JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

pubs.acs.org/JCTC

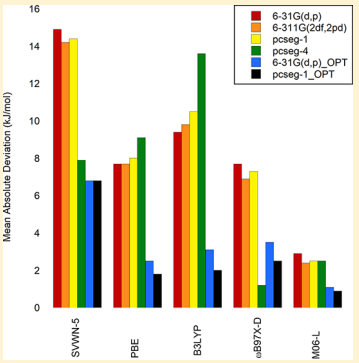
Method Calibration or Data Fitting?

Frank Jensen*

Department of Chemistry, Aarhus University, DK-8000 Aarhus, Denmark

Supporting Information

ABSTRACT: We investigate by explicit parameter optimization to what extent basis sets of polarized double-ζ quality can introduce compensating errors in five different density functional methods. It is shown that minor changes in the contraction coefficients of the valence functions in the basis sets can have a significant impact and allow different density functional methods to achieve very similar performances. This holds for nuclear magnetic shielding constants and for isomerization energies, barrier heights, and noncovalent interactions. It is furthermore shown that errors due to neglect of vibrational and solvent effects can be absorbed in the combined method and basis set errors. These findings hold for data sets consisting of 50−150 data points. This raises the question of whether the common practice of identifying combinations of density functional methods and basis sets that have a good performance against a selected set of reference data should be considered as data fitting in the combined parameter space spanned by the method and basis set.



INTRODUCTION

The inference of molecular structure from spectroscopic data is traditionally done by empirical correlations, but electronic structure calculations are increasingly used for predicting spectra for possible molecular structures and assigning the structure by comparison of calculated and observed spectra.[1−5] The success of the latter approach is critically dependent on the accuracy of the calculated spectra, and method calibration is therefore an essential component. Two flavors of method calibration are in common use, a "purist" and an "application" approach. In the purist approach, the theoretical model is systematically improved with respect to electron correlation, basis sets, vibrational, relativistic, solvent, etc. effects, with the aim of assessing the importance of each component and, thus, establishing the required theoretical level for a given target accuracy.[6−11] The computational requirements in this approach are often high and dictate that only a limited number of relatively small molecules can be used as the reference data, and the number of reference systems is often further restricted by the requirement of accurate experimental results. The application approach aims at identifying computationally efficient methods that for a given class of molecules can be used as a general tool for correlating structure and spectra, and this in addition often employs explicit linear correlation methods for reducing systematic errors.[12] The requirement of computational efficiency is rooted in the desire to handle actual experimental problems, which often involve large molecules. A popular venture in this category is searching for combinations of density functional theory (DFT) methods and medium sized basis sets that minimize the error relative to a selected set of experimental data and identify specific functional and basis set combinations as general purpose tools for correlating molecular structure and spectra.[1,12−24] It is recognized that well-performing combinations often benefit from error cancellation between the different components,[25] and automated methods have been proposed for identifying low-error combinations.[26,27]

A specific example that has attracted significant attention in recent years is the use of calculated NMR chemical shifts for structure identification.[28,29] Correlating calculated and experimental ^{1}H and ^{13}C chemical shifts has for example been used for distinguishing diastereomeric isomers.[29−36] The target accuracy is in these cases ∼0.1 ppm for ^{1}H and ∼1 ppm for ^{13}C, and this is difficult to achieve by standard computational methods. The DP4[37−39] and CHESHIRE[1,32,40] methodologies have been developed with the aim of providing general tools for differentiating between different possible structural candidates, and these methods rely on calibration against an experimental set of ^{1}H and ^{13}C chemical shifts. A large number of theoretical models have been examined, with slightly different performance.[19,40] We have in other work developed a family of basis sets specifically for calculating nuclear magnetic shielding constants, denoted pcSseg-*n*,[41] but these often perform worse than standard basis sets, which indicates that method and basis set error cancellation plays an important role. This raises the question whether the "application" approach of benchmarking perhaps should be viewed as data fitting in the combined parameter space of method and basis set variables, and this is the issue we address in the present work. We will focus on calculated nuclear magnetic shielding constants but show that the findings carry over to energetic features and likely to other properties as well.

COMPUTATIONAL DETAILS

We have used five representative DFT methods, the SVWN5[42] local density, the PBE[43] and M06-L[44] gradient corrected, the

Received: May 22, 2018
Published: July 27, 2018

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Benchmark Reference Nuclear Magnetic Shielding Constants (ppm)[a]

	^{1}H		^{13}C		^{15}N		^{17}O		^{19}F	
molecule	*Theo*	*Exp*	*Theo*	*Exp*	*Theo*	*Exp*	*Theo*	*Exp*	*Theo*	*Exp*
C_2H_4	26.05	25.43	69.71	64.4						
C_3H_4 (cyclopropen)	24.37	24.0	192.10	190.4						
	30.64	30.1	83.69	84						
H_2CO	21.99	18.3	1.53	−0.5			−378.61	−427		
CH_3F	27.35	26.6	122.15	116.7					482.88	470.6
CH_4	31.30	30.61	198.93	195.0						
CO			2.24	0.9			−55.05	−62.74		
FCCH	30.49	31.9	100.06	93.9					423.55	446.05
			179.86	168.9						
FCN			82.24	(80.80)	117.89	(109.89)			374.10	344.70
H_2CCO	29.19	31.29	193.32	184.5			−5.92	(−10.87)		
			−6.34	−7.0						
H_2O	30.65	30.05				338.01	323.6			
C_2H_4O (oxirane)	29.14	30.95	153.20	147.9			363.23	336.5		
HCN	29.01	27.78	84.58	82.0	−14.11	−20.4				
HFCO	23.86	(23.38)	39.63	(37.20)			−94.33	(−107.53)	165.27	147.7
HF	28.83	28.51							420.31	409.6
HOF	19.57	18.51					−68.92	(−101.44)	192.21	169.6
N_2O					106.45	99.5	199.02	178.3		
					12.56	11.3				
N_2					−60.43	−61.6				
NH_3	31.44	30.68			270.66	264.54				
OF_2							−447.09	−495.3	−23.95	−60.3
F_2									−192.76	−233.2
CF_4			65.94	64.6					267.58	258.6

[a] *Theo* indicates fixed nuclei CCSD(T) results using aug-cc-pCV[TQ]Z basis set extrapolated values,[53] except for F_2 and CF_4 which are CCSD(T)/pcSseg-4 values.[54] *Exp* indicates experimental values with pseudo-experimental values in parentheses being the *Theo* result with added calculated vibrational corrections.

B3LYP[45,46] hybrid, and the ωB97X-D[47] range separated functionals, where the latter includes an empirical dispersion correction. The basis sets employed are the Pople style 6-31G(d,p)[48] and 6-311G(2df,2pd),[49] as well as the pcseg-*n*[50] and pcSseg-*n*[41] basis sets, where pcseg-*n* are optimized for DFT energetic features while pcSseg-*n* are optimized for calculating nuclear magnetic shielding constants. Molecular geometries have been taken from the respective benchmarks as given in the references for each case. Parameter optimizations have been done using a pseudo-Newton−Raphson algorithm in connection with gradients of the error function with respect to basis set contraction coefficients generated by numerical differentiation, using the gauge including atomic orbital (GIAO) procedure implemented in the Gaussian 09 program package for nuclear magnetic shielding constants.[51] The error function for the data set in Tables 2 and 3 has been defined as the root mean squared (RMS) error over molecules, where the error for each molecule is the RMS of all the chemical shielding constants for the given nucleus. Note that the minimization is of the RMS error while the errors reported in the Tables are mean absolute deviation (MAD) values. It is thus possible that the MAD can increase when the RMS error decreases, as is observed for the ^{1}H MAD values with the SVWN5/pcSseg-1 combination in Table 3. The error function for the CHESHIRE data set (Tables 4 and 5) has been defined as the RMS error for the best linear correlation over chemical shifts for the molecules, where the nuclear shielding constants are averaged over nuclei that are conformationally interchangeable. The error function for the relative energies in Tables 6−8 has been defined as the RMS error relative to the reference results.

RESULTS AND DISCUSSION

We investigate the performance of five representative DFT methods, the SVWN5[42] local density, the PBE[43] and M06-L[44] gradient corrected, the B3LYP[45,46] hybrid, and the ωB97X-D[47] range separated functionals. For calculating nuclear magnetic shielding constants these functionals are used in combination with the Pople style 6-31G(d,p)[48] and 6-311G(2df,2pd),[49] as well as the pcSseg-1 and aug-pcSseg-4[41] basis sets, where pcSseg-1 is comparable in size to 6-31G(d,p) while aug-pcSseg-4 provides results essentially at the basis set limit. The 6-31G(d,p) is 3s2p1d/2s1p in composition (10s4p1d/4s1p in terms of primitive functions), while the pcSseg-1 is 3s3p1d/2s1p (8s5p1d/4s2p in terms of primitive functions). The pcSseg-*n* family of basis sets differs from the pcseg-*n* energy optimized basis[50] sets by inclusion of a p-function with large exponent and less contraction of the p-functions. The pcSseg-*n* basis sets have in previous work been shown to provide a systematic and stable convergence toward the basis set limit, and the pcSseg-1 provides significantly lower basis set errors than other basis sets of DZP quality.[41,52] We will in the following only discuss the isotropic nuclear magnetic shielding, since this is the commonly observed quantity experimentally. For the relative energies, the regular pcseg-*n* energy optimized basis sets[50] are used, with the pcseg-1 having the composition 3s2p1d/2s1p (7s4p1d/4s1p in terms of primitive functions).

Calibration against Absolute Nuclear Magnetic Shielding Constants. Teale et al.[53] and Stoychev et al.[54]

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Mean Absolute Deviations of 15 H and 40 Non-H Nuclear Magnetic Shielding Constants against Non-Relativistic Fixed Nucleus Values for the Molecules in Table 1 (ppm)[a]

	H						C, N, O, F					
XC	*Standard*				*Optimized*		*Standard*				*Optimized*	
Basis set	6-31G(d,p)	6-311G(2df,2pd)	pcSseg-1	aug-pcSseg-4	6-31G(d,p)	pcSseg-1	6-31G(d,p)	6-311G(2df,2pd)	pcSseg-1	aug-pceg-4	6-31G(d,p)	pcSseg-1
SVWN5	0.48	0.49	0.46	0.59	0.48	0.44	15.2	32.3	38.4	38.8	12.9	16.0
PBE	0.62	0.48	0.36	0.41	0.52	0.34	15.5	25.3	32.8	32.8	12.0	12.9
B3LYP	0.60	0.41	0.27	0.27	0.36	0.18	9.8	24.3	31.8	31.2	7.4	7.7
ωB97X-D	0.51	0.36	0.23	0.24	0.30	0.16	8.8	19.1	26.1	26.7	5.7	6.7
M06-L	0.72	0.54	0.45	0.32	0.45	0.21	22.1	12.6	14.1	18.4	13.1	11.8

[a]Optimized indicates that valence p-contraction (s-contraction for H) coefficients have been optimized against theoretical reference results, two parameters for each atom for 6-31G(d,p) and pcSseg-1.

have calculated CCSD(T) nuclear magnetic shielding constants for the systems in Table 1 using basis sets close to the basis set limit (aug-cc-pCV[TQ]Z extrapolated in ref 53 and pcSseg-4 in ref 54). These data provide theoretical reference values for ^{1}H, ^{13}C, ^{15}N, ^{17}O, and ^{19}F absolute shielding constants at the nonrelativistic, fixed nuclei approximation, and accurate experimental data are also available for most of the nuclei. We have assigned pseudoexperimental results for the few missing nuclei by addition of the vibrational corrections from the work of Teale et al.[53] for use as discussed below; these are listed in parentheses in Table 1.

Table 2 shows the mean absolute deviations (MAD) relative to the best theoretical estimates (Table 1) with combinations of functionals and basis sets, where the SVWN5, PBE, B3LYP, ωB97X-D, and M06-L order reflects the performance for non-hydrogen nuclei at the basis set limit, represented by the aug-pcSseg-4 results. The ^{1}H shielding constants have a smaller range (20−32 ppm) than the non-hydrogen nuclei (−450 to +500 ppm) and are treated separately. The C, N, O, and F results in Table 2 are represented graphically in the left-hand panel of Figure 1. The pcSseg-1 results under the header *Standard* are in general close to the aug-pcSseg-4 results, which confirms the previous finding that the pcSseg-1 and pcSseg-2 basis sets have low basis set errors and will often be sufficiently accurate for routine uses.[41] Focusing on the non-hydrogen results, Table 2 shows that the performance of the SVWN5, PBE, B3LYP, and ωB97X-D functionals deteriorates when the basis set quality is increased, while the M06-L is the best performing at the basis set limit. The good performance of M06-L when using large basis sets has been noted in other work as well.[54,55] The lowest MAD values are obtained with the ωB97X-D and B3LYP functionals and the 6-31G(d,p) basis set (9 and 10 ppm, respectively), and these combinations would in an application type calibration then be the recommended methodologies.

The difference between the results with the 6-31G(d,p) and aug-pcSseg-4 basis sets is a measure of basis set errors canceling functional errors and amounts to ∼20 ppm for some of the functionals. If the basis set exponents and contraction coefficients are considered as parameters, we can ask how these

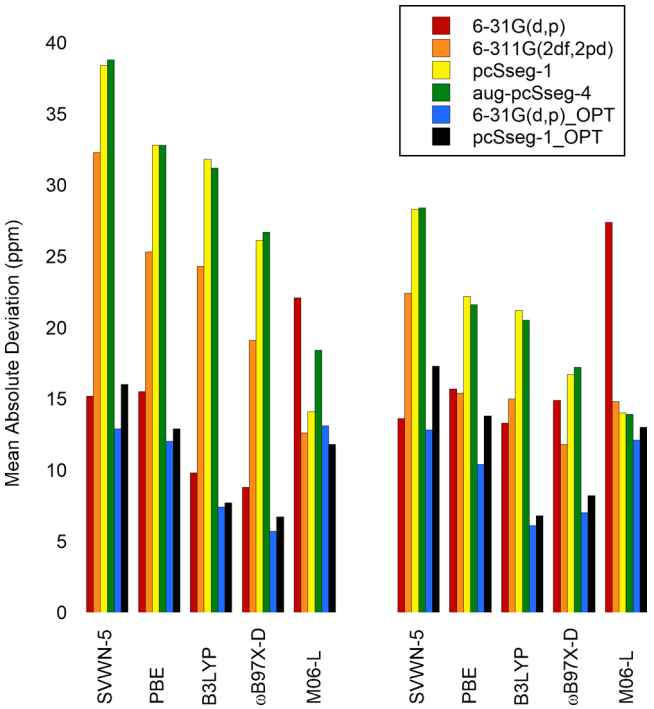


Figure 1. Graphical representation of the C, N, O, and F magnetic shielding constant results in Tables 2 and 3, with mean absolute deviations (ppm) from the CCSD(T) (Table 2, left set of graphs) and experimental (Table 3, right set of graphs) values in Table 1 as a function of DFT and basis set combinations.

parameters can be tweaked to improve the performance. Inspection of the gradient of an error function with respect to the basis set parameters indicated that the contraction coefficients for the valence p-function were the most important, and since the number of data points (15, 16, 6, 9, and 9 for H, C, N, O, and F) is relatively small, only those were selected for optimization. The results in Table 2 under the header *Optimized* are obtained by minimizing the root-mean-squared (RMS) deviation against the reference values with respect to the two p-contraction coefficients in the 6-31G(d,p)

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. Mean Absolute Deviations of 15 H and 40 Non-H Nuclear Magnetic Shielding Constants against Experimental Values for the Molecules in Table 1 (ppm)[a]

	H						C, N, O, F					
XC	*Standard*				*Optimized*		*Standard*				*Optimized*	
Basis set	6-31G(d,p)	6-311G(2df,2pd)	pcSseg-1	aug-pcSseg-4	6-31G(d,p)	pcSseg-1	6-31G(d,p)	6-311G(2df,2pd)	pcSseg-1	aug-pcSseg-4	6-31G(d,p)	pcSseg-1
SVWN5	1.06	0.99	0.92	0.88	0.98	0.93	13.6	22.4	28.3	28.4	12.8	17.3
PBE	1.38	1.22	1.17	1.00	1.03	0.99	15.7	15.4	22.2	21.6	10.4	13.8
B3LYP	1.43	1.28	1.22	1.05	0.97	0.94	13.3	15.0	21.2	20.5	6.1	6.8
ωB97X-D	1.39	1.25	1.20	1.05	0.97	0.94	14.9	11.8	16.7	17.2	7.0	8.2
M06-L	1.49	1.36	1.37	1.16	0.98	1.04	27.4	14.8	14.0	13.9	12.1	13.0

[a]Optimized indicates that valence p-contraction (s-contraction for H) coefficients have been optimized against experimental reference results, two parameters for each atom for 6-31G(d,p) and pcSseg-1.

and pcSseg-1 basis sets (and the corresponding two s-contraction coefficients for hydrogen). The nuclear magnetic shielding constant is a sufficiently local property that the optimization can be done atom-wise; i.e., the optimization of the carbon contraction coefficients are done only against the ^{13}C shielding constants etc. The hydrogen MAD values in Table 2 are obtained using the optimized hydrogen basis set and standard basis sets for C, N, O, and F, while the C, N, O, and F MAD values are obtained by combining the individually optimized C, N, O, and F basis sets with the standard basis set for H. The last two columns in Table 2 show that the 6-31G(d,p) and pcSseg-1 basis sets with optimized p-contraction coefficients perform almost identically with MAD values of ∼7 ppm for the B3LYP and ωB97X-D functionals. The 6-31G(d,p) and pcSseg-1 reductions of ∼3 and ∼20 ppm upon parameter optimization to the same final MAD values suggest that the 6-31G(d,p) basis set parameters simply are closer to the optimum error canceling values from the onset.

Table 3 shows the corresponding results where the performance is measured against the experimental nuclear magnetic shielding constants (Table 1), with the C, N, O, and F results represented graphically in the right-hand panel of Figure 1. The experimental values differ from the CCSD(T) values used as references in Table 2 by inclusion of vibrational effects, residual electron correlation (beyond CCSD(T)), and relativistic corrections, of which the vibrational corrections are the most important. We have generated pseudoexperimental values for the few nuclei (Table 1) where experimental results are missing by addition of the calculated vibrational corrections to the CCSD(T) results, as given by Teale et al.[53] The MAD between the experimental and theoretical reference values in Tables 2 and 3 is 1.10 ppm for H and 12.7 ppm for C, N, O, and F, and these values are thus a measure of effects ignored by the nonrelativistic fixed nucleus CCSD(T) model. Table 3 shows that the performance against hydrogen shielding constants deteriorates, which is not surprising since these are significantly affected by vibrational corrections. The performance for non-hydrogen shielding constants deteriorates for some method/basis set combinations but improves for other combinations, and the B3LYP/6-31G(d,p) combination again emerges as one of the best performing methodologies with a MAD of 13 ppm. Optimizing the two p-contraction coefficients for either the 6-31G(d,p) or pcSseg-1 basis sets against the experimental values (last two columns in Table 3, MAD values of ∼7 ppm for the B3LYP and ωB97X-D functionals) leads to almost identical performance as optimizing against the theoretical reference (last two columns in Table 2). This shows that even a simple two-parameter tuning of the basis sets introduces enough flexibility to accommodate changes due to, e.g., vibrational effects, thus effectively absorbing the 12.7 ppm average difference between the two sets of reference data. The 6-31G(d,p) and pcSseg-1 error reductions of ∼8 and ∼9 ppm upon parameter optimization for the ωB97X-D functional to essentially the same final MAD values suggest that, for this set of reference data, the original two sets of basis set parameters are equally poor/good with respect to canceling errors in the functional.

From a QSAR perspective, a specific method and basis set combination can be considered as producing a descriptive vector that can be used for modeling the reference data, and a QSAR model can be obtained, for example, by multivariable linear correlation. Using only results from the nonoptimized 6-31G(d,p) and pcSseg-1 basis sets, the best single descriptive vector for predicting the ^{1}H shielding reference data in Table 2 is ωB97X-D/pcSseg-1 with a MAD of 0.23 ppm. A linear combination of the PBE/6-31G(d,p) and B3LYP/6-31G(d,p) results, however, can achieve a MAD of 0.12 ppm, and including also a descriptive variable in terms of the SVWN5/pcSseg-1 results allows construction of a three-component QSAR model with a MAD of 0.08 ppm. A linear combination of the PBE/6-31G(d,p), B3LYP/6-31G(d,p) and M06-L/pcSseg-1 results can reduce the MAD to 4.8 ppm for the C, N, O, and F data in Table 2, and the combination of B3LYP/6-31G(d,p) and ωB97X-D/pcSseg-1 results has a MAD of 6.5 ppm against the experimental results in Table 3. One can easily envision machine learning models operating on a larger variety of descriptive vectors that could achieve even lower errors.[56,57]

Calibration against Experimental Chemical Shifts. The application type calibration utilizes experimental data as references, and these include finite temperature and solvation

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Mean Absolute Deviations against 167 Experimental ^{13}C Chemical Shifts Taken from the CHESHIRE Data Set (ppm)[40]

		functional				
	basis	SVWN5	PBE	B3LYP	ωB97X-D	M06-L
vs TMS[a]	6-31G(d,p)	3.31	5.33	4.20	2.74	8.24
	pcSseg-1	7.21	4.08	5.28	5.19	3.43
	pcSseg-4	10.25	7.22	8.16	8.12	3.90
Lin. Corr.[b]	6-31G(d,p)	2.04	2.09	1.63	1.34	2.15
	pcSseg-1	2.01	1.87	1.54	1.37	1.71
	pcSseg-4	2.13	2.09	1.67	1.36	1.85
Lin. Corr. Opt[c]	6-31G(d,p)	1.74	1.59	1.23	1.12	1.52
	pcSseg-1	1.90	1.60	1.29	1.15	1.47

[a]"vs TMS" indicates that calculated nuclear magnetic shielding constants are converted to chemical shifts by subtracting the TMS value calculated with the same method and basis set. [b]"Lin. Corr." indicates that mean absolute deviations are obtained from the best linear correlation model between calculated nuclear magnetic shielding constants and experimental chemical shifts. [c]"Lin. Corr. Opt." indicates that mean absolute deviations are obtained from the best linear correlation model between calculated nuclear magnetic shielding constants and experimental chemical shifts, where valence contraction coefficients for C, N, and O have been optimized to give best linear correlation, four parameters for 6-31G(d,p), and three parameters for pcSseg-1.

Table 5. Mean Absolute Deviations against 134 Experimental ^{1}H Chemical Shifts Taken from the CHESHIRE Data Set (ppm)[40]

		functional				
	basis	SVWN5	PBE	B3LYP	ωB97X-D	M06-L
vs TMS[a]	6-31G(d,p)	0.20	0.17	0.15	0.19	0.17
	pcSseg-1	0.20	0.17	0.15	0.19	0.16
	pcSseg-4	0.27	0.23	0.19	0.22	0.24
Lin. Corr.[b]	6-31G(d,p)	0.16	0.15	0.13	0.13	0.14
	pcSseg-1	0.15	0.14	0.13	0.14	0.14
	pcSseg-4	0.14	0.14	0.12	0.13	0.13
Lin. Corr. Opt.[c]	6-31G(d,p)	0.16	0.15	0.13	0.12	0.13
	pcSseg-1	0.14	0.14	0.12	0.13	0.13

[a]"vs TMS" indicates that calculated nuclear magnetic shielding constants are converted to chemical shifts by subtracting the TMS value calculated with the same method and basis set. [b]"Lin. Corr." indicates that mean absolute deviations are obtained from the best linear correlation model between calculated nuclear magnetic shielding constants and experimental chemical shifts. [c]"Lin. Corr. Opt." indicates that mean absolute deviations are obtained from the best linear correlation model between calculated nuclear magnetic shielding constants and experimental chemical shifts, where valence contraction coefficients for hydrogen have been optimized to give best linear correlation and two parameters for 6-31G(d,p) and pcSseg-1.

effects, as well as potentially being averages over multiple conformations. The calculated results are typically obtained at a single geometry, and solvent effects may or may not be included by a continuum model. The choice of geometry is a further variable, and both force field and quantum mechanical optimized structures have been employed.[38,58] Conformationally mobile systems require a Boltzmann averaging over thermally accessible structures,[11] and this is sensitive to the accuracy for calculating relative conformational energies. Experimental chemical shifts are defined relative to a standard reference compound, which for ^{1}H and ^{13}C is tetramethylsilane (TMS), and calculated absolute shielding constants can be converted to chemical shifts by subtracting the corresponding TMS values calculated at the same theoretical level. Methodologies that utilize linear correlations between calculated and experimental results often perform the correlation between calculated absolute shielding constants and experimental chemical shifts, since the TMS value is absorbed in the correlation parameters. The combination of method, basis set, and solvation model for calculating the magnetic shielding constant and method, basis set, and solvation model or a force field method for calculating the geometry leads to a very large number of combinatorial possibilities. Further variations can be introduced based on choices for the grid employed for calculating the exchange-correlation part and the choice of method for handling the gauge dependence of the nuclear magnetic shielding constant. All of these variables can be considered to span a parameter space in which a search can be performed for identifying error canceling combinations.

Tables 4 and 5 show the performance against the CHESHIRE reference set of data (167 ^{13}C and 134 ^{1}H experimental chemical shifts for a collection of small and medium sized organic compounds, omitting Cl-substituted systems, as they are known to be poorly predicted due to relativistic effects).[1,32,40] The first three rows are the MAD values when calculated nuclear magnetic shielding constants are converted into chemical shifts by subtracting the corresponding calculated TMS value, while the next three rows are the values obtained when performing a linear correlation of the absolute magnetic shielding constants against the experimental chemical shifts. The ωB97X-D/6-31G(d,p) combination is the best for predicting ^{13}C chemical shifts when using the TMS reference with a MAD value of 2.7 ppm, while the corresponding value with the pcSseg-4 basis set is 8.1 ppm. The M06-L functional again emerges as the best performing at the basis set limit (pcSseg-4) for ^{13}C chemical

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 6. Mean Absolute Deviations against 34 Experimental Values for Molecular Isomerization Energies Taken from the ISO34 Data Set (kJ/mol)[61,62a]

basis		functional				
		SVWN5	PBE	B3LYP	ωB97X-D	M06-L
6-31G(d,p)	standard	14.5	11.7	9.3	8.3	9.9
6-311G(2df,2pd)	standard	11.1	8.7	9.4	5.7	8.1
pcseg-1	standard	10.9	8.3	10.0	5.4	7.6
pcseg-4	standard	11.3	8.4	8.9	6.5	8.8
6-31G(d,p)	optimized	7.1	2.9	3.4	2.3	5.3
pcseg-1	optimized	5.9	3.1	3.5	3.0	4.7

[a]Optimized indicates that valence contraction coefficients have been optimized against the RMS deviation from experimental reference results, four parameters for each atom for 6-31G(d,p) (two for H) and three parameters for each atom for pcseg-1 (two for H).

shifts.[54,55] Employing a linear regression reduces the MAD of all the DFT/basis set combinations to values close to 2 ppm, with ωB97X-D/6-31G(d,p) achieving the lowest value of 1.34 ppm. Performing an optimization of the C, N, and O valence s- and p-contraction coefficients (four and three parameters for 6-31G(d,p) and pcSseg-1, respectively, where s-contraction coefficients were included since the data set is substantially larger than in Table 1) allows the MAD to be further reduced to 1.12 ppm, and all of the five DFT methods with either the 6-31G(d,p) or pcSseg-1 basis sets now display very similar performances with MAD values below 2 ppm. Table 5 shows corresponding results for ^{1}H chemical shifts, with similar trends. The conclusion is that each DFT/basis set combination has significant systematic errors that to a large extent can be removed by a linear correlation, but even for this fairly large set of reference data, there is still room for improved error cancellation by optimization of basis set parameters. While the error reduction of ∼0.2−0.5 ppm by parameter optimization (Table 4) is small in absolute terms, it is of the same magnitude as observed between different method/basis set combinations and is achieved by tuning only four/three basis set parameters against a data set consisting of 167 data points. The CHESHIRE protocol recommends testing the correlation models against a second set of compounds that has not been included in the fitting procedure (78 ^{13}C and 61 ^{1}H chemical shifts for the present case). The performance against this second set of data is for the present selection of methods slightly worse than for the fitting set of data, by both standard and optimized basis sets, and the basis set tweaking thus neither improves nor deteriorates the performance.

Error Cancellation for Energetic Results. Most basis sets have been designed using energetic criteria,[58−60] and it is perhaps not surprising that using these for calculating molecular properties can result in basis set errors that can be used for error compensation. This section, however, shows that the same effects can be observed for relative molecular energies, where basis sets optimized for reproducing atomic energies could be expected to have much smaller errors.

Table 6 shows the results for the ISO34 data set consisting of 34 experimental isomerization reaction energies of medium sized H, C, N, and O organic molecules, and the results are represented graphically in the left-hand panel of Figure 2.[61,62] Using the standard 6-31G(d,p), 6-311(2df,2pd), pcseg-1, and pcseg-4 basis sets, where the latter for practical purposes can be considered as the basis set limit, the ωB97X-D with either the 6-311G(2df,2pd) or pcseg-1 basis set emerges as the best performing combinations with MAD values of ∼6 kJ/mol. Using the valence contraction coefficients as free parameters (four parameters for 6-31G(d,p) and three for pcseg-1 for each non-hydrogen atom and two parameters for hydrogen) allows the MAD to be reduced to ∼3 kJ/mol, and the PBE, B3LYP, and ωB97X-D functionals now perform almost identically. We note that the reference data in this case, analogous to the CHESHIRE NMR data set, include finite temperature effects, while the theoretical data are relative energies at a single geometry.

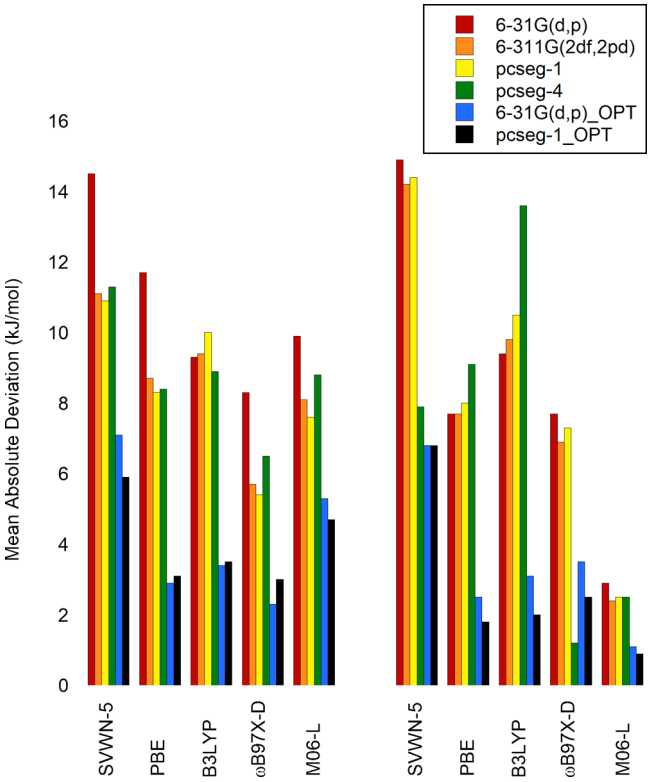


Figure 2. Graphical representation of the relative energy results in Tables 6 and 8, with mean absolute deviations (kJ/mol) for the ISO34 benchmark (Table 6) on the left set of graphs and for the S66 benchmark (Table 8) on the right set of graphs as a function of DFT and basis set combinations.

Table 7 shows corresponding results for the 43 unique activation energies for hydrogen transfer involving C, N, O, and F atoms in the BH76 data set[63,64] where the anionic systems have been omitted. The latter systems require diffuse basis functions for an adequate description and furthermore have no well-defined basis set limit for some DFT functionals.[65] The reference values have been derived from high-level theoretical methods, and being relative energies at fixed geometries, are directly comparable with the DFT results. It is

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 7. Mean Absolute Deviations against 43 Unique Theoretical Values for Barrier Heights in Proton Transfer Reactions of Non-Charged Systems Involving C, N, O, and F Atoms Taken from the BH76 Data Set (kJ/mol)[63,64a]

basis		functional				
		SVWN5	PBE	B3LYP	ωB97X-D	M06-L
6-31G(d,p)	standard	74.5	44.5	25.3	14.8	20.1
6-311G(2df,2pd)	standard	72.3	42.5	22.5	12.3	18.6
pcseg-1	standard	73.6	43.8	23.7	13.9	20.0
pcseg-4	standard	69.9	39.7	19.7	9.6	16.1
6-31G(d,p)	optimized	59.3	32.4	17.2	7.8	10.0
pcseg-1	optimized	55.6	28.9	16.3	8.0	13.0

[a]Optimized indicates that valence contraction coefficients have been optimized against the RMS deviation from high-level theoretical reference results, four parameters for each atom for 6-31G(d,p) (two for H), and three parameters for each atom for pcseg-1 (two for H).

Table 8. Mean Absolute Deviations of 66 Theoretical Values for Non-Covalent Interaction Energies Taken from the S66 Data Set (kJ/mol)[66a]

basis		functional				
		SVWN5	PBE	B3LYP	ωB97X-D	M06-L
6-31G(d,p)	standard	14.9	7.7	9.4	7.7	2.9
6-311G(2df,2pd)	standard	14.2	7.7	9.8	6.9	2.4
pcseg-1	standard	14.4	8.0	10.5	7.3	2.5
pcseg-4	standard	7.9	9.1	13.6	1.2	2.5
6-31G(d,p)	optimized	6.8	2.5	3.1	3.5	1.1
pcseg-1	optimized	6.8	1.8	2.0	2.5	0.9

[a]Optimized indicates that valence contraction coefficients have been optimized against the RMS deviation from high-level theoretical reference results, four parameters for each atom for 6-31G(d,p) (two for H), and three parameters for each atom for pcseg-1 (two for H).

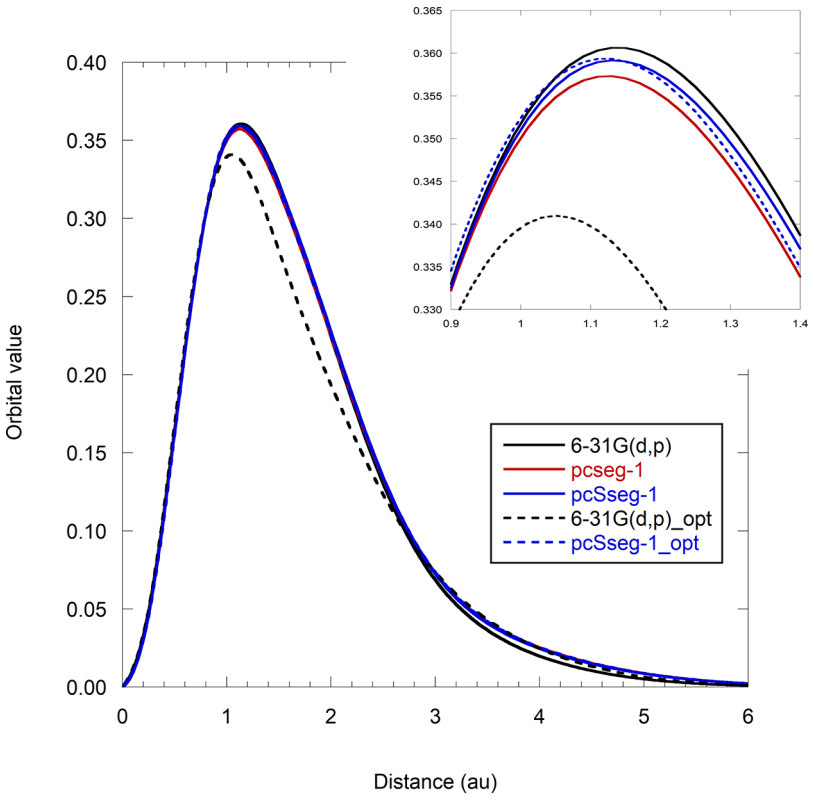


Figure 3. Radial behavior of the atomic p-orbital for carbon with the ωB97X-D functional. 6-31G(d,p), pcseg-1, and pcSseg-1 denote the regular basis sets, while 6-31G(d,p)_opt and pcSseg-1_opt correspond to basis sets derived by optimizing the p-contraction coefficients against the reference results in Table 3.

again seen that an optimization of the valence contraction coefficients of the 6-31G(d,p) and pcseg-1 basis sets significantly reduces the MAD values, in some cases by almost a factor of 2.

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 8 shows results for the S66 data set containing 66 noncovalent interaction energies with reference values provided by high-level theoretical methods, and the results are represented graphically in the right-hand panel of Figure 2.[66] The M06-L/6-31G(d,p) combination produces a MAD value of 2.9 kJ/mol, and only the large pcseg-4 basis set with the ωB97X-D functional is substantially better with a MAD of 1.2 kJ/mol. Optimization of the valence contraction coefficients of the 6-31G(d,p) and pcseg-1 basis sets, however, leads to almost identical performance for the PBE, B3LYP, ωB97X-D, and M06-L functionals (MAD ∼ 2−3 kJ/mol), and the lowest MAD of 0.9 kJ/mol is obtained by optimizing the pcseg-1 basis set for the M06-L functional. It is worth noting that functionals that inherently do not describe dispersion, like PBE and B3LYP, are capable of producing low MAD values (2−3 kJ/mol) when the 6-31G(d,p) and pcseg-1 basis sets are optimized to produce error compensation. The ability of these small basis sets to produce an attractive interaction is most likely associated with basis set superposition errors.[25]

Analysis of Optimized Contraction Coefficients. Optimizing basis set parameters for minimizing the error against a selected set of reference results is not a recommended procedure but used in the present case to illustrate that basis set differences arising from details in their construction can lead to significant differences in error cancellations in combination with a given DFT method. The changes in the contraction coefficients upon optimization are small but different for each functional, showing that the parameter optimization attempts to compensate for errors in the functional and (possible) neglect of for example vibrational and solvent effects, rather than being a general basis set deficiency. The increase in the total atomic energies by the optimization of two p-function contraction coefficients for the results in Tables 2 and 3, for example, is on average ∼40 milli-Hartree, which is comparable to the ∼37 milli-Hartree error in the 6-31G(d,p) or pcSseg-1 basis sets relative to the basis set limiting energy. These tweaked basis sets would thus appear perfectly valid from an energetic point of view. Furthermore, in a molecular calculation, the contracted p-function is only one of many independent functions, and the variational procedure for determining the wave function allows combination with other uncontracted p-functions in the basis sets and with basis functions on other atoms. To illustrate the effect of changing the contraction coefficients, Figure 3 shows the radial behavior of the atomic p-function for a carbon atom with the ωB97X-D functional. The 6-31G(d,p), pcseg-1, and pcSseg-1 are the regular basis sets, while the 6-31G(d,p)_opt and pcSseg-1_opt are the basis sets corresponding to the optimized contraction coefficients from Table 3. This has been chosen as an example where the performance in terms of MAD is similar for the 6-31G(d,p) and pcSseg-1, and they display similar MAD reductions upon reoptimization of the p-contraction coefficients (MAD values change from 14.9 to 7.0 ppm and from 16.7 to 8.2 ppm, respectively). The regular 6-31G(d,p) basis set has a slightly faster decay at long distance compared to pcseg-1 and pcSseg-1, while the 6-31G(d,p)_opt brings the long-range behavior more in line with the other basis sets, at the expense of the behavior near the maximum. The effect of optimizing the pcSseg-1 contraction coefficients has very little effect on the overall shape, as the pcSseg-1_opt is well within the variations of the regular 6-31G(d,p), pcseg-1, and pcSseg-1 basis sets. While the shapes of the 6-31G(d,p) and 6-31G(d,p)_opt basis sets are more different than the pcSseg-1 and pcSseg-1_opt basis sets, the performance against the reference data (Table 3) is very similar, and the effect of optimizing the two contraction coefficients leads to almost the same improvement of the MAD value (8 and 9 ppm). This illustrates that minor differences in the construction details of basis sets can lead to significantly different error compensations when used in connection with, for example, different DFT functionals. It is perhaps surprising that such small details in the basis set can affect results of the magnitudes shown in Tables 2−8.

Basis sets differ in the number of primitive functions, how the exponents of the primitive functions are chosen, the contraction pattern, and how the contraction coefficients are determined.[58−60] Details in these choices can lead to different basis sets that in an energetic sense are almost of the same quality but may lead to different error cancellations when used in combination with a particular DFT method for selected properties against a particular reference set of data. Given that the number of proposed basis sets and functionals each is in the hundreds,[67] the combinatorial space for identifying error cancellations is very large. We have in the present case tailored medium sized basis sets toward individual functionals and specific reference data to assess the magnitude of error cancellation. One could alternatively tailor parameters in the functional toward reproducing selected reference data with a given (selection of) basis sets, such as, for example, in the KT*n*[68] and WP04/WC04[69] functionals for calculating nuclear magnetic shielding constants. This will, analogous to the basis set optimizations in Tables 2−5, lead to method parameters trying to absorb errors from incomplete basis sets and effects not included in the calculation, such as solvent and vibrational contributions. We note that functionals containing many parameters that are adjusted to reproduce reference results may be especially prone to incorporate error compensation in the parameter space.

SUMMARY

The present work shows that there is a fine line between method calibration and data fitting. Calibration performed at the basis set limit can be used to assess the inherent accuracy of a given method[70] when comparing to reference results under the same conditions, such as a nonrelativistic fixed nucleus approximation. Using results from methods with a medium sized basis set and comparing to experimental data that include, for example, solvent and finite temperature effects, which the computational model may ignore or treat incompletely, should perhaps better be considered as data fitting, as a good performance in many cases is a result of error cancellation between the components defining the model. While error cancellation is expected to diminish as the size of the calibration data increases, the present results show that the effects are clearly visible for reference data sets containing ∼100 points, which is a typical size used for calibration purposes. Designing complex fitting functions by optimization of functional and/or basis set parameters is a possibility, and such models may even have predictive capabilities, but one may question whether this is a road that should be traveled.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00477.

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Descriptions of the XLSX data files. Optimized basis set contraction coefficients are not provided, in order to prevent proliferation of data fitting procedures using method and application tweaked basis sets (PDF)
Suppl_1 containing calculated nuclear magnetic shielding constants for the results in Table 2 and QSAR models (XLSX)
Suppl_2 containing calculated nuclear magnetic shielding constants for the results in Table 3 (XLSX)
Suppl_3 containing calculated nuclear magnetic shielding constants for the results in Tables 4 and 5 (XLSX)
Suppl_4 containing calculated total energies for the results in Table 6 (XLSX)
Suppl_5 containing calculated total energies for the results in Table 7 (XLSX)
Suppl_6 containing calculated total energies for the results in Table 8 (XLSX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: frj@chem.au.dk.

ORCID

Frank Jensen: 0000-0002-4576-5838

Funding

This work was supported by Grant No. 4181-00030B from the Danish Council for Independent Research.

Notes

The author declares no competing financial interest.

REFERENCES

(1) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational Prediction of H-1 and C-13 Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012**, *112* (3), 1839−1862.

(2) Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. Fully Automated Quantum-Chemistry-Based Computation of Spin-Spin-Coupled Nuclear Magnetic Resonance Spectra. *Angew. Chem., Int. Ed.* **2017**, *56* (46), 14763−14769.

(3) Joyce, L. A.; Nawrat, C. C.; Sherer, E. C.; Biba, M.; Brunskill, A.; Martin, G. E.; Cohen, R. D.; Davies, I. W. Beyond optical rotation: what's left is not always right in total synthesis. *Chemical Science* **2018**, *9* (2), 415−424.

(4) Reinscheid, U. M. Determination of the absolute configuration of two estrogenic nonylphenols in solution by chiroptical methods. *J. Mol. Struct.* **2009**, *918* (1−3), 14−18.

(5) Reinscheid, U. M.; Kock, M.; Cychon, C.; Schmidts, V.; Thiele, C. M.; Griesinger, C. The Absolute Configuration of Dibromopalau'amine. *Eur. J. Org. Chem.* **2010**, *2010* (36), 6900−6903.

(6) Astrand, P. O.; Ruud, K. Zero-point vibrational contributions to fluorine shieldings in organic molecules. *Phys. Chem. Chem. Phys.* **2003**, *5* (22), 5015−5020.

(7) Ruud, K.; Astrand, P. O.; Taylor, P. R. Zero-point vibrational effects on proton shieldings: Functional-group contributions from ab initio calculations. *J. Am. Chem. Soc.* **2001**, *123* (20), 4826−4833.

(8) Manninen, P.; Ruud, K.; Lantto, P.; Vaara, J. Leading-order relativistic effects on nuclear magnetic resonance shielding tensors. *J. Chem. Phys.* **2005**, *122* (11), 114107.

(9) Sabzyan, H.; Buzari, B. Theoretical study of the contribution of vibrational motions to nuclear shielding constants. *Chem. Phys.* **2008**, *352* (1−3), 297−305.

(10) Flaig, D.; Maurer, M.; Hanni, M.; Braunger, K.; Kick, L.; Thubauville, M.; Ochsenfeld, C. Benchmarking Hydrogen and Carbon NMR Chemical Shifts at HF, DFT, and MP2 Levels. *J. Chem. Theory Comput.* **2014**, *10* (2), 572−578.

(11) Kwan, E. E.; Liu, R. Y. Enhancing NMR Prediction for Organic Compounds Using Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11* (11), 5083−5089.

(12) Pierens, G. K. H-1 and C-13 NMR Scaling Factors for the Calculation of Chemical Shifts in Commonly Used Solvents Using Density Functional Theory. *J. Comput. Chem.* **2014**, *35* (18), 1388−1394.

(13) Toomsalu, E.; Burk, P. Critical test of some computational methods for prediction of NMR H-1 and C-13 chemical shifts. *J. Mol. Model.* **2015**, *21* (9), 244.

(14) Latypov, S. K.; Polyancev, F. M.; Yakhvarov, D. G.; Sinyashin, O. G. Quantum chemical calculations of P-31 NMR chemical shifts: scopes and limitations. *Phys. Chem. Chem. Phys.* **2015**, *17* (10), 6976−6987.

(15) Sarotti, A. M.; Pellegrinet, S. C. A Multi-Standard Approach for GIAO C-13 NMR Calculations. *J. Org. Chem.* **2009**, *74* (19), 7254−7260.

(16) Sarotti, A. M.; Pellegrinet, S. C. Application of the Multi-standard Methodology for Calculating H-1 NMR Chemical Shifts. *J. Org. Chem.* **2012**, *77* (14), 6059−6065.

(17) Buss, A.; Koch, R. Simulation of NMR chemical shifts in heterocycles: a method evaluation. *J. Mol. Model.* **2017**, *23* (1), 9.

(18) Benassi, E. Benchmarking of Density Functionals for a Soft but Accurate Prediction and Assignment of H-1 and C-13 NMR Chemical Shifts in Organic and Biological Molecules. *J. Comput. Chem.* **2017**, *38* (2), 87−92.

(19) Iron, M. A. Evaluation of the Factors Impacting the Accuracy of C-13 NMR Chemical Shift Predictions using Density Functional Theory-The Advantage of Long-Range Corrected Functionals. *J. Chem. Theory Comput.* **2017**, *13* (11), 5798−5819.

(20) Saunders, C.; Khaled, M. B.; Weaver, J. D.; Tantillo, D. J. Prediction of F-19 NMR Chemical Shifts for Fluorinated Aromatic Compounds. *J. Org. Chem.* **2018**, *83* (6), 3220−3225.

(21) Xin, D. Y.; Sader, C. A.; Chaudhary, O.; Jones, P. J.; Wagner, K.; Tautermann, C. S.; Yang, Z.; Busacca, C. A.; Saraceno, R. A.; Fandrick, K. R.; Gonnella, N. C.; Horspool, K.; Hansen, G.; Senanayake, C. H. Development of a C-13 NMR Chemical Shift Prediction Procedure Using B3LYP/cc-pVDZ and Empirically Derived Systematic Error Correction Terms: A Computational Small Molecule Structure Elucidation Method. *J. Org. Chem.* **2017**, *82* (10), 5135−5145.

(22) Hill, D. E.; Vasdev, N.; Holland, J. P. Evaluating the accuracy of density functional theory for calculating H-1 and C-13 NMR chemical shifts in drug molecules. *Comput. Theor. Chem.* **2015**, *1051*, 161−172.

(23) Konstantinov, I. A.; Broadbelt, L. J. Regression Formulas for Density Functional Theory Calculated H-1 and C-13 NMR Chemical Shifts in Toluene-d(8). *J. Phys. Chem. A* **2011**, *115* (44), 12364−12372.

(24) Jain, R.; Bally, T.; Rablen, P. R. Calculating Accurate Proton Chemical Shifts of Organic Molecules with Density Functional Methods and Modest Basis Sets. *J. Org. Chem.* **2009**, *74* (11), 4017−4023.

(25) Kruse, H.; Goerigk, L.; Grimme, S. Why the Standard B3LYP/6-31G*Model Chemistry Should Not Be Used in DFT Calculations of Molecular Thermochemistry: Understanding and Correcting the Problem. *J. Org. Chem.* **2012**, *77* (23), 10824−10834.

(26) Shaghaghi, H.; Fathi, F.; Ebrahimi, H. P.; Tafazzoli, M. Quantitative prediction of 13C NMR chemical shifts in solvent using PCM-ONIOM method and optimally selected wave function. *Concepts Magn. Reson., Part A* **2013**, *42* (1), 1−13.

(27) McAnanama-Brereton, S.; Waller, M. P. Rational Density Functional Selection Using Game Theory. *J. Chem. Inf. Model.* **2018**, *58* (1), 61−67.

(28) Bagno, A.; Saielli, G. Addressing the stereochemistry of complex organic molecules by density functional theory-NMR. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2015**, *5* (2), 228−240.

(29) Grimblat, N.; Sarotti, A. M. Computational Chemistry to the Rescue: Modern Toolboxes for the Assignment of Complex

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Molecules by GIAO NMR Calculations. *Chem. - Eur. J.* **2016**, *22* (35), 12246−12261.
(30) Sarotti, A. M. Structural revision of two unusual rhamnofolane diterpenes, curcusones I and J, by means of DFT calculations of NMR shifts and coupling constants. *Org. Biomol. Chem.* **2018**, *16* (6), 944−950.
(31) Grimblat, N.; Zanardi, M. M.; Sarotti, A. M. Beyond DP4: an Improved Probability for the Stereochemical Assignment of Isomeric Compounds using Quantum Chemical Calculations of NMR Shifts. *J. Org. Chem.* **2015**, *80* (24), 12526−12534.
(32) Lodewyk, M. W.; Tantillo, D. J. Prediction of the Structure of Nobilisitine A Using Computed NMR Chemical Shifts. *J. Nat. Prod.* **2011**, *74* (5), 1339−1343.
(33) Wiitala, K. W.; Al-Rashid, Z. F.; Dvornikovs, V.; Hoye, T. R.; Cramer, C. J. Evaluation of various DFT protocols for computing H-1 and C-13 chemical shifts to distinguish stereoisomers: diastereomeric 2-, 3-, and 4-methylcyclohexanols as a test set. *J. Phys. Org. Chem.* **2007**, *20* (5), 345−354.
(34) Wiitala, K. W.; Cramer, C. J.; Hoye, T. R. Comparison of various density functional methods for distinguishing stereoisomers based on computed H-1 or C-13 NMR chemical shifts using diastereomeric penam beta-lactams as a test set. *Magn. Reson. Chem.* **2007**, *45* (10), 819−829.
(35) Grimblat, N.; Kaufman, T. S.; Sarotti, A. M. Computational Chemistry Driven Solution to Rubriflordilactone B. *Org. Lett.* **2016**, *18* (24), 6420−6423.
(36) Nguyen, Q. N. N.; Tantillo, D. J. Using quantum chemical computations of NMR chemical shifts to assign relative configurations of terpenes from an engineered Streptomyces host. *J. Antibiot.* **2016**, *69* (7), 534−540.
(37) Smith, S. G.; Goodman, J. M. Assigning the Stereochemistry of Pairs of Diastereoisomers Using GIAO NMR Shift Calculation. *J. Org. Chem.* **2009**, *74* (12), 4597−4607.
(38) Ermanis, K.; Parkes, K. E. B.; Agback, T.; Goodman, J. M. Doubling the power of DP4 for computational structure elucidation. *Org. Biomol. Chem.* **2017**, *15* (42), 8998−9007.
(39) Smith, S. G.; Goodman, J. M. Assigning Stereochemistry to Single Diastereoisomers by GIAO NMR Calculation: The DP4 Probability. *J. Am. Chem. Soc.* **2010**, *132* (37), 12946−12959.
(40) http://cheshirenmr.info.
(41) Jensen, F. Segmented Contracted Basis Sets Optimized for Nuclear Magnetic Shielding. *J. Chem. Theory Comput.* **2015**, *11* (1), 132−138.
(42) Vosko, S. H.; Wilk, L.; Nusair, M. ACCURATE SPIN-DEPENDENT ELECTRON LIQUID CORRELATION ENERGIES FOR LOCAL SPIN-DENSITY CALCULATIONS - A CRITICAL ANALYSIS. *Can. J. Phys.* **1980**, *58* (8), 1200−1211.
(43) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865−3868.
(44) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125* (19), 194101.
(45) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. AB-INITIO CALCULATION OF VIBRATIONAL ABSORPTION AND CIRCULAR-DICHROISM SPECTRA USING DENSITY-FUNCTIONAL FORCE-FIELDS. *J. Phys. Chem.* **1994**, *98* (45), 11623−11627.
(46) Becke, A. D. DENSITY-FUNCTIONAL THERMOCHEMISTRY 0.3. THE ROLE OF EXACT EXCHANGE. *J. Chem. Phys.* **1993**, *98* (7), 5648−5652.
(47) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615−6620.
(48) Hehre, W. J.; Ditchfield, R.; Pople, J. A. SELF-CONSISTENT MOLECULAR-ORBITAL METHODS 0.12. FURTHER EXTENSIONS OF GAUSSIAN-TYPE BASIS SETS FOR USE IN MOLECULAR-ORBITAL STUDIES OF ORGANIC-MOLECULES. *J. Chem. Phys.* **1972**, *56* (5), 2257−2261.
(49) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. SELF-CONSISTENT MOLECULAR-ORBITAL METHODS 0.20. BASIS SET FOR CORRELATED WAVE-FUNCTIONS. *J. Chem. Phys.* **1980**, *72* (1), 650−654.
(50) Jensen, F. Unifying General and Segmented Contracted Basis Sets. Segmented Polarization Consistent Basis Sets. *J. Chem. Theory Comput.* **2014**, *10* (3), 1074−1085.
(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian-09*, Revision D-01; Gaussian Inc.: 2009.
(52) Jensen, F. Basis set convergence of nuclear magnetic shielding constants calculated by density functional methods. *J. Chem. Theory Comput.* **2008**, *4* (5), 719−727.
(53) Teale, A. M.; Lutnaes, O. B.; Helgaker, T.; Tozer, D. J.; Gauss, J. Benchmarking density-functional theory calculations of NMR shielding constants and spin-rotation constants using accurate coupled-cluster calculations. *J. Chem. Phys.* **2013**, *138* (2), 024111.
(54) Stoychev, G. L.; Auer, A. A.; Izsak, R.; Neese, F. Self-Consistent Field Calculation of Nuclear Magnetic Resonance Chemical Shielding Constants Using Gauge-Including Atomic Orbitals and Approximate Two-Electron Integrals. *J. Chem. Theory Comput.* **2018**, *14* (2), 619−637.
(55) Zhao, Y.; Truhlar, D. G. Improved description of nuclear magnetic resonance chemical shielding constants using the M06-L meta-generalized-gradient-approximation density functional. *J. Phys. Chem. A* **2008**, *112* (30), 6794−6799.
(56) Sarotti, A. M. Successful combination of computationally inexpensive GIAO C-13 NMR calculations and artificial neural network pattern recognition: a new strategy for simple and rapid detection of structural misassignments. *Org. Biomol. Chem.* **2013**, *11* (29), 4847−4859.
(57) Zanardi, M. M.; Sarotti, A. M. GIAO C-H COSY Simulations Merged with Artificial Neural Networks Pattern Recognition Analysis. Pushing the Structural Validation a Step Forward. *J. Org. Chem.* **2015**, *80* (19), 9371−9378.
(58) Jensen, F. Atomic orbital basis sets. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2013**, *3* (3), 273−295.
(59) Nagy, B.; Jensen, F. *Reviews in Computational Chemistry* **2017**, *30*, 93.
(60) Hill, J. G. Gaussian basis sets for molecular applications. *Int. J. Quantum Chem.* **2013**, *113* (1), 21−34.
(61) Grimme, S.; Steinmetz, M.; Korth, M. How to compute isomerization energies of organic molecules with quantum chemical methods. *J. Org. Chem.* **2007**, *72* (6), 2118−2126.
(62) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J. Chem. Theory Comput.* **2008**, *4* (2), 297−306.
(63) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Development and assessment of a new hybrid density functional model for thermochemical kinetics. *J. Phys. Chem. A* **2004**, *108* (14), 2715−2719.
(64) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. Benchmark database of barrier heights for heavy atom transfer, nucleophilic

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

substitution, association, and unimolecular reactions and its use to test theoretical methods. *J. Phys. Chem. A* **2005**, *109* (9), 2012−2018.
(65) Jensen, F. Describing Anions by Density Functional Theory: Fractional Electron Affinity. *J. Chem. Theory Comput.* **2010**, *6* (9), 2726−2735.
(66) Rezac, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427−2438.
(67) 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* (19), 2315−2372.
(68) Keal, T. W.; Tozer, D. J. A semiempirical generalized gradient approximation exchange-correlation functional. *J. Chem. Phys.* **2004**, *121* (12), 5654−5660.
(69) Wiitala, K. W.; Hoye, T. R.; Cramer, C. J. Hybrid density functional methods empirically optimized for the computation of C-13 and H-1 chemical shifts in chloroform solution. *J. Chem. Theory Comput.* **2006**, *2* (4), 1085−1092.
(70) Kupka, T.; Stachow, M.; Nieradka, M.; Kaminsky, J.; Pluta, T. Convergence of Nuclear Magnetic Shieldings in the Kohn-Sham Limit for Several Small Molecules. *J. Chem. Theory Comput.* **2010**, *6* (5), 1580−1589.

DOI: 10.1021/acs.jctc.8b00477
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX