory. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6595−6607.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was originally published ASAP on January 28, 2020. Due to a production error, eq 3 was incorrect. The revised version was reposted on January 29, 2020.

https://dx.doi.org/10.1021/acs.jpca.9b09982
J. Phys. Chem. A XXXX, XXX, XXX−XXX