- Thank you Jan Jensen.
- Thank you Casper Steinmann

Chapter 1

Introduction

This is the introduction

Chapter 2

Chemical shifts in a probabilistic framework

This section introduces the formalism for Monte Carlo simulations which includes both physical energy terms as well as a probabilistic energy terms based on experimentally observed chemical shifts. These equations presented are not new, but have not been published in the form in which they are presented here. The intention is to present the equations in the form in which they are implemented in PHAISTOS, so that they can easily be re-implemented in other programs by others.

2.1 Defining an energy function from Bayes' theorem

A simplistic approach to this problem is to is to define a hybrid energy by defining a penalty function that describes the agreement between experimental data and data calculated from a proposed protein structure with a physical energy (such as from a molecular mechanics force field). A structure is then determined by minimizing

$$E_{\text{hybrid}} = w_{\text{data}}\ E_{\text{data}} + E_{\text{physical}}. \tag{2.1}$$

This approach, however, does not uniquely define neither shape nor weight of E_{data}. Chemical shifts have been combined with physical energies in a multitude of ways, e.g., weighted RMSD values or harmonic constraints. Vendruscolo and co-workers implemented a "square-well soft harmonic potential", and corresponding molecular gradients and were able to run a chemical shift-biased MD simulation. In all cases the parameters and weights of E_{data} had to be carefully tweaked by hand, and it is not clear how to choose optimal parameters.

The inferential structure determination (ISD) principles introduced by Rie-ping, Habeck and Nigles [Rieping et al., 2005]defines a Bayesian formulation of Eq XX. In the following section the equations of an ISD approach are derived for combining the knowledge of experimental chemical shifts with a physical energy. First remember Bayes' theorem which relates a conditional probability (A given B) with its inverse:

$$p\left(A|B\right) = \frac{p\left(B|A\right)p\left(A\right)}{p\left(B\right)} \tag{2.2}$$

Now consider a set of chemical shifts $\{\delta_i\}$, and the uncertainty to which these can be predicted $\{\sigma_i\}$ from a structure, $\boldsymbol{X}$ (the experimental uncertainty is negligibly small compared to this). We have to make the basic assumption, that the error, given as $\Delta\delta_i = \left|\delta_i^{\text{predicted}} - \delta_i^{\text{experimental}}\right|$, approximately

follows a Gaussian distribution with some standard deviation, but we need not hand-pick and assign any numeric value to the standard deviation.Furthermore, the Gaussian distribution is the least biasing distribution according to the principle of maximum entropy.

In this case, the most likely structure, $\boldsymbol{X}$, and optimal choice of $\{\sigma_i\}$ is found by maximizing (via Bayes' theorem)

$$p\Big(\boldsymbol{X},\{\sigma_i\}\Big|\{\delta_i\}\Big) = \frac{p\Big(\{\delta_i\}\Big|\boldsymbol{X},\{\sigma_i\}\Big)p\Big(\boldsymbol{X},\{\sigma_i\}\Big)}{p\Big(\{\delta_i\}\Big)}. \tag{2.3}$$

Here, *marginal distribution* of $p\left(\{\delta_i\}\right)$ merely serves as a normalizing factor and the *likelihood* of $p\left(\{\delta_i\}|\boldsymbol{X},\{\sigma_i\}\right)$, is obtained as the product of the individual, Gaussian probabilities over all n single chemical shift measurements. Nuclei of the same atom-type, here denoted by index j, (e.g. C^{α}, H^{α}, etc.) are assumed to carry the same uncertainly denoted by σ_j:

$$\begin{aligned}
p\Big(\{\delta_i\}\Big|\boldsymbol{X},\{\sigma_i\}\Big) &\simeq \prod_{i=0}^{n} p\left(\Delta\delta_i|\boldsymbol{X},\sigma_i\right) && (2.4)\\
&= \prod_{j=0}^{m}\prod_{i_j=0}^{n_j} p\left(\Delta\delta_{i_j}|\boldsymbol{X},\sigma_j\right) && (2.5)\\
&= \prod_{j=0}^{m}\prod_{i_j=0}^{n_j} \frac{1}{\sigma_j\sqrt{2\pi}}\exp\left(-\frac{\Delta\delta_{i_j}^2}{2\sigma_j^2}\right) && (2.6)\\
&= \prod_{j=0}^{m}\left(\frac{1}{\sigma_j\sqrt{2\pi}}\right)^{n_j}\exp\left(\sum_{i_j=0}^{n_j}-\frac{\Delta\delta_{i_j}^2}{2\sigma_j^2}\right) && (2.7)\\
&= \prod_{j=0}^{m}\left(\frac{1}{\sigma_j\sqrt{2\pi}}\right)^{n_j}\exp\left(\frac{-\chi_j^2(\boldsymbol{X})}{2\sigma_j^2}\right) && (2.8)
\end{aligned}$$

Furthermore, $p\left(\boldsymbol{X},\{\sigma_j\}\right)$ can be simplified as

$$\begin{aligned}
p\Big(\boldsymbol{X},\{\sigma_j\}\Big) &\propto p\Big(\{\sigma_j\}\Big|\mathbf{X}\Big)p\Big(\mathbf{X}\Big) && (2.9)\\
&= p\Big(\{\sigma_j\}\Big)p\Big(\boldsymbol{X}\Big), && (2.10)
\end{aligned}$$

where it is assumed that the errors in the chemical shift prediction model are independent of the particular protein structure and *vice versa*. The *prior* distribution of $p\left(\{\sigma_j\}\right)$ is accounted for by proposing updates from a log-normal distribution (see next subsection). $p(\boldsymbol{X})$ of the molecular protein structure is here simply the Boltzmann distribution, i.e.

$$p(\boldsymbol{X}) = \frac{1}{Z}\exp\left(-\frac{E(\boldsymbol{X})}{k_{\mathrm{B}}T}\right) \tag{2.11}$$

where $E(\boldsymbol{X})$ is the (physical) potential energy of the protein structure, most often described by a molecular mechanics force field. k_{B} is the Boltzmann constant and T is the temperature of interest. Luckily we need not calculate the partition function, Z, because the relative energy landscape

is invariant under choice of normalization constant. Note that $p(\boldsymbol{X})$ also can be introduced via conformational sampling from a biased distribution, such as for example TorusDBN or BASILISK (mimicking the Ramachandran plot and side chain rotamer distributions, respectively).

Neglecting normalization constants, the total probability to be maximized is thus proportional to:

$$p\Big(\boldsymbol{X}, \{\sigma_i\}\Big|\{\delta_i\}\Big) \propto p\Big(\{\delta_i\}\Big|\boldsymbol{X}, \{\sigma_i\}\Big)p\Big(\boldsymbol{X}\Big)p\Big(\{\sigma_i\}\Big) \tag{2.12}$$

$$\propto \prod_{j=0}^{m}\left(\frac{1}{\sigma_j\sqrt{2\pi}}\right)^{n_j}\exp\left(-\frac{1}{2\sigma_j^2}\chi_j^2\right)\exp\left(-\frac{E(\boldsymbol{X})}{k_{\mathrm{B}}T}\right)p\Big(\{\sigma_j\}\Big) \tag{2.13}$$

When $p(\{\sigma_j\})$ is introduced via biased sampling, the associated hybrid-energy to be evaluated is (again neglecting constant terms)

$$E_{\text{hybrid}}\left(\boldsymbol{X}\right) = -k_{\mathrm{B}}T\ln\left(p\Big(\boldsymbol{X}, \{\sigma_i\}\Big|\{\delta_i\}\Big)\right) \tag{2.14}$$

$$= E(\boldsymbol{X}) - k_{\mathrm{B}}T\sum_{j=0}^{n_j} n_j \ln\left(\frac{1}{\sigma_j\sqrt{2\pi}}\right) + \frac{\chi_j^2}{2\sigma_j^2} \tag{2.15}$$

2.1.1 Sampling of nuisance parameters

Since the nuisance parameters of the energy functions are unknown, they too must be sampled.

Jeffreys' prior (general, one-parameter case)

The prior distribution of the nuisance parameter is inherently unknown. In such cases, it is necessary to use a prior distribution that will have only very little influence on the sampled value. One such *uninformative prior* could for instance be a flat distribution over the positive real line. The concept of Jeffreys' priors are a generalization of flat priors. In the one parameter case the Jeffrey's prior is given as

$$p(\theta) \propto \sqrt{\mathbf{I}(\theta)}, \tag{2.16}$$

where $\mathbf{I}(\theta)$ is the *Fisher information* defined (in the one parameter case) as

$$\mathbf{I}(\theta) = \left\langle\left(\frac{\partial}{\partial\theta}\ln p(x|\theta)\right)^2\right\rangle. \tag{2.17}$$

Jeffreys' prior (Gaussian and Cauchy distributions)

Here we derive Jefferys' prior for the uncertainty of a Gaussian distribution, i.e. a distribution on the form

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)}{2\sigma^2}\right). \tag{2.18}$$

This immediately gives us the Jeffreys' prior:

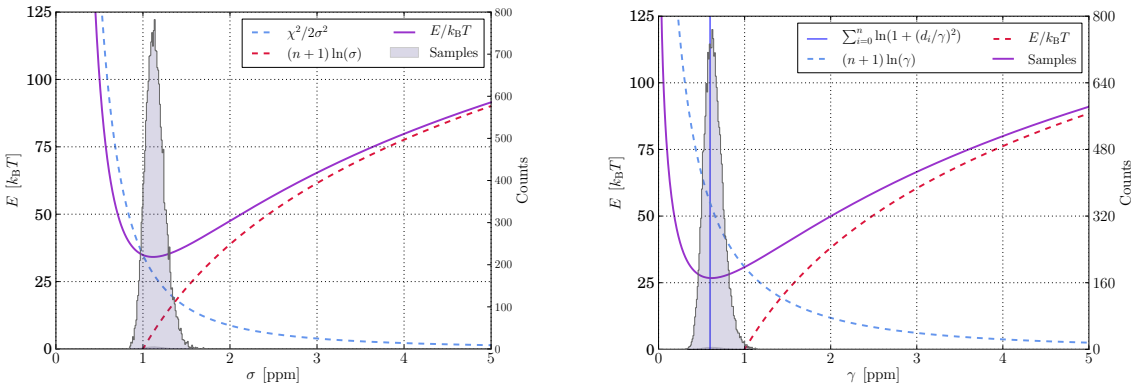
$$\begin{aligned} p(\sigma) &\propto \sqrt{\left\langle \left(\frac{\partial}{\partial \sigma} \ln p(x|\mu,\sigma) \right)^2 \right\rangle} \\ &= \sqrt{\left\langle \left(\frac{\partial}{\partial \sigma} \ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)}{2\sigma^2} \right) \right] \right)^2 \right\rangle} \\ &= \sqrt{\left\langle \left(\frac{(x-\mu)-\sigma^2}{\sigma^3} \right)^2 \right\rangle} \\ &= \sqrt{\int_{-\infty}^{\infty} p(x|\mu,\sigma) \left(\frac{(x-\mu)-\sigma^2}{\sigma^3} \right)^2 dx} \\ &= \sqrt{\frac{2}{\sigma^2}} \propto \frac{1}{\sigma} \end{aligned} \tag{2.19}$$

Similarly for the γ parameter of the Cauchy distribution of the form

$$p(x|x_0,\gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma} \right)^2 \right]}, \tag{2.20}$$

we obtain the following Jeffreys' prior:

$$\begin{aligned} p(\gamma) &\propto \sqrt{\left\langle \left(\frac{\partial}{\partial \gamma} \ln p(x|x_0,\gamma) \right)^2 \right\rangle} \\ &= \sqrt{\left\langle \left(\frac{\partial}{\partial \gamma} \ln \left[\frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma} \right)^2 \right]} \right] \right)^2 \right\rangle} \\ &= \sqrt{\left\langle \left(-\frac{\gamma^2 - (x-x_0)^2}{\gamma^3 + \gamma(x-x_0)^2} \right)^2 \right\rangle} \\ &= \sqrt{\int_{-\infty}^{\infty} p(x|x_0,\gamma) \left(-\frac{\gamma^2 - (x-x_0)^2}{\gamma^3 + \gamma(x-x_0)^2} \right)^2 dx} \\ &= \sqrt{\frac{1}{2\gamma^2}} \propto \frac{1}{\gamma} \end{aligned} \tag{2.21}$$



(a) Gaussian distribution

(b) Cauchy distribution

Figure 2.1: Sampling of σ and γ for 2OED for Ca-chemical shifts. $n = 55$ and $\chi^2 = 69.7$.

JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Definitive Benchmark Study of Ring Current Effects on Amide Proton Chemical Shifts

Anders S. Christensen,* Stephan P. A. Sauer, and Jan H. Jensen*

Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

Supporting Information

ABSTRACT: The ring current effect on chemical shifts of amide protons ($\Delta\delta_{RC}$) is computed at the B3LYP/6-311++G(d,p)//B3LYP/aug-cc-pVTZ level of theory for 932 geometries of dimers of *N*-methylacetamide and aromatic amino acid side chains extracted from 21 different proteins. These $\Delta\delta_{RC}$ values are scaled by 1.074, based on MP2/cc-pVQZ//B3LYP/aug-cc-pVTZ chemical shift calculations on four representative formamide/benzene dimers, and are judged to be accurate to within 0.1 ppm based on CCSD(T)/CBS//B3LYP/aug-cc-pVTZ calculations on formamide. The 932 scaled $\Delta\delta_{RC}$ values are used to benchmark three empirical ring current models, including the Haigh−Mallion model used in the SPARTA, SHIFTX, and SHIFTS chemical shift prediction codes. Though the RMSDs for these three models are below 0.1 ppm, deviations up to 0.29 ppm are found, but these can be decreased to below 0.1 ppm by changing a single parameter. The simple point-dipole model is found to perform just as well as the more complicated Haigh−Mallion and Johnson−Bovey models.

1. INTRODUCTION

Prediction of chemical shifts in proteins based on protein structure serves many uses in structure verification or fast generation of structures in accordance with relatively inexpensive experimental nuclear magnetic resonance (NMR) data.[1−5] The most popular protein chemical prediction software packages include the SHIFTX,[6] SHIFTS,[7,8] SPARTA,[9] and PROSHIFT[10] servers. These programs use empirical models that relate various features of protein structure, such as secondary structure, hydrogen bond geometry, and ring current effects, to changes in chemical shifts. The contributions from several different sources of chemical shift perturbations are in all cases assumed to work additively. These small additive terms are in many cases given as classical approximations to well-known physical interactions. For example, SPARTA, SHIFTX, and SHIFTS use the approximation due to Haigh and Mallion[11] to model the changes in chemical shifts due to ring current effects in the aromatic side chains of phenylalanine, tyrosine, tryptophan, and histidine residues. The Haigh−Mallion model contains a single adjustable parameter for each side chain, which must be parametrized from a data set. In the case of SHIFTX, these parameters are obtained by a data mining approach, which correlates experimental chemical shifts with corresponding, known X-ray protein structures, based on a series of predefined physical and empirical terms.[6] In the case of SHIFTS, the parameters are obtained by fitting parameters for a set of known physical terms, which relates protein structure and chemical shifts to a data set which combines empirical chemical shift data as well as data obtained through quantum chemical calculations.[12] Ultimately, these fitting methods include uncertainties from many terms in the underlying approximations of the fit, as well as the uncertainties connected to the experimental data. It is thus unclear how accurate the obtained parameters are in reproducing the underlying physics of the system.[13]

Other methods exist to approximate the ring current effect, most notably[11,14] the Johnson−Bovey model[15] and the simpler point-dipole model due to Pople.[16,17] In general, the three methods describe the change in chemical shift due to a nearby aromatic ring, formally as

$$\Delta\delta_{RC} = i\,B\,G \tag{1}$$

where G is a geometric factor, which depends on the spatial orientation and distance of the ring relative to the proton, B is a "natural constant" denoting the ring current intensity for a benzene ring, and i is the ring current intensity relative to that of a benzene ring, such that $i_{benzene} \equiv 1$. A thorough description of the three models mentioned above can be found in the Supporting Information. It has previously been attempted to derive analytical expressions for the values of i and B for different functional forms of G, but these have not been successful in reproducing experimental results.[11] Consequently, various numerical methods have been widely used to obtain the intensity values.

The study by Case[18] is one of the few that has addressed these issues using *ab initio* calculations. A methane molecule is used to probe the chemical shift perturbation due to a nearby aromatic ring, with the chemical shift calculated at the CSGT/PW91/IGLO-III level of theory. However, it is not clear whether the level of theory used at that time (1995) is sufficiently accurate, and second, it is unknown whether the ring current parameters obtained for a methane proton can directly be transferred to amide protons. The parameters obtained by Case are used in the SPARTA program.[9]

In this study, we use CCSD(T) and MP2 methods to benchmark the accuracy of $\Delta\delta_{RC}$ calculations at the B3LYP/6-311++G(d,p)//B3LYP/aug-cc-pVTZ level of theory. This level of theory is then used to compute nearly 1000 representative

Received: April 15, 2011

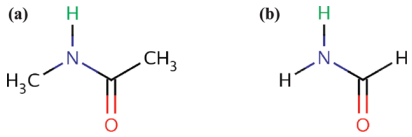


Figure 1. The molecules used as probes for the ring current effects on amide protons: *trans*-*N*-methylacetamid (NMA) (a) and formamide (FMA) (b). The probe nucleus for which the shielding constants are calculated are the amide proton of NMA and the amide proton *trans* to the C═O bond in FMA, here marked in green color.

$\Delta\delta_{RC}$ values, which, in turn, are used to obtain parameters for three empirical $\Delta\delta_{RC}$ models.

The paper is organized as follows. First, we describe the computational methodology used to isolate the ring current effect and to obtain data sets, against which the ring current intensity parameters are fitted. Then, we benchmark various levels of theory, in order to estimate error bounds on our data. This is followed by a presentation of the obtained intensities and a comparison to the intensities obtained by other authors.

2. COMPUTATIONAL METHODOLOGY

2.1. Isolating the Ring Current Effect. As a probe nucleus, for which the isotropic shielding is calculated using quantum chemical methods, the amide proton of a *trans*-*N*-methylacetylamide (NMA; Figure 1a) molecule is used in order to provide a small and inflexible molecule with a high degree of resemblance to the amide group of the protein backbone. For more computationally demanding calculations, such as MP2, CCSD, and CCSD(T) calculations, the two methyl groups are removed in order to save computational time, and the probe nucleus is then the amide proton *trans* to the C═O bond in formamide (FMA; Figure 1b).

A large number of dimer systems (see next subsection), consisting of an amide probe molecule and a nearby aromatic ring in different conformations is constructed. Following the general approach of Boyd and Skrynnikov,[19] the absolute shielding of the probe hydrogen in the dimer system can be written as

$$\sigma_H^{Dimer} = \Delta\sigma_{Conformation} + \Delta\sigma_{Local} + \Delta\sigma_{RC} \quad (2)$$

The chemical shift of the probe hydrogen atom in the dimer system will, apart from the ring current effect, also be influenced by the exact geometry and type of the probe molecule (described in the $\Delta\sigma_{Conformation}$ term) as well as any possible interactions with the aromatic moiety, such as electrostatic forces, possible hydrogen bonding, spin−spin repulsion, and other effects which can be difficult to quantify and separate (described in the $\Delta\sigma_{Local}$ term). Finally, the chemical shift perturbation due to the aromatic ring is approximated as $\Delta\sigma_{RC}$.

Reference systems for each dimer system, which have approximately identical local interactions between the molecules, apart from the ring current effect, are constructed in order to filter out these hard-to-quantify effects (see Figure 2). These are modeled as corresponding dimer systems, where the aromatic ring has been replaced by an olefinic analogue. The definition of an olefinic analogue here is an aromatic ring which, by the addition of two hydrogen atoms, has lost its aromaticity. The protonation is done such that the planar geometry of the ring is still enforced, causing deviations in the spatial positions relative to the corresponding aromatic ring dimer to be negligible. The olefinic

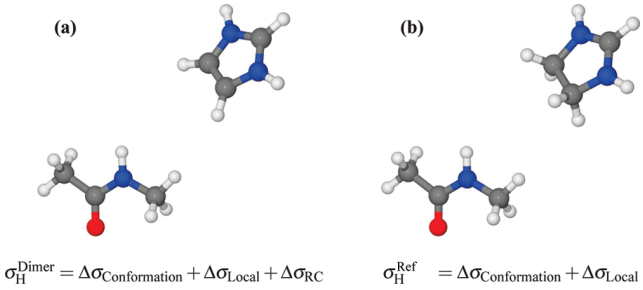


Figure 2. Two example geometries demonstrating the two different dimers used in the calculation scheme to isolate the ring current effect for one amide-ring conformation. The shown geometries correspond to the amide proton of ILE64 and the side chain of HIS69 in the HIV-1 protease, PDB-code 2I4V. In part a, the chemical shift of the probe nucleus is determined by the conformation of the NMA molecule, electrostatic and spin−spin repulsion interactions to the positively charged imidazolium molecule, and, by comparison, a small ring current interaction. In part b, the aromaticity of the imidazolium is broken, but the spatial distribution of charge as well as the internal conformation of NMA is approximately identical to those found in part a.

analogue is placed such that the ring center corresponds to the center of the aromatic ring, relative to the probe hydrogen, and the coordinates of the carbon/nitrogen atoms are matched as closely as possible. This approach ensures that $\Delta\sigma_{Conformation}$ and $\Delta\sigma_{Local}$ are largely retained, while $\Delta\sigma_{RC}$ is removed. Using this substitution scheme, the absolute shielding of the hydrogen atom in the reference system can be written as

$$\sigma_H^{Ref} = \Delta\sigma_{Conformation} + \Delta\sigma_{Local} \quad (3)$$

which enables us to estimate the ring current contribution to the chemical shift due to the aromatic ring as

$$\Delta\delta_{RC} = -\Delta\sigma_{RC} \approx \sigma_H^{Ref} - \sigma_H^{Dimer} \quad (4)$$

The aromatic rings studied here are equivalent to the rings found in the aromatic protein side chains. See Table 1 for an overview of the used molecules. Sketches are shown in the Supporting Information as well.

2.2. Construction of Test Systems. Dimers consisting of an amide probe and an aromatic ring were generated from a data set of 21 protein structures obtained from the RCSB Protein Data Bank (PDB),[20] in order to to ensure that only realistic conformations were used in the QM calculations.

The structures used were (listed by PDB code): 1F94, 1GK1, 1IGD, 1JYQ, 1JYR, 1JYU, 1Q3E, 1QJP, 1VJC, 1XA5, 1ZJK, 2ACO, 2B6C, 2D57, 2DRJ, 2ETL, 2F47, 2FZG, 2GOL, 2I4D, and 2I4V. Since the used structures were experimental X-ray structures, no hydrogen atoms were present in the structures, and PDB2PQR 1.5[21,22] was used to protonate the structures in order to obtain hydrogen atom positions. From all of these protein structures, we selected systems where the center of an aromatic ring was within a cutoff distance of 7 Å from an amide proton. This resulted in a total of 932 different dimer conformations (see Table 1). For each of these conformations, a dimer was created with a simpler aromatic ring in place of the aromatic side chain with the ring centers at identical coordinates and in the same plane. A directional vector was used to align the rotation of the ring in the plane, in order to have closely matching coordinates for the heavy atoms. For tyrosine, the center to oxygen vector was used. For benzene, the center to C^{γ} to C^{ζ} vector was used.

Table 1. List of the Side Chain Approximations Used in This Work and Their Olefinic Analogues[a]

side chain	analogue	olefinc analogue	# dimers
phenylalanine	benzene	1,4-cyclohexadiene	276
tyrosine	phenol	cyclohexa-1,4-diene-1-ol	172
tryptophan	indole	2,3,5,6-tetrahydroindole	113
histidine	imidazole	4,5-dihydroimidazole	185
histidine$^+$	imidazolium	4,5-dihydroimidazolium	174

[a] The residue type is listed along with the aromatic and olefinic analogues, as well as the total number of different NMA/ring dimer conformations in each data set.

For histidine (both in the charged and neutral state), the center to $C^{\varepsilon 1}$ vector was used. The same histidine−amide group pairs were used to generate the dimers for both charged and neutral histidine conformations. The $N^{\varepsilon 2}$ nitrogen atom was in all cases of neutral histidine assumed to be the deprotonated nitrogen.

A given backbone amide group within the 7 Å range from the aromatic ring was substituted by an NMA or an FMA molecule, with the nitrogen atoms at identical coordinates. Furthermore, the N−H vector and the C(=O)−N−H plane were also aligned. See Table 1 for the number of dimers for each ring type.

If the aromatic ring corresponded to the ring of the side chain of the previous residue, with respect to the investigated amide group, the dimer construction scheme occasionally caused a clash between the extra hydrogens, where the C^{β} atom was previously located. Other conformations also gave rise to unphysical conformations, due to clashes between the inflexible subunits of the constructed dimers. To avoid computational artifacts from these, all dimers with a shortest intermolecular distance of 3.4 Å or less were discarded, since 3.4 Å is twice the van der Waal radius of the largest atom (carbon) in the system.[23] NMR shielding constants were then calculated for the dimer systems.

2.3. Basis Set Extrapolations and Correlation Effects. In this work, density functional theory (DFT; and the very popular B3LYP functional[24,25]) is used to obtain NMR shielding constants. Due to the partly semiempirical nature of the approximated exchange-correlation functionals used, B3LYP data cannot in general be expected to show convergence toward experimental shielding values or values obtained at very accurate levels of theory when increasing the basis set size.[26] It is, however, often the case that a small error can be obtained in calculated DFT chemical shielding constants, compared to high-level correlated wave function methods, if a simple linear correction or scaling factor is applied to the DFT data.[27] In this work, a comparison of B3LYP to high-level correlated methods is used to obtain such a linear scaling factor.

Unfortunately, CCSD(T) calculations with an appropriate basis set are still not possible for the FMA/benzene dimer, which has 70 electrons. Instead, we benchmark the shielding constants for FMA alone at the B3LYP, MP2, CCSD, and CCSD(T) levels of theory, in order to allow us to estimate error bounds to shielding constants obtained at levels of theory less accurate than CCSD(T).

For complete basis set limit (CBS) estimates, we use the approach of Moon and Case[26] and Kupka et al.[27,28] By using Dunning's correlation consistent basis sets[29] (cc-pVxZ; where $x \in \{D, T, Q, 5, 6, ...\}$ is the valence orbital splitting in the basis set), it is possible to carry out calculations using a sequence of basis sets of well-defined, systematic increasing quality. Kupka et al. extrapolate calculated NMR shielding values toward infinite basis set size with a three parameter exponential decreasing function:

$$\sigma(x) = \sigma(\infty) + A \exp(-x/B) \quad (5)$$

where $\sigma(x)$ is the shielding obtained using a basis set with the valence orbital splitting number of x and $\sigma(\infty)$, A and B are the fitting parameters, with $\sigma(\infty)$ being the estimated shielding in the complete basis limit. A nonlinear least-squares Marquardt−Levenberg algorithm[30,31] is used to fit the parameters.

Jensen has constructed a set of basis sets for the purpose of Hartree−Fock (HF) and DFT NMR shielding calculations, called the polarization consistent pcS-n basis sets.[32] For basis sets of similar valence orbital splitting, the pcS-n basis sets contain more basis functions of low angular momentum, compared to the Dunning-type basis sets. pcS-1 is a double-ζ quality basis set, pcS-2 is triple-ζ, and so forth. When estimating the complete basis limit based on the pcS-n basis sets, a value of $x = n + 1$ is thus used in eq 5.

Last, we compare the proton chemical shift of the amide proton trans to the C=O bond in FMA at the CCSD(T)/CBS level of theory to the experimental value, in order to verify that CCSD(T)/CBS is, in fact, a reliable method. Inferring the experimental gas-phase ^{1}H shielding values from CH_4 (σ_H = 30.61 ppm[33]), an experimental value of σ_H = 26.24 ppm in the gas phase[34] at 483 K is obtained. At this temperature, thermal motion cause rapid switching of the two N-amide protons and the peaks are not separable, so this value has to be considered as an average over the two proton chemical shifts.[34]

It is well-known[35−38] that a zero-point vibrational correction (ZPVC) has to be added to equilibrium geometry *ab initio* shielding constants in order to obtain close agreement to experimental data. This vibrational averaging correction can easily be calculated using the method of Kern and Matcha.[39] While we had preferred to carry out a ZPVC calculation at the same level of theory as the geometry optimization of the molecules used throughout this work, no program is currently capable of automatically computing a ZPVC at the DFT level of theory with Gaussian-type basis sets. In the following, the ZPVC is calculated at the MP2/cc-pVQZ level of theory instead.

2.4. Software. All geometries were minimized at the B3LYP/aug-cc-pVTZ level of theory using Gaussian 03,[40] except when otherwise noted. All DFT calculations of NMR shielding constants were carried out using Gaussian 03. MP2/6-311++G(d,p) NMR calculations were also carried out in Gaussian 03, while the calculation of MP2 shielding constants using Dunning's correlation consistent basis sets[29] and the polarization consistent pcS-n and aug-pcS-n basis sets of Jensen[32] were carried out with Turbomole 6.2.[41] All calculations at the CCSD and CCSD(T) levels of theory were carried out using CFOUR 1.0.[42] All NMR shielding constants are calculated using the Gauge-Including Atomic Orbital formulation.[43−45] For the calculations of the ZVPC to the FMA chemical shifts, the equilibrium geometry of a planar FMA molecule was obtained at the MP2/cc-pVQZ level of theory with CFOUR 1.0, exploiting the C_S symmetry of the molecule. From this equilibrium geometry, the ZVPC to the NMR isotropic shielding was subsequently calculated at the MP2/cc-pVQZ level of theory using the method of Kern and Matcha[39] as implemented in CFOUR 1.0.

Table 2. Absolute Isotropic Chemical Shielding of the *cis*-*N*-amide Proton in Gas-Phase FMA at the B3LYP, MP2, CCSD, and CCSD(T) Levels of Theory Using Dunning's Correlation Consistent Basis Sets and the Polarization Consistent Shielding Basis Sets of Jensen[a]

		method			
basis set	size	CCSD(T)	CCSD	MP2	B3LYP
cc-pVDZ	57	28.06	28.09	27.90	27.67
cc-pVTZ	132	27.29	27.35	27.16	27.17
cc-pVQZ	255	26.92	27.00	26.80	26.94
cc-pV5Z	438	26.78	26.86	26.65	26.83
$\sigma_{\text{cc-pVxZ}}(\infty)$		26.64	26.73	26.50	26.73
pcS-0	44	29.32	29.36	29.31	28.88
pcS-1	66	27.55	27.58	27.40	27.29
pcS-2	141	27.02	27.09	26.89	26.91
pcS-3	321	26.75	26.83	26.62	26.77
$\sigma_{\text{pcS-}n}(\infty)$		26.67	26.78	26.57	26.75
$\sigma_{\text{exptl(gas)}}$			26.24		

[a] All values are given as ppm. The experimental value is obtained at 483 K.[34] $\sigma(\infty)$ is obtained using eq 5 and fitted over all values in the corresponding series of basis sets. The size indicates the number of basis functions in the system at the given basis set size. All shielding constants are given in ppm.

3. RESULTS

3.1. Correlation and Basis Set Effects on the Chemical Shift of the (N−)H Proton in Formamide. It is currently not feasible to perform a complete basis set study at the CCSD(T) level for $\Delta\delta_{RC}$ of a FMA/benzene dimer. Instead, we perform such a study of the chemical shielding of the *cis*-*N*-proton in FMA and use the results to benchmark more approximate methods that can be applied to FMA/benzene dimers (as described in the next subsection).

Table 2 lists CCSD(T) chemical shielding values computed using a B3LYP/aug-cc-pVTZ optimized geometry of FMA and two different, systematic series of basis sets (cc-pV*x*Z and pcS-*n*). Each set of calculations is used to extrapolate shielding constants to the complete basis set limit (as described in the previous section) and lead to very similar results: 26.64 and 26.67 ppm for cc-pV*x*Z and pcS-*n*, respectively. In the following, we will refer to 26.64 ppm as CCSD(T)/CBS, since this value is extrapolated using the largest basis set (cc-pV5Z) and since the Dunning-type basis sets are constructed for the purpose of correlated wave function calculations, whereas the pcS-*n* basis sets are constructed specifically for shielding constant calculations at the HF and DFT levels of theory.

The CCSD(T)/CBS value is 0.40 ppm higher than the experimental gas phase value obtained at 483 K of 26.24 ppm. However, this experimental value includes vibrational effects and is an average of the chemical shieldings of both amide protons. The effect of vibrations at 0 K (i.e., the zero-point vibrational correction) can be estimated relatively easily, as described in the previous section. At the MP2/cc-pVQZ level of theory, the ZPVC is −0.26 ppm, which, when used to correct the CCSD(T)/CBS value, results in a chemical shielding of 26.38 ppm—within 0.14 ppm of experiment. The ZPVC correction is unlikely to contribute significantly to $\Delta\delta_{RC}$, because it is a shielding difference between two molecular systems with very similar vibrational normal modes of FMA. Thus, we in the following focus on the electronic contribution to the chemical shielding alone.

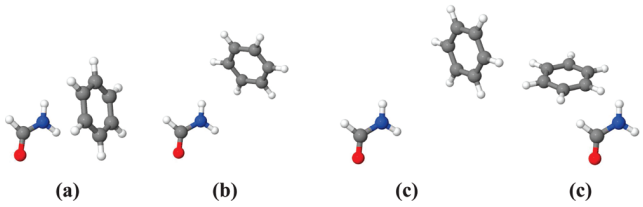


Figure 3. Pictures of the four FMA/benzene dimers used in this study. The resulting $\Delta\delta_{RC}$ calculated at various levels of theory for each conformation can be found in Table 3, where $\Delta\delta^1_{RC}$ corresponds to conformation a, $\Delta\delta^2_{RC}$ to conformation b, and so forth.

The CBS values computed using CCSD, MP2, and B3LYP are all within 0.14 ppm of the CCSD(T)/CBS value, suggesting that the amide proton chemical shielding is relatively insensitive to electronic correlation effects in the limit of large basis sets. However, it is quite basis-set-dependent as at least the cc-pV5Z or the pcS-3 basis set is needed to get within 0.2 ppm of the CCSD(T)/CBS, with the exception of MP2/cc-pVQZ, which deviates by 0.16 ppm. We therefore choose MP2/cc-pVQZ for the $\Delta\delta_{RC}$ calculations using the FMA/benzene dimers described in the next subsection. Since $\Delta\delta_{RC}$ is a relative shielding value between two very similar systems, we expect that the MP2/cc-pVQZ results are well within 0.1 ppm of what would be computed with CCSD(T)/CBS and measured experimentally. A factor not investigated here was the dependence on the used geometry, which is known to cause deviations in calculated 1H shielding constants on the order of ±0.1 ppm—see for instance Rablen et al.[46]

3.2. Scaling B3LYP Results to Those Obtained with Correlated Wave Function Methods. In this section, high-level correlated wave function methods are used to obtain a linear scaling correction to the chemical shift contribution due to ring current effects, obtained at the B3LYP/6-311++G(d,p)//B3LYP/aug-cc-pVTZ level of theory.

Four dimer systems were selected from the large data set of NMA/benzene dimers (see Figure 3), in such a way that the ring current contributions ($\Delta\delta_{RC}$) in the four dimer conformations cover a range from −0.72 ppm to +0.15 ppm, at the B3LYP/6-311++G(d,p) level of theory, in even sized steps. In these dimers, the NMA molecule was replaced with the much smaller FMA molecule, and the isotropic shielding was calculated using various methods and basis sets. Here, the chemical shift is modeled by

$$\Delta\delta_{RC}^{(\text{uncorrected})} \approx \sigma_H^{\text{Probe}} - \sigma_H^{\text{Dimer}} \qquad (6)$$

where σ_H^{Probe} is the shielding of the probe nucleus in the probe molecule alone and σ_H^{Dimer} is the shielding of the probe nucleus in the probe molecule in the dimer. Note that an NMR calculation for a reference dimer is not carried out, and the linear scaling factor is unaffected, whether a reference calculation is carried out, since this calculation would also have to be scaled by the same factor. The results are collected in Table 3. We observe the following:

1. Regardless of the basis set or method, the obtained $\Delta\delta_{RC}$'s have a linear correlation to B3LYP/6-311++G(d,p) data of 0.992 or better (see the Supporting Information). It is thus demonstrated that applying a linear correction based

dx.doi.org/10.1021/ct2002607 |J. Chem. Theory Comput. XXXX, XXX, 000–000

Table 3. Shielding Constant of the FMA Probe Proton in a Vacuum for Each Method and Basis Set Used in This Section, As Well As the Chemical Shift Ring Current Interaction ($\Delta\delta^{n}_{RC}$) for Each of the Four Different Conformations Used[a]

method	σ_{FMA}	$\Delta\delta^{1}_{RC}$	$\Delta\delta^{2}_{RC}$	$\Delta\delta^{3}_{RC}$	$\Delta\delta^{4}_{RC}$	scaling
B3LYP/6-311++G(d,p)	27.64	−0.72	−0.43	−0.21	0.15	
MP2/6-311++G(d,p)	27.70	−0.76	−0.45	−0.22	0.15	1.052
CCSD/6-311++G(d,p)	27.90	−0.75	−0.44	−0.22	0.15	1.033
CCSD(T)/6-311++G(d,p)	27.85	−0.73	−0.43	−0.21	0.15	1.012
MP2/cc-pVDZ	27.90	−0.74	−0.43	−0.22	0.17	1.033
MP2/cc-pVTZ	27.16	−0.77	−0.46	−0.23	0.17	1.076
MP2/cc-pVQZ	26.80	−0.81	−0.45	−0.22	0.13	1.074
B3LYP/pcS-0	28.88	−0.75	−0.42	−0.21	0.16	1.042
B3LYP/pcS-1	27.29	−0.76	−0.44	−0.22	0.16	1.056
B3LYP/pcS-2	26.91	−0.80	−0.47	−0.24	0.14	1.087
B3LYP/pcS-3	26.77	−0.80	−0.47	−0.24	0.16	1.097
B3LYP/pcS-4	26.76	−0.80	−0.47	−0.25	0.16	1.095
CCSD(T)/CBS	26.64					
B3LYP/6-311++G(d,p) (NMA)		−0.74	−0.43	−0.22	0.14	1.004

[a] Furthermore, the resulting scaling factor relative to data obtained at the B3LYP/6-311++G(d,p) level of theory is noted. The B3LYP/6-311++G(d,p) shieldings for the identical conformation with NMA as a probe are given in the bottom row. All shielding constants and $\Delta\delta_{RC}$ values are given in ppm.

on correlated methods is a very good approximation. No constant offset (intercept) greater that 0.01 ppm was found, so the relationship between data obtained at the B3LYP level of theory and data obtained using a correlated wave function method is effectively a simple scaling factor. Thus, the fit was carried out as

$$\Delta\delta^{Other}_{RC} = k_{scaling}\Delta\delta^{B3LYP}_{RC} \quad (7)$$

where $\Delta\delta^{B3LYP}_{RC}$ is the $\Delta\delta_{RC}$ obtained at the B3LYP/6-311++G(d,p) level of theory, $\Delta\delta^{Other}_{RC}$ is the $\Delta\delta_{RC}$ obtained using another method and/or basis set, and $k_{scaling}$ is the fitted scaling constant. Here, the $\Delta\delta_{RC}$ values are obtained via eq 6.

2. All $\Delta\delta_{RC}$ values are within 0.1 ppm of one another, including B3LYP/6-311++G(d,p), which is used for the 932 dimer calculations. This supports our previous assertion that it is easier to compute $\Delta\delta_{RC}$ accurately compared to the computation of absolute shielding constants. Therefore, the $\Delta\delta_{RC}$ values listed in Table 3 are very likely within 0.1 ppm of what would be computed with CCST(T)/CBS and measured experimentally.
3. As a result, all scaling factors are within 10% and will all yield very similar results. However, on the basis of the results in Table 2, we pick the scaling factor computed at the MP2/cc-pVQZ level, where $k_{scaling}$ = 1.074.
4. The difference between ring current effects acting on either an FMA or an NMA probe was found at the B3LYP/6-311++G(d,p) level of theory to be a factor of 1.004 (see Table 3). This suggests that results obtained using FMA as a probe are, to a very good approximation, transferable to systems where NMA is used as a probe.

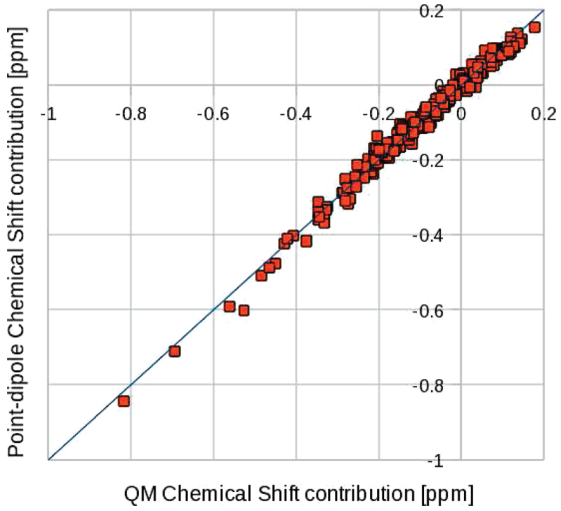


Figure 4. Correlation between the chemical shift predictions of the point-dipole model and the chemical shifts obtained by eq 4 for a set of NMA/benzene dimers using a best fit value of B_{PD} = 30.42 ± 0.16 ppm Å^3. The blue line represents the best fit between the two methods. The linear correlation of the data set is 0.993.

3.3. Expressions for the *B* Factors. In the point-dipole model, the definition of $i_{Benzene} \equiv 1$ is used, and the *B* factor in the point-dipole model (B_{PD}) is obtained via a fit using the chemical shifts obtained for all NMA/benzene dimers to their corresponding *G* values in the point-dipole model ($G_{PD}(\vec{r},\theta)$, see Supporting Information), using the following formula:

$$\Delta\delta_{RC} = B_{PD}G_{PD}(\vec{r},\theta) \quad (8)$$

where $\Delta\delta^{QM}$ is the scaled QM calculated chemical shifts of the amide protons using eq 4 and $G_{PD}(\vec{r},\theta)$ is the geometric term of the NMA/benzene dimers in the point-dipole model. This gives a value of B_{PD} = 30.42 ± 0.16 ppm Å^3. The linear correlation of this fit is r = 0.993. See Figure 4 for a scatter plot of the fitted data set.

In the literature, the trend has been to use formally derived *B* factors in the Johnson−Bovey model (B_{JB}) and scale the relative intensities accordingly.[18] Following this, the analytical values of B_{JB} are used in the Johnson−Bovey model in this work. These evaluate to −3.79 ppm and −3.25 ppm for five- and six-membered rings, respectively. To facilitate an easy comparison of the ring current intensities to those found by Case,[18] a *B* factor in the Haigh−Mallion model of B_{HM} = 5.455 ppm Å is adopted.

3.4. Fitting the Relative *i* Factors. Using the *B* factor obtained in the previous subsection, the relative ring current intensities of all ring types in the three ring current models are obtained as the best fit of *i* when fitting the right-hand side of eq 4 to the right-hand side of eq 1.

The relative intensities of the two rings in tryptophan were trivially fitted using a two parameter fitting routine, although the contributions from the five- and six-membered rings were somewhat correlated. The fitted relative ring current intensities can be found in Table 4, which also features a comparison to the *i* factors found in other studies. A comparison is made to the values used in the SHIFTX and SHIFTS programs and to the values obtained for methane hydrogen by Case[18] as used in SPARTA. The linear correlation between the B3LYP/6-311++G(d,p) ring current contributions and the predictions of the three approximations

dx.doi.org/10.1021/ct2002607 |J. Chem. Theory Comput. XXXX, XXX, 000–000

Table 4. Relative Ring Current Intensity Factors of the Different Side Chains, As Found in This Study, Compared to the Value of Other Studies[a]

model	point-dipole	Haigh−Mallion				Johnson−Bovey	
reference	This Work	SHIFTX[6]	SHIFTS[47]	Case[18] SPARTA[9]	This Work	Case[18]	This Work
PHE	1.00 (0.02, 0.07)	1.05 (0.05, 0.18)	1.00 (0.05, 0.17)	1.46 (0.07, 0.17)	1.18 (0.03, 0.06)	1.27 (0.03, 0.14)	1.13 (0.02, 0.06)
TYR	0.81 (0.02, 0.10)	0.92 (0.02, 0.09)	0.84 (0.02, 0.08)	1.24 (0.06, 0.22)	0.93 (0.02, 0.09)	1.10 (0.04, 0.10)	0.91 (0.02, 0.07)
HIS+	0.69 (0.02, 0.05)	0.43 (0.08, 0.29)	0.90 (0.06, 0.12)	1.35 (0.05, 0.07)	1.26 (0.03, 0.05)	1.40 (0.03, 0.08)	1.27 (0.03, 0.05)
HIS	0.68 (0.03, 0.06)	0.43 (0.08, 0.28)	0.90 (0.07, 0.11)	1.35 (0.06, 0.08)	1.22 (0.03, 0.07)	1.40 (0.04, 0.09)	1.25 (0.03, 0.06)
TRP5	0.57 (0.03, 0.08)	0.90 (0.03, 0.11)	1.04 (0.03, 0.08)	1.32 (0.04, 0.15)	0.97 (0.02, 0.10)	1.02 (0.02, 0.10)	1.06 (0.02, 0.09)
TRP6	1.02	1.04	1.02	1.24	1.18	1.27	1.18
B-factor	30.42 ppm Å^3	5.13 ppm Å	5.455 ppm Å	5.455 ppm Å	5.455 ppm Å	−3.25 ppm[b]	−3.25 ppm[b]
						−3.79 ppm[b]	−3.79 ppm[b]

[a] The RMSD associated with using the given intensity factor and *B* value is given as the first entry in the parentheses, and MAD as the second entry, for each intensity factor. The RMSD and MAD are calculated over all dimer systems used in the fits to obtain intensity factors. The RMSD given for tryptophan is the RMSD for a sum of both rings with the given intensities. [b] In the Johnson−Bovey model, values of −3.25 ppm and −3.79 ppm are used for six- and five-membered rings, respectively.

were found to be r = 0.980 or better (see Supporting Information), so using linear fits to determine the i factors is evidently a very good approximation. For the data sets for each ring type, chemical shift predictions of different sets of i factors are compared to our QM data. We present the *root-mean-square deviation* (RMSD) and the *maximum absolute deviation* (MAD) of the data set. The RMSDs to our QM values are seemingly very small for all sets of intensity parameters. However, this is mostly due to the magnitude of the ring current effect in the dimers used in the data set being on average very small, and RMSD is thus not a very good measure in this case. Our intensities, however, do have the lowest maximum RMSD of up to just 0.03 ppm, while the competing methods have RMSDs of up to 0.08 ppm (SHIFTX), 0.07 ppm (SHIFTS), 0.07 ppm (Case, Haigh−Mallion), and 0.04 ppm (Case, Johnson−Bovey) for a residue type. A much better metric than RMSD is in this case the maximum average deviation (MAD) to QM values, which loosely corresponds to the largest error one can expect from using a certain set of intensities. In this metric, our method has a MAD of 0.05−0.10 ppm or better, while the corresponding numbers for the competing methods are 0.09−0.29 ppm (SHIFTX), 0.08−0.17 ppm (SHIFTS), 0.09−0.29 ppm (Case, Haigh−Mallion), and 0.09−0.14 ppm (Case, Johnson−Bovey), for all residue types.

We note that the SHIFTX and SHIFTS predictions are, on average, slightly lower than our prediction and those of Case. One possible explanation is that the SHIFTX and SHIFTS predictions are based on empirical parameters fitted to chemical shifts measured for solution phase structures. These structures may exhibit larger conformational fluctuations, leading to a net larger average distance between the ring and the amide proton compared to the X-ray structure as well as fluctuations in the direction of the magnetic dipole arising from the aromatic side chains and, therefore, a smaller ring current effect.

4. SUMMARY

We have presented sets of ring current intensity parameters for chemical shift predictions with the point-dipole, Haigh−Mallion, and Johnson−Bovey models. The maximum errors arising from use of the presented parameters are judged to be within ±0.1 ppm from what would have been computed at the CCSD(T)/CBS for a set of 932 test cases. Further improvements in computational methodology are thus not expected to yield any significant qualitative or quantitative improvement in chemical shift prediction in proteins. Preliminary calculations at the B3LYP/6-311++G(d,p) level using methane as a probe and intermolecular geometries corresponding to those in Table 3 suggest that the current parameters can be used to predict ring current effects on CH protons to within 0.2 ppm.

The presented parameters are rigorously based on the underlying physical properties of aromatic molecules. Parameters based on empirical models were found to perform worse on our amide proton test set. Our report of the superiority of a physics-based method over empirical methods is backed up by the fact that the parameters obtained by Case[18] through QM methods have the same disagreements with the empirical methods as our parameters, despite the fact that the computational methodology used by Case was somewhat different than ours.

Finally, we have made a detailed numerical comparison between the point-dipole, Haigh−Mallion, and Johnson−Bovey models. The chemical shift predictions of the three models were nearly identical, and no outliers compared to our quantum mechanical calculations were found in any of the three models. Apart from reported problems with predictions of ring current effects in macrocyclic rings, such as those found in porphyrins,[48] which should be a nonissue for most uses in protein chemical shift predictions, the three methods should yield results of identical accuracy. Hence, we suggest that the point-dipole model should be used in future chemical shift prediction software, since (1) it both is computationally faster than competing models, since it does not require any integral evaluation as opposed to the Johnson−Bovey model, and contains significantly fewer geometric terms than the Haigh−Mallion model and (2) it is much easier to implement than the competing models.

■ ASSOCIATED CONTENT

Supporting Information. Thorough descriptions of the point-dipole, Haigh−Mallion, and Johnson−Bovey models and our implementations are described. The material also includes sketches of the used molecules, linear correlation values and scatter plots of the fits used to obtain the intensity values of Table 4, and an additional investigation of the equilibrium geometry dependence of FMA NMR shielding calculations.

This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Authors
*E-mail: andersx@nano.ku.dk; jhjensen@chem.ku.dk.

ACKNOWLEDGMENT

A.S.C. is recipient of a Ph. D. scholarship funded by the Novo Nordisk STAR program. S.P.A.S. thanks the Danish Center for Scientific Computing (DCSC), the Danish Natural Science Research Council/The Danish Council for Independent Research, and the Carlsberg Foundation for support.

REFERENCES

(1) Raman, S.; Lange, O.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* **2010**, *327*, 1014–1018.
(2) Meiler, J.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15404–15409.
(3) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270–1272.
(4) Vila, J. A.; Scheraga, H. A. *Acc. Chem. Res.* **2009**, *42*, 1545–1553.
(5) Robustelli, P.; Kohlhoff, K. J.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923–933.
(6) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Am. Chem. Soc.* **1991**, *111*, 9436–9444.
(7) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321–333.
(8) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408–423.
(9) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289–302.
(10) Meiler, J. *J. Biomol. NMR* **2003**, *26*, 25–37.
(11) Haigh, C. W.; Mallion, R. B. *Prog. NMR Spectrosc.* **1980**, *13*, 303–344.
(12) Moon, S.; Case, D. A. *J. Biomol. NMR* **2007**, *38*, 139–150.
(13) Mulder, F. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *395*, 578–590.
(14) Monya, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702–709.
(15) Johnson, C. E.; Bovey, F. A. *J. Chem. Phys.* **1958**, *29*, 1012–1014.
(16) Pople, J. A. *J. Chem. Phys.* **1956**, *24*, 1111.
(17) Pople, J. A. *Mol. Phys.* **1958**, *1*, 175–180.
(18) Case, D. A. *J. Biomol. NMR* **1995**, *6*, 341–346.
(19) Boyd, J.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2002**, *124*, 1832–1833.
(20) Berman, J.; Westbrook, H. M.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *106*, 16972–16977.
(21) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, 522–525.
(22) Dolinsky, T. J.; Nielsen, J.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, 665–667.
(23) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.
(24) Becke, A. D. *J. Phys. Chem.* **1993**, *98*, 5648–5652.
(25) Stephens, P.; Devlin, F.; Chabalowski, C.; Frisch, M. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
(26) Moon, S.; Case, D. A. *Nucleic Acids Res.* **2004**, *32*, 665–667.
(27) Kupka, T.; Ruscic, B.; Botto, R. E. *J. Phys. Chem. A* **2002**, *106*, 10396–10407.
(28) Kupka, T.; Ruscic, B.; Botto, R. E. *Solid State Nucl. Magn. Reson.* **2003**, *23*, 145–167.
(29) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
(30) Marquardt, D. W. *J. Appl. Math.* **1963**, *11*, 431–441.
(31) Levenberg, K. *Q. Appl. Math.* **1944**, *2*, 164–168.
(32) Jensen, F. *J. Chem. Theory Comput.* **2008**, *5*, 719–727.
(33) Jameson, K. A.; Jameson, C. J. *Chem. Phys. Lett.* **1987**, *134*, 461–466.
(34) Vaara, J.; Kaski, J.; Joksaari, J.; Diehl, P. *J. Phys. Chem. A* **1997**, *101*, 5069–5081.
(35) Ruud, K.; Astrand, P.-O.; Taylor, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 4826–4833.
(36) Sauer, S. P. A.; Spirko, V.; Paidarová, I.; Kraemer, W. P. *Chem. Phys.* **1997**, *214*, 91–102.
(37) Wigglesworth, R. D.; Raynes, W. T.; Sauer, S. P. A.; Oddershede, J. *Mol. Phys.* **1999**, *96*, 1595–1607.
(38) Wigglesworth, R. D.; Raynes, W. T.; Kirpekar, S.; Oddershede, J.; Sauer, S. P. A. *J. Chem. Phys.* **2000**, *112*, 736–746.
(39) Kern, C. W.; Matcha, R. L. *J. Phys. Chem.* **1968**, *49*, 2081–2092.
(40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, A. J., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
(41) Ahlrichs, R.; Baer, M.; Haeser, M.; Horn, H.; Koelmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(42) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P.; Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Juselius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Stopkowicz, S.; Tajti, A.; Vazquez, J.; Wang, F.; Watts, J. D.; Almlöf, J.; Taylor, P. R.; Taylor, P. R.; Helgaker, T.; Jensen, H. J. A.; Jorgensen, P.; Olsen, J.; Mitin, A. V.; van Wüllen, C. CFOUR, a quantum chemical program package. For the current version, see http://www.cfour.de (accessed June 2010).
(43) London, F. *J. Phys. Radium* **1937**, *8*, 397–409.
(44) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789–807.
(45) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497–5509.
(46) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357–7363.
(47) Ösapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *111*, 9436–9444.
(48) Perkins, S. J. *Applications of ring current calculations to proton NMR of proteins and transfer RNA*; Plenum Press: New York, 1982; Vol. 4, Chapter 4, pp 193−336.

Bibliography

[Rieping et al., 2005] Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. *Science*, 308:303–306.