JOC The Journal of Organic Chemistry

Article

pubs.acs.org/joc

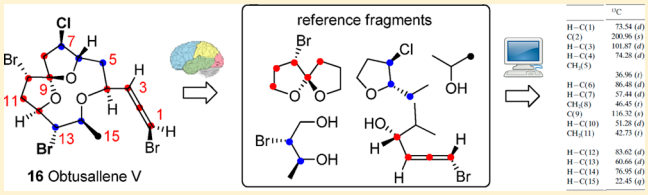
Improving the Accuracy of Computed ^{13}C NMR Shift Predictions by Specific Environment Error Correction: Fragment Referencing

Keith G. Andrews and Alan C. Spivey*

Department of Chemistry, Imperial College, London SW7 2AZ, United Kingdom

Supporting Information

ABSTRACT: The accuracy of both Gauge-including atomic orbital (GIAO) and continuous set of gauge transformations (CSGT) ^{13}C NMR spectra prediction by Density Functional Theory (DFT) at the B3LYP/6-31G** level is shown to be usefully enhanced by employing a 'fragment referencing' method for predicting chemical shifts without recourse to empirical scaling. Fragment referencing refers to a process of reducing the error in calculating a particular NMR shift by consulting a similar molecule for which the error in the calculation is easily deduced. The absolute accuracy of the chemical shifts predicted when employing fragment referencing relative to conventional techniques (e.g., using TMS or MeOH/benzene dual referencing) is demonstrated to be improved significantly for a range of substrates, which illustrates the superiority of the technique particularly for systems with similar chemical shifts arising from different chemical environments. The technique is particularly suited to molecules of relatively low molecular weight containing 'non-standard' magnetic environments, e.g., α to halogen atoms, which are poorly predicted by other methods. The simplicity and speed of the technique mean that it can be employed to resolve routine structural assignment problems that require a degree of accuracy not provided by standard incremental or hierarchically ordered spherical description of environment (HOSE) algorithms. The approach is also demonstrated to be applicable when employing the MP2 method at 6-31G**, cc-pVDZ, aug-cc-pVDZ, and cc-pVTZ levels, although none of these offer advantage in terms of accuracy of prediction over the B3LYP/6-31G** DFT method.



INTRODUCTION

The use of density functional theory (DFT) NMR prediction methods to differentiate isomers and elucidate the structures of complex compounds (e.g., natural products) is an active and burgeoning area of research.[1−3] Rablen[4] and Bifulco[5] have been credited with pioneering these aspects of the field, building upon seminal advances in DFT NMR prediction methods by several groups,[6] including those of Schindler,[7] Schleyer,[8] Rasul,[9] and Pulay.[10] Of particular recent note, Goodman and Smith have developed procedures for assigning probabilities to sets of isomers when one set of experimental data is available (DP4)[11] or distinguishing two isomers when two sets of data are available (CP3).[12]

Although the structure and/or stereochemistry of many complex natural products have been reassigned using these techniques,[13] routine use of DFT calculations to solve seemingly less complex, day-to-day problems is rare.[2,3] Very rapid NMR spectrum predictors based on incremental or hierarchically ordered spherical description of environment (HOSE) algorithms, such as those provided with some structure editing software suites (e.g., in ChemOffice and ACD/Laboratories, respectively), are widely used resources in this regard but provide only an approximate guide to what the ^{1}H and/or ^{13}C NMR spectra of a molecule may look like.[14] In particular, their accuracy in distinguishing stereo- and certain regioisomers is poor by comparison with semiempirical and quantum mechanical methods, particularly for less commonly encountered structural motifs,[3] although integration with artificial neural networks (ANNs) holds promise for improving this situation.[15]

Gauge-including atomic orbital (GIAO),[16,17] individual gauge for localized orbitals (IGLO),[18] and continuous set of gauge transformations (CSGT)[19] are the most frequently employed methods in DFT-NMR calculations, and it is recognized that accurate predictions require the gauge-invariant procedures that these methods implement.[1,20,21] However, notwithstanding the DP4 and CP3 statistical data analysis parameters referred to above that allow convenient processing of GIAO or CSGT raw data, more widespread implementation of DFT NMR predictions is impeded by the lack of a consensus regarding the most effective data analysis techniques and parameters to apply when distinguishing isomers.[1−3,12]

The removal of systematic errors from the calculation process is also problematic. One approach is to use empirically scaled parameters extracted from the analysis of data sets from large numbers of compounds for each computational method/configuration. However, this necessitates the use of distinct values for each new method/solvent model and is dependent on the availability of an appropriate reference data set on a case-by-case basis to provide, on average, improved shift predictions.[22−24] More commonly, referencing of calculated

Received: August 21, 2013
Published: October 20, 2013

   dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

shifts to tetramethylsilane (TMS) is applied. Subsequent analysis then involves data fitting (empirical scaling), using the relevant experimental shifts to determine the best fitting set of calculated shifts. This approach benefits from universal applicability but is unable to account for the differing errors associated with specific chemical environments such as sp^3 versus sp^2 and sp hybridized carbons, which will bias the fitting process.[25] Moreover, when the number of data points is small, there is an increased risk of false positives that arise when, e.g., competing data points happen to scatter evenly about a good fit and data scaling may bias values toward the rogue data set. To minimize these risks, it is particularly important to remove environmental (e.g., hybridization and electronic) bias in the raw data before the data fitting stage; vide infra.

We recently sought to employ GIAO NMR prediction to solve a structural assignment problem encountered during the exploration of a ring-closing/ring-opening metathesis approach to a natural product core.[26] However, we found current methods for ^{13}C NMR predictions in particular to give conflicting outcomes, thereby making it difficult to a reach confident structural assignment; vide infra.

We considered that the root cause of the difficulties we encountered was the dependences in current literature methods upon empirical scaling to remove systematic error and a reliance on TMS as a reference molecule. Our experience indicates that empirical scaling techniques can be damaging to an analysis, due to the assumption that the error in a calculation is proportional to the chemical shift in question. Although this assumption may hold within, e.g., the GIAO calculation method at a gross level, it does not hold when attempting to tease apart multiple shifts occupying a narrow shift range. Thus, in this work, we demonstrate that the correlation between chemical shift and calculated error is not a precise one and that if environment-specific error is directly accounted for by improved referencing, then better accuracies can be attained without substantive additional work.

It has long been recognized that TMS is a poor reference for generating predicted shifts,[1−3] particularly given that heavy atoms such as silicon are known to introduce specific perturbations to proximal atoms![27] Although other single shift references have been advocated,[28−30] we contend that *any single environment* is a poor reference choice particularly as any subsequent scaling negates any arbitrary offset.[31,32] Sarotti and Pellegrinet have reported the use of a dual reference system that they showed can give improved absolute accuracies relative to using a TMS reference.[25] Their dual standard approach employs methanol as a reference for sp^3 environments and benzene for sp^2 and sp environments. These standards were suggested to be the most suitable solvents for a range of compounds and theory methods. However, even dual standard referencing can only be expected on average to improve the fortunes of a range of users with a range of molecules and cannot be relied upon to improve the prediction for every case.

Specific sources of error in DFT ^{13}C NMR predictions include[3] electron correlation,[33] solvent interactions,[34] conformational mobility,[2,35] vibrational phenomena,[36] and relativistic/heavy atom effects.[37] While various approaches have been developed to minimize these individually,[3] including use of correlated computational methods, solvation modeling,[38] conformational[39] and vibrational[40] averaging, heavy atom corrections,[41] and the aforementioned linear regression methods, our approach has been to minimize error in a more holistic and pragmatic manner. Specifically, we have developed a general approach to multi standard referencing, which we have termed 'fragment referencing'. Theoretically, this approach is applicable to all molecules, basis sets, theories, and solvent models and, most importantly, is straightforward to implement on a case by case basis even by the non-specialist.

While we recognize that employing empirical scaling or a TMS reference regularly gives an acceptable result, we have sought to exemplify our method to show its applicability, both in improving results successfully predicted using existing methods and also in scenarios where these traditional methods are unsatisfactory. We have found our method to give more accurate and confident results and to do so more consistently.

COMPUTATIONAL METHODS

The key methodology employed is similar to that employed by Goodman and Burton.[42] The structures of interest were optimized to their lowest energy gas phase conformers with the MM2[43] or MMFF94[44] forcefields within ChemBio 3D Ultra (version 12.0.2.1076). The molecules examined here have few rotatable bonds, and the possible conformers were easily identified manually.[45]

All further calculations were performed using Gaussian (version G09W; release C.01) employing the popular DFT functional, B3LYP,[46,47] which has precedent for both accurate geometry optimization and NMR shielding tensor calculation. The conformers from the initial searches were optimized using the 6-31G** [6-31G(d,p)] basis set[48] for all atoms, and NMR calculations were performed at the same level of theory, using the GIAO[16,17] and CGST[19] method. In the case of the cyanoindoles (Scheme 3), additional comparative calculations were performed employing the MP2 method[49,50] with 6-31G**,[48] cc-pVDZ,[51] aug-cc-pVDZ,[52] and cc-pVTZ[53] basis sets and the GIAO[16,17] procedure.

Structures were optimized in the gas phase, and for unique conformers, their validity as true minima was confirmed with a vibrational analysis calculation to ascertain that no imaginary modes existed. After the key structures had been verified, subsequent similar structures were assumed to be valid by analogy to reduce expensive frequency analysis calculations. The energies were recovered with a single point energy (SPE) calculation in solution, using the conductor-like polarizable continuum model (CPCM) solvation model as implemented in Gaussian, with chloroform ($CHCl_3$) as the solvent. Although it is appreciated that the zero-point energy corrected Gibbs free energies should be used for the Boltzmann weightings, it was found that the error in the final mean absolute error (MAE) values using the SPE values was suitably small for our purposes. The NMR calculations were also completed in $CHCl_3$ at B3LYP/6-31G** level unless otherwise noted (e.g., Tables 9 and 10).

The chemical shift values for each isomer were obtained by using Boltzmann weighted average shifts over the set of conformers, calculated by first averaging the chemical shifts for degenerate symmetry related carbon environments within each conformer. Each conformer then contributed a percentage of its chemical shift value to the final total value per environment depending on its Boltzmann weighted energy contribution at 298 K (eq 1).

$$\sigma^x = \frac{\sum_i \sigma_i^x \exp\left(-\frac{E_i}{RT}\right)}{\sum_i \exp\left(-\frac{E_i}{RT}\right)} \quad (1)$$

Here, σ^x is the weighted average shielding tensor of the atom(s), σ_i^x is the raw shielding tensor of the atom(s) in conformer i, and E_i is the energy of the associated conformer. Finally, TMS was optimized in a symmetry restricted (tetrahedral) geometry optimization calculation in chloroform, and the single NMR shielding tensor for all four carbon atoms was obtained again in chloroform at the B3LYP/6-31G** level. The value obtained was 192.172. All final shifts were then calculated as $\sigma_{TMS} - \sigma_{calc}$. Where multistandard references were used, shifts were calculated according to eq 2:

$$\delta^{x}_{calc} = \sigma^{ref}_{calc} + \delta^{ref} - \sigma^{x}_{calc} \quad (2)$$

Here, δ^{x}_{calc} and δ^{ref} are the experimental shifts of the substrate and the reference, respectively, and σ^{ref}_{calc} and σ^{x}_{calc} are the calculated shielding tensors of the substrate and the reference, respectively, as described by Sarotti and Pellegrinet.[25] Note that unlike TMS, these reference values were not optimized in chloroform.

Data was scaled according to eq 3:

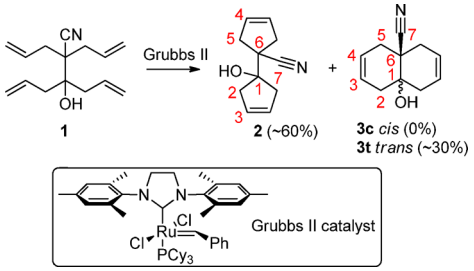
$$\delta_{scaled} = \frac{\delta_{calc} - \text{intercept}}{\text{slope}} \quad (3)$$

Here, δ_{scaled} is the scaled substrate shift (ppm), δ_{calc} is the calculated substrate shift (ppm) and the intercept and slope are the parameters of a linear regression when plotting the experimental shifts against the calculated shifts for a substrate.

RESULTS AND DISCUSSION

Our interest in GIAO NMR prediction was initiated by the need to assign structures to the products of the following alkene metathesis reaction (Scheme 1).

Scheme 1. Alkene Metathesis on Tetraene 1 To Give Biscyclopentene 2 and *trans*-Tetradehydrodecalin 3t



This reaction was envisioned to provide a novel entry to *trans*-tetradehydrodecalin **3t** as we hoped to develop this chemistry for the preparation of the natural product core (−)-euonyminol.[26] In the event, this reaction furnished a mixture of two major products in a 2:1 molar ratio as determined by ^{1}H NMR integration: an unknown product and *trans*-tetradehydrodecalin **3t**.[54] Standard ^{1}H/^{13}C NMR, mass spectrometry (MS) and infrared (IR) data for the unknown product were consistent with either *cis*-tetradehydrodecalin **3c** or biscyclopentene **2**. Although a single crystal X-ray structure determination eventually confirmed this product as biscyclopentene **2**, we concurrently explored DFT NMR to resolve the structural ambiguity.[26]

The molecular structures of both *cis*-tetrahydrodecalin **3c** and biscyclopentene **2** were optimized from MM2 minimized structures at the DFT-B3LYP/6-31G** level, and the NMR shielding tensors were calculated at the same level in chloroform using the GIAO NMR method. The resulting predicted ^{1}H NMR data for both structures was processed in comparison to the actual experimental shift data for compound **2** in order to obtain a range of standard analytical parameters [e.g., linear correlation efficient (LCE), DP4]. Analysis of the data gave a confident result in favor of compound **2** (data not shown, see ref 26). By contrast, analysis of the predicted ^{13}C NMR data did not provide a clear-cut outcome (Table 1).[45]

The most striking result is the corrected mean absolute error (CMAE), best expressed through the BCMAE,[55] which implies the *cis*-tetradehydrodecalin **3c** predicted shifts correlate more strongly with the experimental shifts than do the predicted shifts of **2**. The DP4 parameter, which could overturn this result

Table 1. GIAO DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for Biscyclopentene 2 and *cis*-Tetradehydrodecalin 3c Calculated Using TMS as Reference (Relative to the Experimental Data for Compound 2)[a]**

		biscyclopentene **2**		*cis*-tetradehydrodecalin **3c**	
assign[b]	exptl[c]	calcd	Δ[c]	calcd	Δ[d]
C3	128.5	126.2	2.2	122.6	5.9
C4	128.5	125.7	2.7	120.5	8.0
C7	125.5	118.9	6.6	117.5	8.0
C1	83.1	85.1	−2.0	72.2	10.9
C6	49.2	52.5	−3.2	44.1	5.1
C2	44.9	46.3	−1.5	37.9	6.9
C5	41.6	43.5	−1.9	35.8	5.8
		biscyclopentene **2**		*cis*-tetradehydrodecalin **3c**	
ATE[e]		20.2		50.7	
MAE[e]		2.9		7.2	
SD[e]		3.3		1.8	
RMS[e]		3.3		7.5	
Range[e]		9.9		5.8	
LCE[e]		0.9976		0.9979	
CMAE[e]		1.6		1.3	
DP4[e]		0.38		0.62	
BCMAE[e]		0.44		0.56	

[a]All values are in ppm with the exception of LCE (unitless) and the probabilities DP4 and BCMAE (unitless fractions out of 1). [b]Carbon assignments for compound **2** as in Scheme 1. [c]Experimental ^{13}C NMR shifts for compound **2** recorded at 100 MHz in $CDCl_3$ (see ref 26). [d]Δ = difference in chemical shift relative to experiment; [e]Abbreviations: ATE, absolute total error; MAE, mean absolute error; SD, standard deviation; RMS, root-mean-squared error; Range = Max − Min, where Max and Min are the largest and smallest errors in ppm, respectively; LCE, linear correlation coefficient; CMAE, corrected mean absolute error, DP4, Goodman and Smith's probability parameter;[11] BMAE (Bayes' MAE) and BCMAE (Bayes' CMAE) are Bayes' theorem probabilities of the two MAE or CMAE values {e.g., $1 - [A/(A + B)]$}, respectively.[55]

if the shifts lay closer to the line of best fit (trendline), also gives the erroneous result. In this case, the absolute total error (ATE) is a strong indicator of the correct result, but why then do the purportedly more robust parameters such as CMAE and DP4 predict the incorrect isomer? Structure **3c** appears to have a fortuitously better correlation with the experimental data; had it had a lower absolute error too, the indications in favor of the incorrect result would have been persuasive. We hypothesized that for small data sets such as ours scaling could be influenced to the detriment of the outcome by a single large error. In our system, such a data point seemed likely to be the chemical shift (δ_C) of the nitrile carbon (C7), which was 6.6 ppm too low.

Rittner has performed a study of α-substituted acetonitriles and concluded that (relative to TMS as reference) the CSGT NMR method is superior to the GIAO method for calculating the carbon shifts of the nitrile group, based upon comparison of their standard deviations ($\Delta SD_{GIAO-CSGT}$ = 0.64). The α-carbon is also reported, and the standard deviation of each set this time favors the GIAO method ($\Delta SD_{CSGT-GIAO}$= 0.37).[56] We therefore predicted the ^{13}C NMR shifts of compounds **2** and **3c** at the same level of theory but using the CSGT NMR model (Table 2).

The CSGT predictions are significantly more assured than the previous GIAO ones and now indicate the correct result via all measures both with and without empirical scaling.

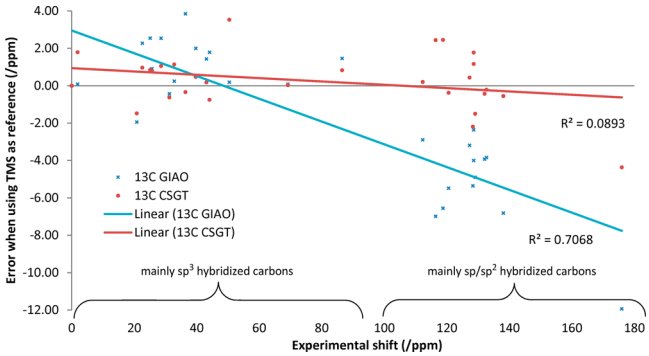
dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 2. CSGT DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for Biscyclopentene 2 and *cis*-Tetradehydrodecalin 3c Calculated Using TMS as Reference (Relative to the Experimental Data for Compound 2)[a]**

		biscyclopentene **2**		*cis*-tetradehydrodecalin **3c**	
assign[b]	exptl[c]	calcd	Δ[d]	calcd	Δ[d]
C3	128.5	130.4	−1.9	125.9	2.6
C4	128.5	130.1	−1.7	124.0	4.5
C7	125.5	128.1	−2.6	126.5	−1.0
C1	83.1	85.9	−2.8	73.3	9.8
C6	49.2	49.8	−0.6	41.7	7.5
C2	44.9	44.9	0.0	36.6	8.2
C5	41.6	41.5	0.0	34.0	7.6
		biscyclopentene **2**		*cis*-tetradehydrodecalin **3c**	
ATE[e]		9.5		41.2	
MAE[e]		1.4		5.9	
SD[e]		1.1		3.5	
RMS[e]		1.7		6.6	
Range[e]		2.8		10.8	
LCE[e]		0.9997		0.9970	
CMAE[e]		0.5		1.6	
DP4[e]		0.64		0.36	
BCMAE[e]		0.75		0.25	

[a]All values are in ppm with the exception of LCE (unitless) and the probabilities DP4 and BCMAE (unitless fractions out of 1). [b]Carbon assignments for compound **2** as in Scheme 1. [c]Experimental ^{13}C NMR shifts for compound **2** recorded at 100 MHz in $CDCl_3$ (see ref 26). [d]Δ = difference in chemical shift relative to experiment. [e]Abbreviations: as for Table 1.

Chart 1. Correlation between Isotropic Chemical Shift and Error in the Predicted Chemical Shift Relative to TMS for the GIAO DFT (B3LYP/6-31G) ^{13}C NMR Method[a]**



[a]Data from 29 environments from 11 molecules (including TMS) with a range of chemical functionalities. (See Supporting Information).

As discussed above, empirical scaling is often applied during the processing of predicted NMR data, but is its use always justified? Examining the scaling formula applied (eq 3), it is apparent that first a fixed value (the intercept value) is subtracted from each predicted shift in order to compensate for any arbitrary, invariant chemical shift difference that a given reference may have introduced. Each point is then adjusted in proportion to its magnitude by dividing by a fixed value (the slope). Consequently, points at low chemical shifts receive a smaller correction compared to points at high chemical shifts. How does this treatment compare to the real error associated with each shift? This can be assessed by inspection of data showing the relationship between chemical shift (δ_C/ppm) and the error in predicting that environment relative to a TMS reference using data points generated in this study (Chart 1).

Although there is a notable correlation between isotropic chemical shift and the magnitude and sign of the error when using a TMS reference for the GIAO method, it is apparent that it is not a close linear relationship (R^2 = 0.7). By contrast, the CSGT method shows a relatively weak correlation for the environments we have examined; environments with high chemical shifts have both positive and negative errors associated with them. It would appear therefore that there is a case for correcting the error based on the *chemical/magnetic environment* rather than the chemical shift when applying either the GIAO or CSGT method.

Fragment Referencing. As noted previously, using alternative references to TMS during DFT NMR calculations can give more accurate predictions. The reason is that the more similar the reference environment to the predicted environment, the more relevant is the cancellation of errors associated with calculating that particular environment. Thus the dual referencing of sp^3 centers to methanol and sp^2/sp centers to benzene as advocated by Sarotti and Pellegrinet results in improved corrections to the predicted shift.[25] However, do methanol and benzene constitute optimal references? On average these reference compounds should allow a reasonable approximation of the magnitude of error associated with calculating atoms of various hybridizations. Specifically, for the GIAO method this error is loosely correlated with chemical shift and so the associated empirical scaling provides, on average, improved error correction relative to the use of TMS as a single reference compound. This can be illustrated by applying methanol/benzene dual referencing to the comparison of predicted ^{13}C NMR data for biscyclopentene **2** and *cis*-tetradehydrodecalin **3c** relative to the experimental data for compound **2** using the DP4 analysis (data not shown, see Supporting Information). It improves the GIAO probability for the correct biscyclopentene **2** to 51% and the CSGT confidence to 71%; cf. 38% and 64%, respectively, when referenced to TMS (cf. Tables 1 and 2).

Drawing inspiration from Benzi et al.,[29] who noted that environment similarity is the key to identifying favorable references, we considered that the generation of specific environment corrections for each type of environment present in the molecule, i.e., fragment referencing, would offer significant advantages over dual referencing. Specifically, fragment referencing should allow for improved correlations to the experimental data without resort to empirical scaling. Logically, the most efficient method of generating appropriate fragment references is to identify small molecules that contain a large proportion of the chemical groups or 'fragments' in the molecule of interest and for which experimental ^{13}C NMR data is either available in the literature or can be recorded readily in the appropriate solvent. It pertains that little additional computational time is required to implement fragment referencing relative to standard methods, and as we shall see, this is time well invested.

Proof-of-Concept. Taking the structural assignment of biscyclopentene **2** versus *cis*-tetradehydrodecalin **3c** as an initial 'fragment referencing' case study, we initially selected cyclohexene, *tert*-butanol, and acetonitrile as sources of reference shifts (Figure 1).

Chemical shifts for the indicated carbons within each of the three reference fragments were calculated at the DFT-B3LYP/

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

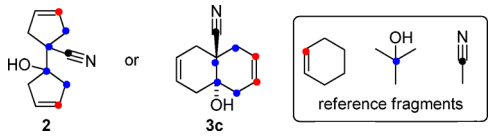
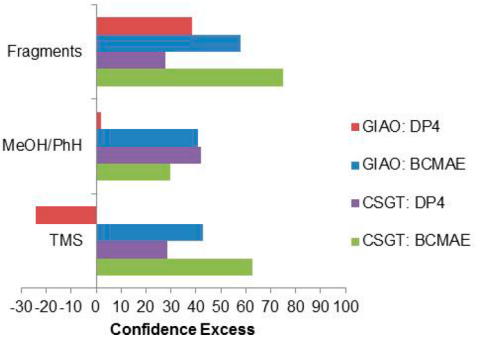


Figure 1. Initial molecules employed for fragment referencing during DFT ^{13}C NMR calculations for biscyclopentene **2** and *cis*-tetradehydrodecalin **3c**.

6-31G** level and converted into reference shifts using published experimental values. The GIAO and CSGT ^{13}C NMR predicted shifts for biscyclopentene **2** and *cis*-tetradehydrodecalin **3c** relative to these references were then computed and compared to the experimental data for compound **2**. (For full data see Supporting Information; comparative data presented in Chart 2.)

Chart 2. Comparison of GIAO and CSGT DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for Biscyclopentene 2 and *cis*-Tetradehydrodecalin 3c Calculated Using Fragments, MeOH/PhH and TMS, Respectively, As References (Relative to Experimental Values for Compound 2; Full Data in Supporting Information)[a]**



[a]Confidence excess (for compound **2**) is defined as DP4(**2**) − DP4(**3c**) × 100 or BCMAE(**2**) − BCMAE(**3c**) × 100, by analogy with the definition of enantiomeric excess (ee).

The BCMAE values for both GIAO and CSGT methods are significantly improved when using these fragments as compared to either methanol/benzene or TMS for referencing (blue and green bars). The GIAO DP4 analysis also gives a clear preference for the correct bis-tetradehydrodecalin **2** structure when employing fragment referencing, by contrast to the ambivalent and incorrect predictions afforded by methanol/benzene and TMS referencing, respectively (red bars). The CSGT DP4 confidence remains approximately constant irrespective of referencing system (purple bars). Whereas the dual standard referencing can be both beneficial and detrimental to prediction confidence as compared to TMS referencing, depending on the analytical parameter employed (cf. red/purple versus green/blue bars), fragment referencing clearly gives the most consistent outcome irrespective of analytical parameter.

Factors Affecting the Absolute Accuracy of Fragment-Based Predictions and Sources of Bias. Of course, the accuracy of the DFT ^{13}C NMR predicted shifts obtained using fragment referencing will depend on the choice of fragment. The best accuracies require fragments of close fit to the anticipated product, but useful levels of correction are clearly applied even using rather approximate fragments. That this is the case is evident from the foregoing proof-of-concept study wherein even when using the quaternary carbon in *tert*-butanol to reference all the sp^3 hybrid carbons in compounds **2** and **3c** (cf. Figure 1) rather than employing the allylic methylene carbon in cyclohexene as a reference for the analogous carbons in compounds **2** and **3c**, good results are obtained (Table 3).

Table 3. Comparison of Chemical Shift Error for Each Predicted Carbon Environment in *trans*-Tetradehydrodecalin 3t Calculated Using a Range of Fragment References (Relative to Experimental Values for Compound 3t; See Ref 54)[a]

fragments[b]	C3	C4	C7	C1	C6	C2	C5
a−c	0.7	0.9		−2.6		−1.0	−2.4
d−g			−0.1	−1.2	−3.9	−1.9	−3.4
h−j				−1.4	−2.4	0.0	−1.4
k, l			−0.9		−4.1		
m, n	0.5	0.7				0.6	−0.9
o−r	−0.7	−0.5		−2.0	−0.3	−0.6	−2.0
s				−3.4	−4.1		
2 (a−g)	−0.8	−0.6	−0.6	−1.5	−1.0	−0.5	−1.5
TMS	0.5	0.7	4.9	−4.6	−5.3	−3.1	−4.6
|Min|[c]	0.5	0.5	0.1	1.2	0.3	0.0	0.9

[a]All values are in ppm. [b]Fragment assignments as in Figure 2, carbon assignments as in Scheme 1. Where a range of fragment atoms are specified for multiple carbon atoms in **3t**, the chemically analogous environments are matched in each case. [c]This is the minimum absolute error in the column, i.e., the minimum absolute error calculated for that environment using the various reference molecules.

To explore this further, we investigated the effect of fragment choices on the MAE when predicting the ^{13}C NMR data for *trans*-tetradehydrodecalin **3t**. Initially, a range of potential fragments were selected (Figure 2).

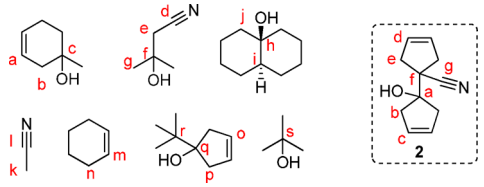


Figure 2. Molecules employed for fragment referencing during DFT ^{13}C NMR calculations for *trans*-tetradehydrodecalin 3t (Table 3), biscyclopentene 2 (Table 4) and *cis*-tetradehydrodecalin 3c (Table 5).

Chemical shifts for the indicated carbons within each of the reference fragments were calculated at the DFT-B3LYP/6-31G** level and converted into reference shifts using published experimental values as before. These were then used to compute predicted chemical shifts for *trans*-tetradehydrodecalin **3t**, and these were compared to the experimental data set[54] (Table 3).

The data indicate that the most 'complex' carbon environments are the most difficult to reference accurately. In general, as expected, the appropriate hybridization is very important, but the substitution number, the types of substituent, and the geometry also play a role. Electronics are clearly important for polarized systems, as shown by comparing 3-hydroxy-3-methylbutanenitrile and *tert*-butanol as fragments for the C1 quaternary carbon adjacent to the hydroxyl group (i.e., carbons f and s, respectively). The use of biscyclopentene **2** as a

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

reference 'fragment' proves to be an excellent choice giving calculated shifts with an MAE of only 0.9 ppm (Table 4). This table also illustrates practically how the fragment referencing method is applied via the tabulation of reference shifts.

Table 4. Comparison of GIAO DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for *trans*-Tetradehydrodecalin 3t Calculated Using Biscyclopentene 2 as the Reference 'Fragment' (Relative to Experimental Values for Compound 3t)[a]**

'fragment' (2)					*trans*-tetradehydrodecalin 3t		
frag[b]	exptl[c]	pred	ref[d]	assign[e]	pred	calcd[f]	exptl[g]
c	128.5	65.9	194.4	C3	69.7	124.7	124.1
g	125.5	73.3	198.8	C7	74.8	124.0	123.4
d	128.5	66.4	194.9	C4	70.7	124.2	123.3
a	83.1	107.1	190.2	C1	120.6	69.6	68.1
b	44.9	145.8	190.7	C2	153.1	37.6	37.1
f	49.2	139.7	189.0	C6	151.9	37.1	36.1[h]
e	41.6	148.7	190.2	C5	155.5	34.7	33.2

[a]All values are in ppm. [b]Fragment assignments as in Figure 2. [c]Experimental ^{13}C NMR shifts for compound **2** recorded at 100 MHz in $CDCl_3$ (see ref 26). [d]Reference shifts (ref) = experimental shifts (exptl) + predicted shifts (Pred). [e]Carbon assignments for compound **3t** as in Scheme 1. [f]Calculated shifts (calcd) = ref − pred. [g]Experimental ^{13}C NMR shifts for compound **3t** recorded at 63 MHz in $CDCl_3$ (see ref 54). [h]The chemical shift for C6 (i.e., the quaternary carbon α to the nitrile) was incorrectly assigned in ref 54 as being at δ 29.7 ppm; this peak is an impurity in the spectrum (see Supporting Information to ref 54), the correct experimental value is δ 36.1 ppm.

The range of errors shown in Table 3 raises a question, particularly in the context of using DFT ^{13}C NMR to distinguish between potential isomers (i.e., using DP4 or BCMAE analysis): could a damaging bias be introduced by referencing using a fragment(s) that better resembles one of the isomers in question? Consider two isomers, A (correct) and B (incorrect), competing for a data set. If we select a fragment that corrects B more effectively than it corrects A, could we incorrectly conclude B best fits the data?

A strength of fragment referencing is that it avoids the 'deliberate' bias introduced in data fitting. In data fitting, we are adjusting the incorrect isomer shifts to get as close to the real data as possible! On the other hand, if we have a fragment that corrects all of the error in the incorrect isomer, it will have converged on *its own real data points*. This means that even a moderate correction of the correct isomer may culminate in the correct result, although clearly the result depends on the exact extent of shift separation and error involved.[57] So if isomer A is chemically different from isomer B, a bias toward B is not necessarily damaging to the analysis. Many examples in this paper show an increase in the absolute error in the incorrect isomer data, as the data points converge on their own real shifts. If isomer A is chemically (and normally structurally also) similar to isomer B, then the bias will be low by definition, as both calculations should introduce a similar environmental error. The question of bias would seem, therefore, to be second to the question of the quality of the fragments. Concern as to the quality of any fragments may, of course, be investigated by consulting other fragments.

Although such errors resulting from even subconscious bias in selection of reference fragments therefore seem unlikely to detract from the utility of the method, we investigated using *trans*-tetradehydrodecalin **3t** as a reference 'fragment' in the prediction of ^{13}C NMR chemical shifts for both biscyclopentene **2** and *cis*-tetradehydrodecalin **3c** (Table 5).

Table 5. GIAO DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for Biscyclopentene 2 and *cis*-Tetradehydrodecalin 3c Calculated Using *trans*-Tetradehydrodecalin 3t as the Reference 'Fragment' (Relative to the Experimental Values for Compound 2)[a]**

		biscyclopentene 2		*cis*-tetradehydrodecalin 3c	
assign[b]	exptl	calcd	Δ[c]	calcd	Δ[c]
c	128.5	127.9	0.6	124.2	4.3
d	128.5	127.6	0.9	122.3	6.1
g	125.5	125.0	0.6	123.6	2.0
a	83.1	81.6	1.5	68.7	14.4
f	49.2	48.3	1.0	40.0	9.3
b	44.9	44.4	0.5	35.9	8.9
e	41.6	40.1	1.5	32.4	9.2
		biscyclopentene 2		*cis*-tetradehydrodecalin 3c	
ATE[d]		6.5		54.2	
MAE[d]		0.9		7.7	
SD[d]		0.4		3.7	
RMS[d]		1.0		8.6	
Range[d]		1.0		12.4	
LCE[d]		0.9999		0.9949	
CMAE[d]		0.3		2.0	
DP4[d]		0.76		0.24	
BCMAE[d]		0.87		0.13	
BMAE[d]		0.89		0.11	

[a]All values are in ppm with the exception of LCE (unitless) and the probabilities DP4, BMAE, and BCMAE (unitless fractions out of 1). [b]Carbon assignments as in Figure 2. [c]Δ = difference in chemical shift relative to experiment. [d]Abbreviations: as for Table 1.

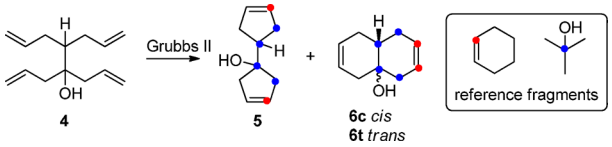
Clearly, *trans*-tetradehydrodecalin **3t** should be an excellent reference 'fragment' for its *cis*-congener **3c** but should also be a reasonably good reference for the biscyclopentene isomer **2**. The data show that, despite using an obviously 'biased' fragment, a much more confident *and correct* prediction in favor of biscyclopentene **2** is obtained via the DP4, BMAE, and BCMAE analyses relative to using TMS (cf. Table 1). Of course, with the increased accuracy that fragment referencing brings, the distinction in MAE between isomers, even for a biased system, should improve. As errors are reduced, shifts in incorrect isomers that lie close to the real data increasingly represent genuine cases of similar shifts at the expense of those due to error-induced coincidence. Here the Δ_{MAE} increases from 4.3 ppm to 6.8 ppm relative to a TMS reference. This divergence of relative accuracies facilitates distinguishing the two isomers with confidence.

It is also worth noting here that when processing raw predicted shifts using empirical scaling, a set of experimental data for the structure must be available. Fragment referencing does not require the real shifts, so the calculated shifts can be processed and improved even when no experimental data is available. Fragment referencing therefore allows improved *predictions* to be computed, whereas empirical scaling (e.g., BCMAE, DP4) simply enables improved *comparisons* between computed and real data sets.

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Fragment Referencing Allows Identification of Components of a Mixture. During the synthesis of the tetraene **1** (scheme 1), a byproduct was isolated that was identified via multiple spectroscopic techniques as the decyanated tetraene **4** in which the nitrile has been replaced by a hydrogen. Decyanated tetraene **4** was subjected to the same alkene metathesis conditions as applied to the parent compound **1** (Scheme 2).

Scheme 2. Alkene Metathesis on Tetraene 4 To Give Biscyclopentene 5, *cis*-Tetradehydrodecalin 6c, and *trans*-Tetradehydrodecalin 6t; Fragment Referencing to Cyclohexene and *tert*-Butanol



A mixture of three compounds was obtained (molar ratio 9:6:2), which could be only partially separated by chromatography, into a mixture of 2 sets of two compounds. It was expected that the products would correspond to the three ring-closed products **5**, **6c**, and **6t**. Careful analysis of the spectra of the two mixed fractions allowed three individual experiment shift sets to be extracted. Were these the expected products, and which compound was which? Overlapping spin systems in the ^{1}H NMR spectra made analysis of the ^{13}C spectra the only viable approach. GIAO and CSGT ^{13}C NMR predicted shifts for all three isomers were therefore computed using TMS, methanol/benzene, and cyclohexene/*tert*-butanol as fragments for referencing (data not shown, see Supporting Information). The aim was not only to identify the products but also to again compare the three referencing techniques on a data set no longer containing the nitrile (the 'unusual' chemical shift of which had originally prompted our investigation of the CGST method and ultimately the fragment referencing approach; vide supra).

Using TMS as reference, both GIAO and CGST now gave good differentiation between all three isomers by DP4 analysis. The methanol/benzene reference system gave poorer levels of differentiation between isomers as measured by DP4 and in general poorer MAE for both GIAO and CGST. Referencing to cyclohexene and *tert*-butanol as fragments gave isomer differentiation by DP4 comparable to that using TMS but superior to that using MAE for both GIAO and CGST (Table 6).

In all cases, the biscyclopentene isomer **5** was most easily distinguished and assignment of the *cis*- and *trans*-tetradehydrodecalin isomers **6c** and **6t** was more delicate. Even in the absence of the possibly distorting nitrile ^{13}C peak, it was interesting to note that the CSGT method gave better fits than the GIAO method for both TMS and fragment referencing, although fragment referencing had superior MAE by either approach. It is also noteworthy that only fragment referencing attributes all three isomers correctly and confidently without recourse to empirical scaling (e.g., using the BMAE).

Given the success of the fragment referencing approach in the context of resolving product structure ambiguities for the two alkene metathesis reactions described above (Schemes 1 and 3), we decided to examine its utility more broadly. In particular, we selected to concentrate on cases involving low molecular weight molecules with relatively few carbon signals rather than more complex structures (e.g., many natural products), as these apparently simple systems are more commonly encountered by practicing synthetic chemists and, somewhat counterintuitively, more challenging from the perspective of current DFT NMR methods. The following examples have been selected to illustrate the use of fragment referencing to assist with product identification and assignment particularly for cases where standard incremental or HOSE

Table 6. CSGT and GIAO DFT (B3LYP/6-31G) ^{13}C NMR MAE Values for Isomers 5, 6c, and 6t Calculated Using TMS, MeOH/PhH, and Cyclohexene/*tert*-Butanol as References[a,b]**

			biscyclopentene **5**		*trans*-tetradehydrodecalin **6t**		*cis*-tetradehydrodecalin **6c**	
		data set	MAE	BMAE	MAE	BMAE	MAE	BMAE
TMS	CSGT	1	1.0	0.77	6.8	0.11	6.6	0.12
		2	8.9	0.07	1.0	0.60	1.8	0.33
		3	8.8	0.06	1.2	0.46	1.2	0.47
	GIAO	1	2.6	0.53	5.9	0.24	5.9	0.23
		2	9.4	0.14	2.9	0.44	3.0	0.42
		3	9.4	0.12	2.5	0.45	2.6	0.43
MeOH/PhH	CSGT	1	3.0	0.58	8.4	0.21	8.3	0.21
		2	7.2	0.15	2.6	0.40	2.4	0.45
		3	7.2	0.16	3.0	0.40	2.7	0.44
	GIAO	1	2.1	0.57	5.4	0.22	5.4	0.22
		2	9.9	0.11	2.5	0.44	2.4	0.45
		3	9.9	0.11	2.6	0.43	2.4	0.46
frag	CSGT	1	0.6	0.86	7.3	0.07	7.1	0.07
		2	8.4	0.05	0.6	0.72	1.8	0.23
		3	8.4	0.05	1.3	0.33	0.7	0.62
	GIAO	1	1.0	0.80	7.6	0.10	7.6	0.10
		2	7.7	0.04	0.5	0.65	1.0	0.31
		3	7.7	0.06	1.4	0.31	0.7	0.63

[a]All values are in ppm. [b]Underlined values represent the correct result. Note that only the fragment referencing is correct for all cases (although definitive proof of correctness via isolation and characterisation of all three compounds was not achieved).

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 7. Selected Parameters for CSGT and GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions for 1-Methyl-5-cyanoindole (8) Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 8)[a]**

	CSGT				GIAO			
	TMS		frag		TMS		frag	
	9	8	9	8	9	8	9	8
MAE[b]	2.5	1.2	2.1	0.7	4.1	3.8	2.4	0.5
BMAE[b]	0.32	0.68	0.26	0.74	0.45	0.55	0.17	0.83
DP4[b]	0.26	0.74	0.21	0.79	0.25	0.75	0.16	0.84

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Abbreviations: as for Table 1.

Table 8. Selected Parameters for CSGT and GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions for 1-Methyl-6-cyanoindole (9) Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 9)[a]**

	CSGT				GIAO			
	TMS		frag		TMS		frag	
	9	8	9	8	9	8	9	8
MAE[b]	1.3	2.3	1.1	2.0	3.6	3.8	0.8	1.9
BMAE[b]	0.63	0.37	0.65	0.35	0.55	0.45	0.71	0.29
DP4[b]	0.70	0.30	0.76	0.24	0.58	0.42	0.77	0.23

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Abbreviations: as for Table 1.

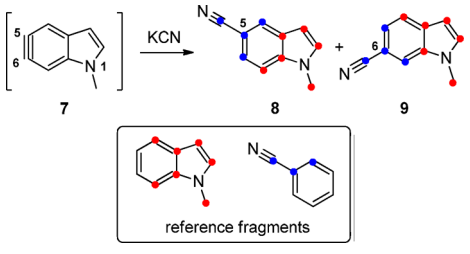
algorithms as implemented within, e.g., structure editing software packages lack the accuracy to resolve apparently routine but difficult problems. In these cases, synthetic chemists regularly resort to resynthesis and/or running complex NMR correlation experiments to aid product identification. Fragment referencing offers an enhanced level of absolute accuracy, which can be directly used to make a rapid and confident assignment, without knowledge of more complex statistical analysis techniques, such as the DP4 parameter.

Reliability of Chemical Shift Assignments Using Fragments. Accurately assigning a product requires accurate shift assignment. For example, assigning aromatic shifts when close together can sometimes require 2D-NMR correlation experiments and detailed analysis. If the predicted shifts are accurate to the extent that they are in the correct order, it is a simple matter to assign them using the predicted data. Empirical scaling or use of a single TMS reference does not allow predicted shifts to be reordered, in terms of numerical value, which can lead to incorrect assignments when no additional data are available.

As a first example, let us take the case of the cyanation of 5,6-indolyne **7** as reported by Cheong (Scheme 3).[58]

Cheong was able to assign compound **8** from the literature,[59] but if the ^{13}C shifts for the two cyanoindole products **8** and **9** are predicted using the CSGT method in conjunction with a TMS reference, the assignments for two pairs of atoms are switched as the errors exceed the shift difference between them.

Scheme 3. Cheong's 5,6-Indolyne Cyanation Reaction;[58] Fragment Referencing to *N*-Methylindole and Benzonitrile

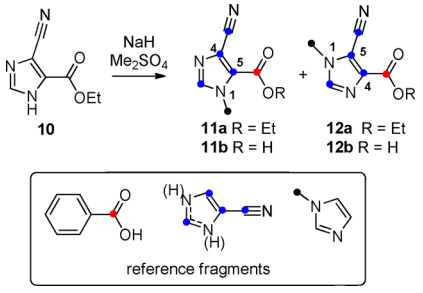


This problem is circumvented by using fragment referencing to *N*-methylindole and benzonitrile (data not shown, see Supporting Information). An analysis of the fragment referenced CSGT and GIAO shift predictions for both isomers is shown below (Tables 7 and 8).

The improved accuracy afforded by the fragment referencing allows not only unambiguous assignments but also improved distinction between isomers as evidenced by, e.g., BMAE analysis. Rapid assignment of isomers without recourse to synthesis or complex NMR correlation experiments is therefore possible using fragment referencing.

As a second example, let us take the case of the *N*-methylation of the ethyl cyanoimidazolecarboxylate **10** as reported by Subrayan (Scheme 4).[60]

Scheme 4. Subrayan's[60] Ethyl Cyanoimidazolecarboxylate *N*-Methylation Reaction[a]



[a]Note that DFT ^{13}C NMR calculations were performed on the acids **11b** and **12b** (for which experimental data was also reported by Subrayan) with fragment referencing to benzoic acid, 4-cyanoimidazole, and *N*-methylimidazole.

Subrayan resorted to hydrolysis, decarboxylation, and then NOE studies to assign the products, but DFT ^{13}C NMR on the acids **11b** and **12b** using fragment referencing to benzoic acid, 4-cyanoimidazole and *N*-methylimidazole, in this case in DMSO rather than $CHCl_3$, can also provide a confident result and with greater consistency than afforded by referencing to TMS (Tables 9 and 10).[61]

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 9. Selected Parameters for CSGT and GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions in DMSO for 4-Cyanoimidazole 11b Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 11b)[a]**

	CSGT				GIAO			
	TMS		frag		TMS		frag	
	11b	**12b**	**11b**	**12b**	**11b**	**12b**	**11b**	**12b**
MAE[b]	3.5	4.3	2.1	4.9	3.3	5.6	1.6	4.2
BMAE[b]	0.55	0.45	0.70	0.30	0.63	0.37	0.72	0.28
DP4[b]	0.48	0.52	0.64	0.36	0.56	0.44	0.63	0.37

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Abbreviations: as for Table 1.

Table 10. Selected Parameters for CSGT and GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions in DMSO for 5-Cyanoimidazole 12b Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 12b)[a]**

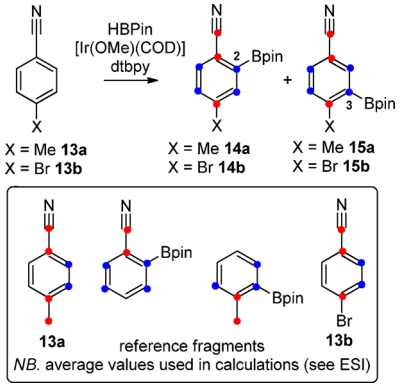
	CSGT				GIAO			
	TMS		frag		TMS		frag	
	11b	**12b**	**11b**	**12b**	**11b**	**12b**	**11b**	**12b**
MAE[b]	6.0	3.0	5.3	1.6	5.9	3.1	5.5	1.4
BMAE[b]	0.34	0.66	0.23	0.77	0.34	0.66	0.21	0.79
DP4[b]	0.37	0.63	0.24	0.76	0.40	0.60	0.30	0.70

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Abbreviations: as for Table 1.

The CSGT method in combination with TMS referencing notably predicts the incorrect 4-CN isomer, whereas fragment referencing predicts the correct isomer via the DP4 method (48% versus 64% probability, respectively, Table 9). Interestingly, there is a large relative improvement in the MAE for the correct isomer when using fragment referencing, while the incorrect isomer MAE is relatively stationary. This is not usually the case after empirical scaling; fragment referencing, however, often enhances the gap between the correct and incorrect isomer [e.g., GIAO for the 5-CN isomer, the 4-CN improves by just 0.3 ppm (5.9 → 5.6), while the correct 5-CN improves by 1.6 ppm (3.1 → 1.5) (Table 10)].

As a third example, let us take the case of iridium-catalyzed borylation of 4-substituted benzonitriles, which results in formation of both 2- and 3-borylated products, as reported by Smith (Scheme 5).[62]

Scheme 5. Smith's Ir-Catalyzed Borylation of 4-Substituted Benzonitriles 13a and 13b[62]



Smith generally identified the products of these reactions by counting the $^{3}J_{C-H}$ gHMBC NMR cross peaks. However, this method was not applicable for products with substituents having benzylic protons, which complicated the analysis. For the case of 4-methylbenzonitrile (**13a**), assignment of the major product as isomer **14a** required Smith to prepare an authentic sample via an alternate procedure.

DFT ^{13}C NMR peak shift prediction using 4-methylbenzonitrile (**13a**), 2-pinacolatobenzonitrile, and 2-pinacolatotoluene as fragments for referencing provides rapid and confident solutions for isomers **14a** and **15a**. Note that the experimental ^{13}C shifts for the carbon atoms *ipso* to boron were not available due to broadening and coupling (Table 11).

Table 11. Selected Parameters for GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions for 2- and 3-Borylated 4-Methyl Benzonitriles 14a and 15a Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 14a)[a]**

	TMS		frag	
	14a	**15a**	**14a**	**15a**
MAE[b]	2.4	3.4	0.4	2.5
BMAE[b]	0.59	0.41	0.86	0.14
DP4[b]	0.70	0.30	0.85	0.15

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Abbreviations: as for Table 1.

It can be seen that fragment referencing gives a much more confident BMAE result than when referencing using TMS without resorting to DP4 analysis, although the DP4 result is also more confident.

Fragment Referencing without DFT Geometry Optimization. Since the most time-limiting aspect of the process of ^{13}C NMR prediction with fragment referencing is the DFT geometry optimization, we were interested to see what could be achieved if this step was omitted and MM2 optimized geometries were used directly in the calculations (including for the fragments). The above borylation reaction (Scheme 5, above) was taken as a test case (Table 12).

As expected, the results were not quite as confident or accurate as with the DFT incorporated but nonetheless gave the correct result (BMAE = 71%) in comparing the 2- and 3-borylated isomers, while the TMS approach could not distinguish between the two. DP4 analysis gave the correct

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 12. Selected Parameters for GIAO ^{13}C NMR Shift Predictions for 2 and 3-Borylated 4-Methyl Benzonitriles 14a and 15a Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 14a) with Initial Geometry Optimization by only MM2 (Not DFT)[a]

	TMS		frag	
	14a	**15a**	**14a**	**15a**
MAE[b]	11.5	12.1	0.7	1.6
BMAE[b]	0.51	0.49	0.71	0.29
DP4[b]	0.53	0.47	0.60	0.40

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Abbreviations: as for Table 1.

result but was not very confident. Clearly, DFT optimizations should be carried out if more accurate absolute shift predictions are required, but if time and/or computational resources are short, then this short-cut method can be useful. Recently, Sarotti has advocated the use of artificial neural network (ANN) pattern recognition to improve the predictive capability of GIAO ^{13}C NMR predictions in order to circumvent the computational cost of DFT geometry optimization.[63] The integration of MM2 fragment referencing with ANNs would likely provide yet more accurate predictions while retaining the benefits of low computational cost.

Fragment Referencing Using a Non-DFT Method and the Influence of the Basis Set. Although DFT is the most widely employed computational method in the area of ^{13}C NMR prediction, we were also keen to evaluate the applicability of the fragment referencing approach in the context of a non-DFT method. We decided therefore to explore how fragment referencing performed when using the MP2 method[49,50] in combination with the 6-31G**[48] basis set (as used for the DFT calculations) and also with a series of correlation consistent (cc) basis sets developed by Dunning at al.: cc-pVDZ,[51] aug-cc-pVDZ,[52] and cc-pVTZ.[53] Unlike the 6-31G** basis set, these polarized valence double-ζ (pVDZ) and triple-ζ (pVTZ) basis sets include polarization functions by definition and in the case of the augmented (aug) variant some additional diffuse polarization terms. The indole cyanation reaction (Scheme 3, above) was selected for this comparative study in which GIAO[16,17] processing was applied in all cases (Tables 13 and 14).

The results confirm that fragment referencing is compatible with the MP2 method; the correct isomer was predicted for all the basis sets studied, with a minimum BMAE of 80% for isomer **8** and 71% for isomer **9**. By contrast, TMS-referenced shift prediction using the MP2 method cannot be used to distinguish between the isomers without empirical scaling for any of the basis sets studied (cf. BMAE values), and at the 6-31G** level, it is notable that the MP2 method delivers significantly poorer results than the corresponding B3LYP functional. Even after empirical scaling of the TMS-referenced data, MP2/6-31G** still produces unsatisfactory confidences (e.g., BCMAE 47% for isomer **8** and 47% for isomer **9**), although the improved basis sets do give confident predictions.

For fragment referencing, the MAE improves with improving basis set, indicating that if a correct spectrum prediction is required (rather than just an isomer comparison), investing in a higher quality basis set will improve the absolute accuracy, although the increasing computational cost does not lead to dramatic improvements in absolute accuracy. The quality of the fragments appears to be more critical in this regard. This is evidenced by examining the sources of greatest error in the predicted shifts (see Supporting Information), which reveals that the junction carbon atoms of the indole, which are referenced to indole and not corrected by incorporating the effect of the nitrile, are the least well predicted.

In all cases, the MAE calculated using fragment referencing is significantly lower than the CMAE generated by empirical scaling TMS-referenced data. On the evidence of this study, it is not worth investing in improved basis sets or methods for distinguishing two isomers: B3LYP/6-31G** in conjunction with fragment referencing gives the best absolute accuracy of all the combinations tested. The 'probability' confidences are fairly static, irrespective of method. Moreover, the BMAE calculated from fragment referenced data is the most confident source of discrimination (i.e., it is generally better than BCMAE or DP4) in the majority of methods used.

Further to the analysis presented in Chart 1, it is notable that the max error in a data set is always for the scaled-TMS data sets and not the fragment referenced data sets (ignoring raw TMS data). This highlights the ability of fragment referencing to overcome the problem that empirical scaling has of failing to cater for errors that fall outside of the assumed error-chemical shift trend.

Fragment Referencing Corrects the Shifts of Halogenated Atoms. The ^{13}C NMR shifts α to halogen atoms are generally very poorly predicted by the GIAO DFT ^{13}C NMR calculation method. This poor prediction is largely due to spin−orbit coupling effects and becomes increasingly disruptive for heavier atoms with large numbers of electrons.[64] Sometimes, an arbitrary correction is applied based on an empirically derived

Table 13. GIAO MP2 ^{13}C NMR Predicted Shifts[a] with 6-31G, cc-pVDZ, aug-cc-pVDZ, and cc-pVTZ Basis Sets for 1-Methyl-5-cyanoindole (8) Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 8), Including Comparison with the DFT (B3LYP/6-31G**) Results (See Table 7, above)[b]**

	MP2/6-31G**		MP2/cc-pVDZ		MP2/aug-cc-pVDZ		MP2/cc-pVTZ		B3LYP/6-31G**	
	frag ref	TMS ref	frag ref	TMS ref	frag ref	TMS ref	frag ref	TMS ref	frag ref	TMS ref
ATE[c]	8.1	96.4	7.7	58.5	7.0	56.2	7.5	88.6	4.8	29.4
MAE[c]	0.8	9.6	0.8	5.9	0.7	5.6	0.7	8.9	0.5	2.9
CMAE[c]	0.9	2.2	0.9	1.1	0.8	1.2	0.8	1.5	0.4	1.1
BMAE[c]	0.80	0.50	0.80	0.50	0.81	0.51	0.80	0.50	0.83	0.55
BCMAE[c]	0.79	0.47	0.78	0.74	0.80	0.71	0.78	0.64	0.85	0.74
DP4[c]	0.77	0.54	0.76	0.71	0.80	0.73	0.78	0.64	0.84	0.75

[a]Values shown are all for the correct isomer only. [b]All values are in ppm with the exception of the probabilities, BMAE, BCMAE, and DP4 (unitless fractions out of 1). [c]Abbreviations: as for Table 1.

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 14. GIAO MP2 ^{13}C NMR Predicted Shifts[a] with 6-31G, cc-pVDZ, aug-cc-pVDZ, and cc-pVTZ Basis Sets for 1-Methyl-6-cyanoindole (9) Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 9), Including Comparison with the DFT (B3LYP/6-31G**) Results (See Table 8, above)[b]**

	MP2/6-31G**		MP2/cc-pVDZ		MP2/aug-cc-pVDZ		MP2/cc-pVTZ		B3LYP/6-31G**	
	frag ref	TMS ref	frag ref	TMS ref	frag ref	TMS ref	frag ref	TMS ref	frag ref	TMS ref
ATE[c]	11.1	92.9	9.7	60.5	9.4	59.5	9.3	91.3	7.6	26.7
MAE[c]	1.1	9.3	1.0	6.0	0.9	5.9	0.9	9.1	0.8	2.7
CMAE[c]	1.2	2.4	1.0	1.5	1.0	1.4	1.0	1.5	0.6	1.2
BMAE[c]	0.71	0.50	0.74	0.50	0.74	0.50	0.75	0.50	0.71	0.55
BCMAE[c]	0.70	0.47	0.73	0.63	0.73	0.67	0.74	0.66	0.78	0.65
DP4[c]	0.59	0.43	0.65	0.63	0.69	0.75	0.71	0.69	0.77	0.58

[a]Values shown are all for the correct isomer only. [b]All values are in ppm with the exception of the probabilities, BMAE, BCMAE, and DP4 (unitless fractions out of 1). [c]Abbreviations: as for Table 1.

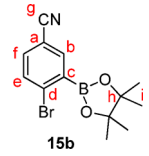
value,[64] but fragment referencing offers a tailored approach to correction for various halogenated environments.

Taking another example from Smith's iridium-catalyzed borylation study (Scheme 5), consider the case when 4-bromo benzonitrile (**13b**) is used as substrate and the major product is the 2-boryl isomer **14b**. The influence of the bromine atom on the aromatic system, particularly the carbon atom *ipso* to the Br atom, is readily corrected for by using 4-bromo benzonitrile (**13b**) as a fragment for referencing in place of its tolyl analogue **13a** (Table 15).

Table 15. GIAO DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts[a] for 3-Borylated 4-Bromo Benzonitrile 15b Calculated Using TMS and Fragment Referencing (Relative to Experimental Values for Compound 15b)[b]**

		fragments		TMS		scaled TMS[c]	
	exptl	calcd	Δ[d]	calcd	Δ[d]	calcd	Δ[d]
c	138.8	139.5	−0.7	135.7	3.1	138.6	0.2
f	134.5	135.0	−0.5	131.8	2.7	134.5	0.0
e	134.2	134.4	−0.2	129.9	4.3	132.5	1.7
d	127.1	126.7	0.4	142.5	−15.4	131.0	3.9
g	118.1	118.4	−0.3	113.8	4.3	115.7	2.4
a	115.8	116.2	−0.4	112.1	3.7	114.0	1.8
h	85.0	85.3	−0.3	87.3	−2.3	88.1	3.1
i	24.7	24.8	−0.1	25.6	−0.9	23.7	1.0
		fragments		TMS		scaled TMS	
ATE[e]		2.9		33.4		14.1	
MAE[e]		0.4		4.2		1.8	

[a]Carbon assignments as shown below. [b]All values are in ppm. [c]The scaled TMS data was processed by first applying a correction of 14 ppm to the C-Br atom (as described by Braddock and Rzepa, ref 64). The MAE value for the scaled TMS data is the same as the CMAE value for this data set. [d]Δ = difference in chemical shift relative to experiment. [e]Abbreviations: as for Table 1.



15b

In this case, the shifts when using TMS as reference have not been realigned into numerical order but matched with the known experimental assignments. The first four assignments are incorrect. The accuracy of the fragment referenced shifts is notable and is based on only two references: 4-bromobenzonitrile and 2-cyano phenylpinacolatoborane (Scheme 5). For the environments that appear in both molecules, the average value was used as the reference. For example, the average value of C3 in both reference molecules is used to calculate C3 in molecule **14b**. A 50:50 weighting was applied here, but a more precise weighting could be empirically derived with sufficient data. Interestingly, DP4 with TMS referencing predicts the wrong isomer (55% probability) for this example; fragment referencing corrects this, at 78% probability for the correct isomer (data not shown; see Supporting Information).

A more complex example is that of the assignment of the positions of the bromine and chlorine atoms within the structures of marine algae natural product obtusallene V (**16**) as studied by Braddock and Rzepa (Figure 3).

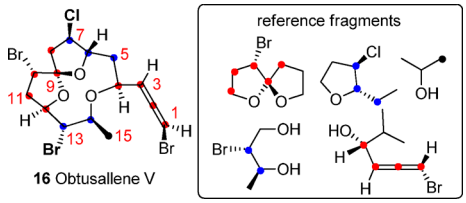


Figure 3. Revised structure of obtusallene V (**16**) as determined by Braddock and Rzepa.[64]

This compound was originally assigned as being the 7-bromo-13-chloro-regioisomer by Guella et al. in 2000[65] but was reassigned as being the 7-chloro-13-bromo-isomer by Braddock and Rzepa[64] in 2008 on the basis of a biosynthetic hypothesis and GIAO DFT ^{13}C NMR calculations. We selected this molecule to examine the utility of fragment referencing for environment-specific halogen corrections including the intriguing bromo-allene group in this natural product and also to explore how the method would perform for a more complex molecule than thus far examined.

The geometry of obtusallene V (**16**) was optimized at the B3LYP/6-31G** level as were the five selected reference fragments (Figure 3). Unlike when using TMS as reference, all of the predicted shifts were in the correct order with respect to the experimental assignments when applying fragment referencing (Table 16).

The halogenated atoms were well corrected for, and the MAE was only 1.5 ppm (cf. MAE_{TMS} = 5.7 ppm). The largest errors seen were within the chloro tetrahydrofuran ring. Just the one conformer used by Braddock and Rzepa was used for the obtusallene V calculation, but six conformers were used to generate reference values for the chloro tetrahydrofuran fragment. It was noticeable that altering the conformation of

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 16. GIAO DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for Obtusallene V (16) Calculated Using Fragment Referencing (Relative to Experimental Values for Compound 16)[a]**

assign (16)[b]	exptl	calcd	Δ[c]
C2	201.0	200.2	0.8
C9	116.4	115.5	0.9
C3	101.4	101.9	−0.5
C6	86.4	89.2	−2.8
C12	83.5	84.2	−0.7
C14	77.4	77.9	−0.5
C4	74.7	77.4	−2.7
C1	73.0	74.9	−1.9
C7	60.3	57.0	3.3
C13	57.0	56.0	1.0
C10	49.7	51.4	−1.7
C8	46.4	44.9	1.5
C11	42.9	43.1	−0.2
C5	37.1	37.7	−0.6
C15[d]	22.5	21.9	0.6
			Δ
ATE[e]			19.1
MAE[e]			1.3

[a]All values are in ppm. [b]Carbon assignments as in Figure 3. [c]Δ = difference in chemical shift relative to experiment. [d]This shift is referenced to isopropanol, referencing to 2-bromo-1,3-butanediol (cf. Figure 3) gave a less good fit (see Supporting Information). [e]Abbreviations: as for Table 1.

the THF ring slightly caused significant variations in the reference values. The internal angle of the chloro tetrahydrofuran ring hosting the alkyl group in obtusallene is 105.5°, while the fragment conformers used to reference this portion of the molecule have values as low as 102.1° for the corresponding angle. Thus a limitation of fragment referencing is the availability of a good fragment for unusual or strained systems.

Cross-Fragment Referencing and Diagnostic Chemical Shift Analysis. We have shown that isomers similar in structure to ones being simulated constitute good reference 'fragments'. In this context, a common scenario is one in which two sets of experimental data are available for two isomers and it seems plausible that one isomer could act as the reference for the other, but at the outset we do not know which experimental and predicted shifts correlate in order to generate the fragments.

Take the case of the two isomeric aurones (*E*)-**17** and (*Z*)-**17**, which were prepared by synthesis to corroborate a proposed natural product structure by Venkateswarlu (Figure 4).[66,67]

In 2001, Rahman assigned a compound isolated from an Arabian sea algae as being aurone (*Z*)-**17**,[67] but in 2007 Venkateswarlu prepared both (*E*)-**17** and (*Z*)-**17** by synthesis and showed that neither displayed data consistent with the natural product.[66] Assignment of the stereochemistry of the two synthetic isomers was, however, made on the basis of ^{1}H and ^{13}C NMR trends derived from related compounds.

Say we designate the two experimental ^{13}C NMR data sets for (*E*)-**17** and (*Z*)-**17** as A and B. We can also compute two predicted ^{13}C NMR data sets. Assuming we are sure the two predicted isomers match the experimental sets one way or the other, there are two possible permutations. Let us define A(*E*) and A(*Z*) as the potential fragment reference shifts for

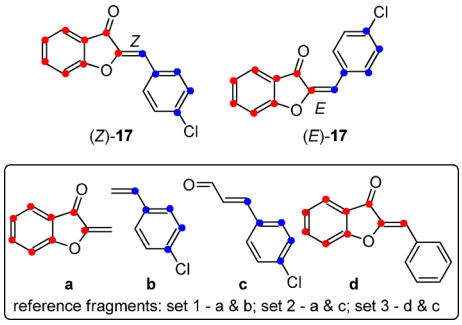


Figure 4. Venkateswarlu's two aurone isomers considered as candidates for a marine natural product.[66]

experimental data set A, where we have used predicted sets for the (*E*) or (*Z*) isomers, respectively, to calculate a set of reference shifts. Conversely, B(*E*) and B(*Z*) are the two possibilities for experimental data set B. As we are sure that one of these must be correct, we can now say that one of the following must be true:

- Permutation 1: A(*E*) gives the true fragment references for A, in which case B(*Z*) is also correct
- Permutation 2: A(*Z*) gives the true fragment references for A, in which case B(*E*) is also correct

We can now test both these fragment sets on the predicted data for the other isomer. So, we can use A(*E*) to reference the (*Z*) predicted shifts and B(*Z*) to reference the (*E*) predicted shifts *and* use B(*E*) to reference the (*Z*) predicted shifts and A(*Z*) to reference the (*E*) predicted shifts. One of these combinations will *both* generate the correct fragments and match the data up correctly (Table 17).

Table 17. Selected Parameters for DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions for Aurones (*E*)-17 and (*Z*)-17 Calculated Using Cross-Fragment Referencing[a]**

	permutation 1		permutation 2	
	a[B(*Z*)]	b[A(*E*)]	a[B(*E*)]	b[A(*Z*)]
	exptl set A[b]			
MAE[c]	0.7	1.5	1.5	2.7
BMAE[c]	0.68	0.32	0.65	0.35
Bayes' method[d]	0.66		0.34	
	exptl set B[b]			
MAE[c]	1.5	0.7	2.7	1.4
BMAE[c]	0.31	0.69	0.34	0.66
Bayes' method[d]	0.66		0.34	

[a]All values are in ppm with the exception of the probability BMAE (unitless fractions out of 1). [b]Fragment sets are defined in Figure 4. [c]Abbreviations: as for Table 1. [d]Probability that each permutation was correctly matched up by comparing the totals of MAE values for each permutation using Bayes' theorem.

It is clear that permutation 1 gives the best accuracy in predicting the shifts, and this is because it generates the best fragment reference values by combining the correlated experimental and predicted data. Additionally, within each experimental set, the normal procedure in competing predicted data (after fragment referencing) confirms that data set A correlates to the (*E*) isomer, and B to the (*Z*) isomer.

The basic principle used in this 'cross-fragment referencing' is very similar to the initial steps used in the CP3 parameter developed by Goodman and Smith.[12] They calculate directly

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

the difference in the predicted shifts (δa − δb) and experimental shifts (δA − δB) and compare this with the opposite assignment, the difference in the predicted shifts (δa − δb) and experimental shifts (δB − δA). They then include these values in various equations to assign a probability that a match is correct or incorrect, by simply observing that for the incorrect assignment, the sign (±) should be inverted for the incorrect match.

Although this approach is based on the principle of differences, we have altered the method. First, we have made an assumption about the data. Even for correct fragment matches (i.e., when using a genuine pair of predicted and experimental data to generate reference shifts), we do not specifically know whether we have aligned the correct experimental shifts with the corresponding predicted shifts. Indeed, we have already observed that shifts are sometimes reordered after fragment referencing. However, for the 'incorrect' permutation, we can expect the error to be even larger as the reference values themselves will likely be incorrect. Additionally, peaks that switch around in numerical order tend to be close together to start with, meaning that the error introduced to a reference peak will also likely be small (note that after referencing, we reorder the peaks into numerical order again, so if two assignments were in the incorrect order by an unreferenced difference of 0.2 ppm, the error associated with using the wrong value will be 0.2 ppm even after faulty referencing). Thus we can be reasonably confident that lining up the two sets of data numerically will lead to a distinction in the accuracy of the calculated shifts, either by method or by assignment, but probably by both.

Although cross-fragment referencing is clearly sufficient to obtain a good result for aurones (*E*)-**17** and (*Z*)-**17**, we also examined how particular fragments influence the accuracy of the analysis. The following pairs of references were selected for study: set 1 = fragments **a** and **b**; set 2 = fragments **a** and **c**; set 3 = fragments **d** and **c** (Figure 4, Table 18).

The accuracy of set 1 (lactone **a** and 4-chlorostyrene **b**) was good, but these fragments lead to poor treatment of the key diagnostic alkene carbons. The predictions for these carbons is improved by employing lactone **a** and (*E*)-4-chlorocinnimaldehyde (**c**) (set 2) and still further by using unsubstituted (*Z*)-aurone **d** and (*E*)-4-chlorocinnimaldehyde (**c**) (set 3) as fragments. It is noticeable that the accuracy of all carbons benefit slightly from the fragment upgrades, indicating more extensive fragments are beneficial for systems with extended conjugation. No apparent bias was observed when using (*Z*)-aurone **d** as a fragment reference for (*E*)-aurone **17**; a confident result was generated by both DP4 and BMAE analysis. Moreover, the errors in predicting the carbons α to the chlorine were only −0.3 and −0.1 ppm for the (*Z*)- and (*E*)-isomers, respectively, using the cinnamaldehyde fragment (data not shown; see Supporting Information).

A technique sometimes used in the literature to assign (or reassign) the structure of natural products is to use one particular diagnostic ^{13}C signal. For example, Williams et al. showed by DFT computation that the experimental ^{13}C chemical shift for the central carbon of a proposed allene-containing isolate from the bark of *Brosimum acutifolium* Huber was inconsistent with this environment and reassigned the isolate as known non-allenic natural product mururin C.[68]

This diagnosic shift approach could in principle be applied to the vinyl methine carbons in the (*E*)-and (*Z*)-aurone isomers (i.e., carbons 'i', below). However, at the outset we have no knowledge of which experimental data points correspond to which carbon environments. Therefore, we have little choice but to arrange the experimental data for isomeric aurones (*E*)-**17** and (*Z*)-**17** in numerical order and align each predicted shift in order alongside. If we do this for the shifts predicted using basic fragment referencing (set 1) we obtain the following isomer specific assignments (Table 19).

Table 18. Selected Parameters for GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions for Aurones (*E*)-17 and (*Z*)-17 Calculated Using Different Fragment Reference Sets (Relative to Experimental Values for Compounds (*E*)-17 and (*Z*)-17)[a]**

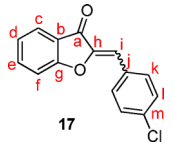
	(*Z*)-**17**					
	frag set 1[b]		frag set 2[b]		frag set 3[b]	
	Z	*E*	*Z*	*E*	*Z*	*E*
MAE[c]	1.5	2.4	1.2	2.2	0.6	1.7
BMAE[c]	0.62	0.38	0.64	0.36	0.76	0.24
DP4[c]	0.43	0.57	0.71	0.29	0.90	0.10
	(*E*)-**17**					
	frag set 1[b]		frag set 2[b]		frag set 3[b]	
	E	*Z*	*E*	*Z*	*E*	*Z*
MAE[c]	1.4	1.4	1.2	1.2	0.6	1.3
BMAE[c]	0.49	0.51	0.50	0.50	0.67	0.33
DP4[c]	0.61	0.39	0.67	0.33	0.69	0.31

[a]All values are in ppm with the exception of the probabilities, BMAE and DP4 (unitless fractions out of 1). [b]Fragment sets are defined in Figure 4. [c]Abbreviations: as for Table 1.

Table 19. Signal Assignment for the Aurone Isomers (*E*)-17 and (*Z*)-17 Following GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Predictions Calculated Using Fragment Reference Set 1[a]**

exptl shifts (refs 66 and 67)		predicted assignment[b]	
(*E*)-**17**	(*Z*)-**17**	(*E*)	(*Z*)
124.7	124.3	i	c
123.2	123.2	b	b
122.9	121.0	c	d
121.1	112.5	d	i
112.6	111.1	f	f

[a]All values are in ppm. [b]Carbon assignments as shown below.



17

Without prior knowledge of which experimental data set is which, it is not possible to compare the predicted shifts for the vinyl methine carbons (i.e., carbons 'i') in the respective isomers; which data point should they compete for, the one at 124.7 ppm or that at 112.5 ppm?

Clearly, using diagnostic shifts is difficult in cases like this where the peak assignments are not secure. Let us therefore compute a selection of potentially diagnostic GIAO predicted shifts for both the (*E*)- and (*Z*)-isomers, competing for the experimental data for aurone (*Z*)-**17** (Table 20).

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

Table 20. GIAO DFT (B3LYP/6-31G) ^{13}C NMR Predicted Shifts for Aurones (*E*)-17 and (*Z*)-17 Calculated Using TMS and Fragment Reference Set 1 (Relative to Experimental Values for Compound (*Z*)-17)[a]**

	(*E*)-prediction	Δ[c]	(*Z*)-prediction	Δ[c]
		TMS ref[b]		
c	122.0	2.3	121.2	3.1
b	121.0	2.2	119.2	4.0
d	121.4	−0.4	119.0	2.0
i	118.1	−5.6	111.0	1.5
f	108.2	2.9	108.5	2.6
		frag ref[b]		
c	125.8	−1.5	124.5	−0.2
b	125.2	−2.0	123.4	−0.2
d	124.7	−3.7	123.3	−2.3
i	122.5	−10.0	114.8	−2.3
f	112.5	−1.4	112.8	−1.7

[a]All values are in ppm. [b]Carbon assignments as for Table 19. [c]Δ = difference in chemical shift relative to experiment.

For the TMS referenced data, the predicted shifts for the vinyl methane carbons in the (*E*)- and (*Z*)-isomers apparently have discrepancies of −5.6 and +1.5 ppm, respectively, relative to the experimental data for (*Z*)-aurone **17**. Not only is this level of discrimination insufficient to make a confident prediction based on this diagnostic shift, it is also erroneous because of the unwitting and incorrect assignment of the predicted shift at 118.1 ppm as being that of the (*Z*)-isomer vinyl methane carbon (when it should be that at 122.0 ppm; which would give a 'true' discrepancy of −9.5 ppm).

For the fragment referenced data, the predicted shifts for the vinyl methane carbons in the (*E*)- and (*Z*)-isomers apparently have discrepancies of −10.0 and −2.3 ppm, respectively, relative to the experimental data for (*Z*)-aurone **17**. With the added confidence of knowing we have reduced environment-sensitive error (and hence also reduced the probability of misassignment), we can now be confident enough, given the good accuracy of the rest of the Z-isomer predicted shifts, to use this shift as a diagnostic tool.

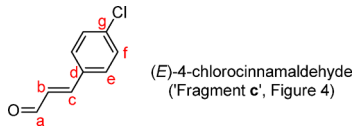
Fragment Selection. The basic requirement for a reference fragment is that experimental ^{13}C NMR data are available in the appropriate solvent, either from a literature reference or from proprietary spectra available within the laboratory in question. Beyond that, selection of suitable references should be driven by the aim to select those available fragments that are perceived to present the chemical/magnetic environments most similar to those of the molecule for which DFT ^{13}C NMR data is being calculated. It is likely the case that particularly good references will be available from prior studies in the laboratory concerned. Typically, during target-driven synthesis, the substrate in a reaction will constitute a good reference for a large part of the product, and similarly, during methodology development multiple analogues are often produced, at least one of which will likely provide a suitable reference compound (e.g., in the iridium-catalyzed borylation of 4-methylbenzonitrile, Scheme 5). This means that fragment referencing is likely to be readily implementable using data available within a laboratory and moreover these data will likely be assigned. This is not always the case with data taken from the literature, so how can we take an unassigned list of potential fragment experimental shifts and combine them with the correct predicted shifts so as to correctly generate the fragment references? For the majority of low molecular weight compounds, there are relatively few signals and they are unambiguous to align (i.e., they do not lie close enough together to be at risk of being predicted in the wrong order), but for more advanced fragments, it is sometimes useful to consult a smaller, less ambiguous fragment first. This gives an approximate value for the error, which can be tuned in the full fragment.

Take for example the case of (*E*)-4-chlorocinnamaldehyde (fragment **c**, Figure 4). It might initially be uncertain which experimental shift is attributable to the carbon *ipso* to the chlorine atom and which to the *β*-alkenyl carbon of the enal. This would be easily solved if 2D NMR spectra were available, but in their absence, a comparison to a simpler fragment, e.g., 4-chloro styrene (fragment **b**, Figure 4) can be made. Knowing that the carbon *ipso* to the chlorine atom in 4-chlorostyrene has a reference shift of 188.0 ppm (with no ambiguity as no enal is present), the 4-chlorocinnamaldehyde experimental shifts at 137.8 and 145.0 ppm can be assigned as shown below (Table 21).

Table 21. Selected Fragment Referenced GIAO DFT (B3LYP/6-31G) ^{13}C NMR Shift Assignments for (*E*)-4-Chlorocinnamaldehyde *after* Correlation with 4-Chlorostyrene To Allow C−Cl Assignment by Expected Error Fitting[a]**

		incorrect		correct	
GIAO calcd	assign[b]	exptl	σ_{REF}	exptl	σ_{REF}
30.6	a	168.8	199.4	168.8	199.4
51.0	g	145.0	196.0	137.8	188.8
51.1	c	137.8	188.9	145.0	196.1
63.2	e	134.5	197.7	134.5	197.7

[a]All values are in ppm. [b]Carbon assignments as shown below.



(*E*)-4-chlorocinnamaldehyde ('Fragment **c**', Figure 4)

We are simply using our knowledge of the expected error range to help with the assignment. This 'knowledge' accumulates rapidly: as more calculations are performed under the same conditions the larger the database becomes. It is likely that if fragment referencing becomes widespread and routine, then many error corrections will be available 'off the shelf'. Toward this end, we have provided a table of specific environment reference values in the Supporting Information for B3LYP/6-31G**, CPCM: $CHCl_3$.

CONCLUSIONS

It is hoped that we have convincingly demonstrated that the accuracy of ^{13}C NMR spectra prediction by Density Functional Theory (DFT) can be usefully enhanced by employing chemical shifts using 'fragment referencing', without recourse to empirical scaling. Fragment referencing refers to a process of reducing the error in calculating a particular NMR shift by consulting a similar molecule for which the error in the calculation is easily deduced. The method can be considered the quantum mechanical counterpart of incremental NMR prediction. Such incremental prediction tools use a base value, which is corrected depending on the circumstances, using empirical data to apply the corrections. Fragment referencing corrects for calculation error based on the same principle;

dx.doi.org/10.1021/jo401833b | J. Org. Chem. 2013, 78, 11302−11317

similar chemical environments are expected to have similar errors. The absolute accuracy of the chemical shifts predicted when employing fragment referencing relative to conventional techniques (e.g., using TMS or MeOH/benzene dual referencing) has been shown to be improved significantly in a range of case studies that illustrate the superiority of the technique particularly for systems with similar chemical shifts arising from different chemical environments. The technique is particularly suited to molecules of relatively low molecular weight containing 'non-standard' magnetic environments, e.g., α to halogen atoms, which are poorly predicted by other methods. The simplicity, speed, and accuracy of the technique mean that it can be employed to definitively resolve routine structural assignment problems that cannot be solved using standard incremental or HOSE algorithms as currently implemented in many chemical drawing packages.

Although fragment referencing is applicable also to non-DFT methods such as MP2 and can be applied with more sophisticated basis sets than the 6-31G** basis set used in our DFT B3LYP calculations, we have shown that at least in the case we have examined, the additional computational expense is not justified in terms of increased accuracy and that time spent identifying a judicious set of fragments is more profitable.

EXPERIMENTAL SECTION

General Experimental Methods. All reactions were performed under an atmosphere of nitrogen and anhydrous conditions in flame- or oven-dried glassware unless stated otherwise. Yields refer to chromatographically homogeneous materials, unless otherwise indicated. All reagents were used as received from commercial suppliers. Anhydrous solvents were used directly following passage under nitrogen through Al_2O_3 columns in a commercial SPS. Flash chromatography (FC) was performed on silica gel. Thin layer chromatography (TLC) was performed on aluminum-backed plates precoated with silica (0.2 mm). Infrared (IR) spectra were recorded neat on a Fourier transform spectrometer. Only selected absorbances (υ_{max}) are reported. ^{1}H NMR spectra were recorded at ambient temperature at 400 MHz. Chemical shifts (δ_H) are quoted in parts per million (ppm), referenced to the residual solvent peak ($CDCl_3$ at δ_H 7.27 ppm). The multiplicities of ^{1}H signals are designated by the following abbreviations: s = singlet; d = doublet; dd = doublet of doublets; ddt = doublet of doublets of triplets; t = triplet; q = quartet; br = board; m = multiplet. Coupling constants, J, are reported to the nearest 0.1 Hz. ^{13}C NMR spectra were recorded at 100 MHz at ambient temperature. Chemical shifts (δ_C) are quoted in ppm referenced to ($CDCl_3$ at δ_C 77.1 ppm). Low resolution mass spectra (m/z) were recorded on a quadrupole spectrometer; molecular ions (M^+, MH^+, MNH_4^+) are reported. High resolution mass spectrometry (HRMS) measurements were recorded on a quadrupole spectrometer and are valid to ±5 ppm.

*1′-Hydroxy-[1,1′-bi(cyclopentane)]-3,3′-diene-1-carbonitrile (**2**)*[26] *and trans-**8a**-Hydroxy-1,4,4a,5,8,8a-hexahydronaphthalene-4a-carbonitrile (**3t**).*[54] For details of the reaction forming these products, see ref 26.

*4,5-Diallylocta-1,7-dien-4-ol (**4**).* Magnesium turnings (252 mg, 10.36 mmol) and a small crystal of iodine were placed in a two-necked round-bottom flask (50 mL) fitted with a reflux condenser and pressure-equalizing dropping funnel charged with allyl bromide (784 mg, 6.48 mmol) in tetrahydrofuran (20 mL). The reaction mixture was heated to reflux, and a few drops of the allyl bromide solution and THF (3 mL) were added to initiate formation of the Grignard reagent. The ethyl 2-allyl-2-cyanopent-4-enoate[69] (500 mg, 2.59 mmol) was then added in one portion followed by the remainder of the allyl bromide (784 mg, 6.48 mmol) dropwise over 2 h. After a further 1 h of stirring at reflux, the reaction mixture was quenched with satd aqueous ammonium chloride solution (40 mL), extracted with CH_2Cl_2, washed with saturate sodium bicarbonate solution, dried ($MgSO_4$), filtered, and concentrated in vacuo. The residue was purified by flash column chromatography eluting with 12:1 pet ether/EtOAc to give 4,5-diallylocta-1,7-dien-4-ol **4** as a yellow oil (150 mg, 25%). IR (neat) ν_{max} 3008, 2978, 2916, 1638 (C=C), 1437, 995, 908 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 5.90 (tdd, J 16.6, 9.9, 7.9, 4H, =CH), 5.20–5.00 (m, 8H, $=CH_2$), 2.42–2.34 (m, 4H, 2 × CH_2), 2.27 (dd, J 14.5, 7.6, 2H, 2 × C*H*H), 2.11 (dt, J 14.5, 7.6, 2H, 2 × CH*H*), 1.75 (tt, J 7.9, 4.4, 1H, CH). ^{1}H NMR (acetone-d_6, 400 MHz): δ 5.93 (dddt, J 28.8, 17.1, 10.2, 7.1, 4H), 5.15–4.90 (m, 8H), 3.17 (s, 1H), 2.48–2.35 (m, 4H), 2.32–2.24 (m, 2H), 2.13–2.03 (m, 4H), 1.71 (tt, J 8.1, 4.1, 1H). ^{13}C NMR ($CDCl_3$, 100 MHz): δ 138.5, 133.9, 118.6, 115.9, 76.3, 44.8, 41.6, 33.7. MS(CI$^+$) m/z 224 (MNH_4^+); HRMS (CI$^+$) calcd for $C_{14}H_{26}NO$ (MNH_4^+) 224.2014, found 224.2023, Δ = 4.0 ppm.

*1-Cyclopent-3-en-1-yl)cyclopent-3-en-1-ol (**5**), cis-1,4,4a,5,8,8a-Hexahydronaphthalen-4a-ol (**6c**), and trans-1,4,4a,5,8,8a-Hexahydronaphthalen-4a-ol (**6t**).*[70] To a stirred solution of 4,5-diallylocta-1,7-dien-4-ol **4** (15 mg, 0.06 mmol) in CH_2Cl_2 (0.5 mL) was added Grubbs II catalyst (2 mol %). The reaction was stirred at rt overnight, and then the solvent was removed by a stream of nitrogen. The residue was analyzed by ^{1}H NMR and found to contain the three products **5**, **6c**, and **6t** in a ratio 9:6:2. Attempted purification of these by flash column chromatography eluting with 1:15 EtOAc/pet ether gave two fractions, each containing two of the products: fraction 1 (6 mg) containing **5** and **6c** and fraction 2 (5 mg) containing **6c** and **6t**.

Careful analysis of the ^{13}C NMR spectra for the above mixtures allowed the following three data sets to be extracted. The assignments followed from the computational analysis; vide infra.

Biscyclopentene **5**: ^{13}C NMR ($CDCl_3$, 100 MHz): δ 130.0, 128.7, 83.2, 46.1, 45.8, 34.9 ppm.

cis-Tetradehydrodecalin **6c**: ^{13}C NMR ($CDCl_3$, 100 MHz): δ 125.3, 124.1, 69.7, 36.7, 36.4, 29.5 ppm.

trans-Tetradehydrodecalin **6t**: ^{13}C NMR ($CDCl_3$, 100 MHz): δ 126.6, 123.9, 67.8, 39.4, 34.6, 29.4 ppm; [cf. lit.[70] ($CDCl_3$, 20 MHz): δ 126.5, 123.9, 67.7, 39.6, 34.8, 29.4 ppm].

ASSOCIATED CONTENT

Supporting Information

^{1}H and ^{13}C NMR spectra for compound **4** and full computational details for all the cases referred to in this manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: a.c.spivey@imperial.ac.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Imperial College London is thanked for financial support of this work. We also thank Prof. Henry S. Rzepa, Imperial College London for useful discussions regarding this work.

REFERENCES

(1) Helgaker, T.; Jaszuński, M.; Ruud, K. *Chem. Rev.* **1998**, *99*, 293–352.
(2) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744–3779.
(3) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839–1862.
(4) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357–7363.
(5) Barone, G.; Gomez-Paloma, L.; Duca, D.; Silvestri, A.; Riccio, R.; Bifulco, G. *Chem.—Eur. J.* **2002**, *8*, 3233–3239.
(6) Li, Y.; Gao, H.; Zhang, J.; Li, S.; Zhou, W. *Magn. Reson. Chem.* **2012**, *50*, 16–21.
(7) Schindler, M. *J. Am. Chem. Soc.* **1987**, *109*, 1020–1033.

(8) Schleyer, P. v. R.; Maerker, C. *Pure Appl. Chem.* **1995**, *67*, 755−760.
(9) Rasul, G.; Chen, J. L.; Prakash, G. K. S.; Olah, G. A. *J. Phys. Chem. A* **2010**, *114*, 4394−4399.
(10) Rauhut, G.; Puyear, S.; Wolinski, K.; Pulay, P. *J. Phys. Chem.* **1996**, *100*, 6310−6316.
(11) Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946−12959.
(12) Smith, S. G.; Goodman, J. M. *J. Org. Chem.* **2009**, *74*, 4597−4607.
(13) Suyama, T. L.; Gerwick, W. H.; McPhail, K. L. *Bioorg. Med. Chem.* **2011**, *19*, 6675−6701.
(14) Blinov, K. A.; Smurnyy, Y. D.; Churanova, T. S.; Elyashberg, M. E.; Williams, A. J. *Chemomet. Intel. Lab. Syst.* **2009**, *97*, 91−97.
(15) Meiler, J.; Maier, W.; Will, M.; Meusinger, R. *J. Magn. Reson.* **2002**, *157*, 242−252.
(16) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789−807.
(17) Gauss, J. *J. Chem. Phys.* **1993**, *99*, 3629−3643.
(18) Schindler, M.; Kutzelnigg, W. *J. Chem. Phys.* **1982**, *76*, 1919−1933.
(19) Keith, T. A.; Bader, R. F. W. *Chem. Phys. Lett.* **1993**, *210*, 223−231.
(20) Facelli, J. C. *Concept. Magn. Reson. A* **2004**, *20A*, 42−69.
(21) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497−5509.
(22) Migda, W.; Rys, B. *Magn. Reson. Chem.* **2004**, *42*, 459−466.
(23) Bagno, A.; Rastrelli, F.; Saielli, G. *Chem.—Eur. J.* **2006**, *12*, 5514−5525.
(24) Aliev, A. E.; Courtier-Murias, D.; Zhou, S. *J. Mol. Struct.: THEOCHEM* **2009**, *893*, 1−5.
(25) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254−7260.
(26) Andrews, K. G.; Frampton, C. S.; Spivey, A. C. *Acta Crystallogr., Sect. C* **2013**, *69*, 1207−1211.
(27) Rozhenko, A. B.; Trachevsky, V. V. *Phosporus, Sulfur Silicon Relat. Elem.* **2009**, *184*, 1386−1405.
(28) Timmons, C.; Wipf, P. *J. Org. Chem.* **2008**, *73*, 9168−9170.
(29) Benzi, C.; Crescenzi, O.; Pavone, M.; Barone, V. *Magn. Reson. Chem.* **2004**, *42*, S57−S67.
(30) Schuler, R. H.; Albarran, G.; Zajicek, J.; George, M. V.; Fessenden, R. W.; Carmichael, I. *J. Phys. Chem. A* **2002**, *106*, 12178−12183.
(31) Dybiec, K.; Gryff-Keller, A. *Magn. Reson. Chem.* **2009**, *47*, 63−66.
(32) Baldridge, K. K.; Siegel, J. S. *J. Phys. Chem. A* **1999**, *103*, 4038−4042.
(33) Gauss, J.; Stanton, J. F. In *Calculation of NMR and EPR Parameters*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH Verlag: Weinheim, 2004; p 123.
(34) Jain, R.; Bally, T.; Rablen, P. R. *J. Org. Chem.* **2009**, *74*, 4017−4023.
(35) Forsyth, D. A.; Sebag, A. B. *J. Am. Chem. Soc.* **1997**, *119*, 9483−9494.
(36) Chesnut, D. B. *Chem. Phys.* **1997**, *214*, 73−79.
(37) d'Antuono, P.; Botek, E.; Champagne, B.; Spassova, M.; Denkova, P. *J. Chem. Phys.* **2006**, *125*, 144309−12.
(38) Dračínský, M.; Bouř, P. J. *Chem. Theory Comput.* **2010**, *6*, 288−299.
(39) Sharma, A.; Reva, I.; Fausto, R. *J. Phys. Chem. A* **2008**, *112*, 5935−5946.
(40) Gortari, I. D.; Portella, G.; Salvatella, X.; Bajaj, V. S.; van der Wel, P. C. A.; Yates, J. R.; Segall, M. D.; Pickard, C. J.; Payne, M. C.; Vendruscolo, M. *J. Am. Chem. Soc.* **2010**, *132*, 5993−6000.
(41) Bagno, A.; Rastrelli, F.; Saielli, G. *J. Phys. Chem. A* **2003**, *107*, 9964−9973.
(42) Smith, S. G.; Paton, R. S.; Burton, J. W.; Goodman, J. M. *J. Org. Chem.* **2008**, *73*, 4053−4062.
(43) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127−8134.
(44) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587−615.
(45) Wiitala, K. W.; Al-Rashid, Z. F.; Dvornikovs, V.; Hoye, T. R.; Cramer, C. J. *J. Phys. Org. Chem.* **2007**, *20*, 345−354.
(46) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(47) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(48) Hehre, W. J. R., L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(49) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−22.
(50) Head-Gordon, M.; Head-Gordon, T. *Chem. Phys. Lett.* **1994**, 220.
(51) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−23.
(52) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358−71.
(53) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−806.
(54) Webber, M. J.; Warren, S. A.; Grainger, D. M.; Weston, M.; Clark, S.; Woodhead, S. J.; Powell, L.; Stokes, S.; Alanine, A.; Stonehouse, J. P.; Frampton, C. S.; White, A. J. P.; Spivey, A. C. *Org. Biomol. Chem.* **2013**, *11*, 2514−2533.
(55) Whilst not strictly a probability, Bayes' CMAE (BCMAE) and Bayes' MAE (BMAE)) are used as a convenient indicator of the relative errors of competing structures. Bayes' method reveals how much of the parameter error in the experiment is attributed to a given candidate structure, relative to the competition (so all values for a given parameter total 100%). Note that there is an inverse operator that ensures a value of greater than 50% (for two data sets) indicates a lower error. Thus, a higher percentage correlates to a higher degree of confidence within that experiment. Note that BCMAE/BMAE values between experiments with different chemical systems should not be compared as relative confidences, as the magnitude of the BCMAE/BMAE value is influenced by the number of data points and the chemical system. Different calculation methods and fragment systems for a given system should however provide comparable BCMAE/BMAE values, allowing *methods* to be analyzed for superiority.
(56) Reis, A. K. C. A.; Oliveira, P. R.; Rittner, R.; Tormena, C. F. *Ann. Magn. Reson.* **2005**, *4*, 1.
(57) If the predicted value for isomer B must pass the real data point for A, there is the unfortunate case in which an under corrected point B may end up close to A; this is an inevitable scenario and applies to the data fitting method too.
(58) Cheong, P. H. Y.; Paton, R. S.; Bronner, S. M.; Im, G. Y. J.; Garg, N. K.; Houk, K. N. *J. Am. Chem. Soc.* **2010**, *132*, 1267−1269.
(59) Bronner, S. M.; Bahnck, K. B.; Garg, N. K. *Org. Lett.* **2009**, *11*, 1007−1010.
(60) Subrayan, R. P.; Thurber, E. L.; Rasmussen, P. G. *Tetrahedron* **1994**, *50*, 2641−2656.
(61) Calculations were performed on acids **11b** and **12b** rather than ethyl esters **11a** and **12a**; first, because the lower conformational mobility of the acids was practically advantageous, and second, because the acids present a greater challenge to the fragment referencing method − compounds with fewer data points are less amenable to empirical scaling methods, making absolute error comparisons more important.
(62) Chotana, G. A.; Rak, M. A.; Smith, M. R. *J. Am. Chem. Soc.* **2005**, *127*, 10539−10544.
(63) Sarotti, A. M. *Org. Biomol. Chem.* **2013**, 4847−4859.
(64) Braddock, D. C.; Rzepa, H. S. *J. Nat. Prod.* **2008**, *71*, 728−730.
(65) Guella, G.; Mancini, I.; Öztunç, A.; Pietra, F. *Helv. Chim. Acta* **2000**, *83*, 336−348.
(66) Venkateswarlu, S.; Panchagnula, G. K.; Gottumukkala, A. L.; Subbaraju, G. V. *Tetrahedron* **2007**, *63*, 6909−6914.
(67) Rahman, A.-U.; Choudhary, M. I.; Hayat, S.; Khan, A. M.; Ahmed, A. *Chem. Pharm. Bull.* **2001**, *49*, 105−107.
(68) Hu, G.; Liu, K.; Williams, L. J. *Org. Lett.* **2008**, *10*, 5493−5496.
(69) Lamberto, M.; Kilburn, J. D. *Tetrahedron Lett.* **2008**, *49*, 6364−6367.
(70) Halton, B.; Boese, R.; Blaser, D.; Lu, Q. *Aust. J. Chem.* **1991**, *44*, 265−276.