

Toward the Quantum Chemical Calculation of NMR Chemical Shifts of Proteins. 2. Level of Theory, Basis Set, and Solvents Model Dependence

Andrea Frank,[†] Heiko M. Möller,[†] and Thomas E. Exner*,^{†,‡}

[†]Department of Chemistry and Zukunftskolleg, University of Konstanz, D-78457 Konstanz, Germany

[‡]Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Eberhard Karls University Tübingen, D-72076 Tübingen, Germany

S Supporting Information

ABSTRACT: It has been demonstrated that the fragmentation scheme of our adjustable density matrix assembler (ADMA) approach for the quantum chemical calculations of very large systems is well-suited to calculate NMR chemical shifts of proteins [Frank et al. *Proteins* 2011, 79, 2189–2202]. The systematic investigation performed here on the influences of the level of theory, basis set size, inclusion or exclusion of an implicit solvent model, and the use of partial charges to describe additional parts of the macromolecule on the accuracy of NMR chemical shifts demonstrates that using a valence triple- ζ basis set leads to large improvement compared to the results given in the previous publication. Additionally, moving from the B3LYP to the mPW1PW91 density functional and including partial charges and implicit solvents gave the best results with mean absolute errors of 0.44 ppm for hydrogen atoms excluding H^N atoms and between 1.53 and 3.44 ppm for carbon atoms depending on the size and also on the accuracy of the protein structure. Polar hydrogen and nitrogen atoms are more difficult to predict. For the first, explicit hydrogen bonds to the solvents need to be included and, for the latter, going beyond DFT to post-Hartree–Fock methods like MP2 is probably required. Even if empirical methods like SHIFTX+ show similar performance, our calculations give for the first time very reliable chemical shifts that can also be used for complexes of proteins with small-molecule ligands or DNA/RNA. Therefore, taking advantage of its ab initio nature, our approach opens new fields of application that would otherwise be largely inaccessible due to insufficient availability of data for empirical parametrization.

INTRODUCTION

NMR chemical shifts are the central property in NMR spectroscopy. Therefore, an accurate and reliable prediction even for very large molecular systems is highly desired for a number of applications. For proteins, which are the main focus of the study presented here, such applications include, besides others, structure determination, investigation on the dynamic behavior, and binding studies with other proteins, DNA/RNA, and/or small-molecule ligands. Beside many empirical methods,^{1–7} quantum chemical methods start to reappear in the recent literature.^{8–29} All these are based on the idea that the large system can be divided into small fragments, for which the NMR chemical shifts are determined separately. Especially, automated fragmentation quantum mechanics/molecular mechanics (AF-QM/MM)²⁰ has been shown to be able to give excellent agreement with the full calculations and also a good correlation with experimental values.^{30–33} In recent publications, AF-QM/MM was also used for the calculation of chemical shift anisotropies in peptides and proteins by Tang and Case.^{23,24} Oldfield and his co-workers extensively used small model systems to determine chemical shifts in peptides and proteins (for a review of this work, see ref 8). This work was then continued and extended by de Dios and co-workers.^{10,26} The group of Scheraga^{13–16,34} developed a method for protein NMR structure determination, refinement, and validation based on quantum chemical ¹³C^α chemical shifts calculated for a large number of different conformations of the central residues in tripeptides.

Jacob and Visscher¹⁷ calculated NMR chemical shifts with the frozen-density embedding (FDE) scheme originally introduced by Wesolowski and Warshel.³⁵ Finally, Johnson and DiLabio applied their mixed quantum mechanical/molecular mechanical (QM/MM) algorithm to calculate chemical shifts of Gly39 of the fungal dockerin domain.¹⁹ Our method is based on the adjustable density matrix assembler (ADMA)³⁶ belonging to the fragment-based quantum mechanical approaches. Since this paper describes the specific application in the field of NMR spectroscopy, we will not review the newest developments in this field here. But we hope that these will be utilized in NMR chemical shift calculation in the very near future and point the interested reader to ref 37 and especially refs 38–41. Besides the just mentioned advances, NMR chemical shift calculations will definitely also profit from the very active developments in linear scaling quantum mechanical approaches (for the latest developments see also ref 37 and references therein). With the help of approximations like local correlation methods,^{42–45} tighter multipole-based integral estimates,^{46,47} and/or density fitting approaches,⁴² calculations for up to a few hundreds of atoms are now possible. Nevertheless, there is still a gap between these calculations and the system sizes of thousands to tens of thousands of atoms seen in proteins anticipated for the calculations presented here. Even with the

Received: December 20, 2011

Published: March 14, 2012



fragment-based approaches, 3D protein structure optimization guided by quantum chemical NMR chemical shift predictions is nowadays far away from being applicable on a standard basis due to the computational demand and the large amount of generated structures during optimization. But other applications require only one known three-dimensional structure of the protein as a starting point, e.g., from X-ray crystallography, and could, therefore, profit even from such sophisticated calculations. In comparison to empirical methods, which are privileged by their much shorter runtime, ab initio methods have the advantage of being independent of experimental data sufficient for parametrization and have, thus, inherently a much larger flexibility regarding the molecular system treated.

Recently, we showed that quantum chemical calculations based on the fragmentation scheme of our adjustable density matrix assembler (ADMA)^{30–33,36} are in principle able to obtain NMR chemical shifts for large systems like proteins.²¹ For carbon atoms a reasonable correlation between theoretical and experimental chemical shifts was obtained. However, carbonyl carbons, some hydrogens, and particularly nitrogens were not predicted with adequate accuracy and it was concluded that additional improvements are needed for the applications mentioned above. Many influences determining the accuracy of the calculations, as demonstrated by small-molecule calculations, have not been studied in detail or have been neglected so far in calculation of such large molecular systems with thousands of atoms: level of theory,^{18,48–53} basis set size,^{10,48–53} conformational averaging including zero-point vibrations,^{23,26,50–52,54–57} and solvents effects.^{52,53,57,58} Therefore, we will perform here the first (at least to our knowledge) systematic investigation of a subset of these aspects. These are the level of theory, basis set size, inclusion of solvent effects by an implicit model, and the use of partial charges around the fragments, which was shown to significantly increase the accuracy of ADMA. In the next section, first the principle of the ADMA fragmentation scheme and the chemical shift calculation based on it will be shortly outlined before the details of the parameters changed in the systematic study are described. The best combination of the parameters is then determined as described in the Results and Discussion using a small polypeptide, the HA2 domain of the influenza virus glycoprotein hemagglutinin, and is further supported by calculations of two larger system. In the Conclusion, we will finally point out possibilities for further improvements, and a scheme to combine the advantage of empirical and ab initio prediction methods is presented.

MATERIALS AND METHODS

In our previous publication,²¹ the fragmentation scheme of the adaptable density matrix assembler (ADMA)^{30–33,36} was used for accurate calculations of NMR chemical shifts for carbon atoms in proteins. For doing so the large protein is divided into small fragments such that each nucleus is included in exactly one of these fragments. These fragments are then used in independent quantum chemical calculations. But to get reasonable results, not only the fragments but also the environment of these have to be considered in these calculations. Therefore, the fragment is surrounded by additional regions (called surroundings) with the same local nuclear geometry as in the macromolecule up to a specified distance d and the resulting molecules or supermolecules are capped by hydrogen atoms to fill unsaturated bonds resulting from the fragmentation procedure. An example of a fragment with its corresponding surroundings and capping atoms, together called the parent molecule in the following, is shown in Figure 1. The accuracy of ADMA is

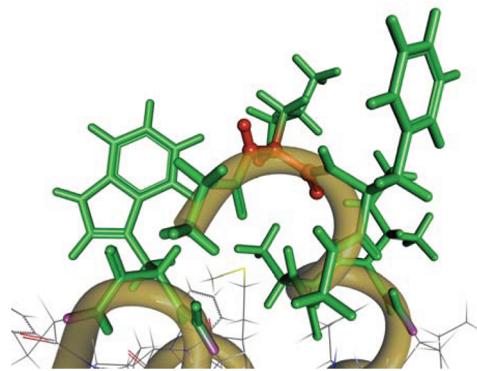


Figure 1. Representative parent molecule of the HA2 domain of the influenza virus glycoprotein hemagglutinin (PDB code 2kxa⁶⁰): The backbone atoms of Leu2 (red, balls-and-sticks) build the central fragment, for which the NMR chemical shifts are calculated using this parent molecule. All side-chain and backbone fragments (green, sticks), for which at least one atom is closer than 5 Å to the central fragment, are added as surroundings. All other atoms (color-coded, wire frame) are removed and the broken bonds are saturated with hydrogen atoms (magenta, sticks). For a more detailed description, see Frank et al.²¹

solely determined by the distance d and with $d \rightarrow \infty$ the results of the complete-molecule calculation are recovered.^{32,59}

Since the parent molecules are normal chemical entities of relatively small size, standard quantum chemical programs can be used to perform calculations on them and the results for the nuclei of the fragment (not of the surroundings and capping atoms) can then be extracted. All calculations were performed with the Gaussian 09 program package.⁶¹ In the original ADMA approach, the fragment electron density matrices are combined to build the complete density matrix of the macromolecule, from which many molecular properties can be derived.^{32,59} Unfortunately, NMR chemical shifts cannot be obtained in this way, since their calculation requires the computation of the mixed second derivative of the total electronic energy of the molecule (E) with respect to the external magnetic field \mathbf{B} and the magnetic moment of the nucleus of interest μ :

$$\sigma_{\alpha\beta} = \frac{\partial^2 E}{\partial \mu_\alpha \partial B_\beta} \quad (1)$$

where σ is the chemical shielding tensor and α and β are all possible combinations of the Cartesian coordinates x , y , and z . From this shielding tensor, the chemical shift tensor δ is then derived using σ_{iso} , the isotropic value or trace of the chemical shielding of the standard reference used in the NMR experiment, which is obtained from a separate calculation of the reference

$$\delta = \mathbf{1}\sigma_{iso} - \sigma \quad (2)$$

($\mathbf{1}$ is the unit matrix; tetramethylsilane for ^1H and ^{13}C and ammonia for ^{15}N calculated at the same level of theory were taken as reference). Getting the derivatives is even more demanding in terms of computer power and memory than the calculation of the energy; thus, that is not possible for larger proteins, for which even the energy is only available in the fragment-based approximation in a reasonable amount of time. Therefore, we decided to use the NMR chemical shifts of the atoms in a specific fragment directly from the corresponding parent molecule for our studies.²¹ In this way, our approach, even if started from a different concept, becomes similar to others based

on the QM/MM methodology^{8–29} but with the advantage of full automation paired with the flexibility to adapt, e.g., the size of the surroundings, level of theory, basis set, as well as border region description^{82,63} needed for the systematic test performed here.

The previous results showed²¹ that with reasonable surroundings of 5–6 Å the NMR chemical shifts of entire polypeptides, for which these calculations are possible and can, therefore, be used for comparison with the fragment calculations, can be reproduced with reasonable accuracy. A distance criterion of 5 Å will be used throughout this publication if not indicated otherwise. When comparing to the experiment, the GIAO (gauge invariant/including atomic orbitals) approach with B3LYP/6-31g(d) theory/basis set combination gave good correlations for all carbon atoms except carbonyl carbons but bad correlations for hydrogen, especially hydrogen atoms involved in hydrogen bonds and nitrogen. To further improve these results, we investigate here the influences of different aspects on our chemical shift calculations. First, the level of theory/basis set combination was varied. Due to the still relatively large size of the parent molecules, density functional theory is the method of choice and we compare here four different functionals. Nevertheless, MP2 calculations for some fragments were also performed since MP2 has been shown to be advantageous for nuclei located in close proximity to aromatic systems.^{52,53} But for these, smaller surroundings of 3 Å had to be used to limit the computational memory as well as scratch space demand to an amount available to us. This limitation could probably be circumvented by using newest developments like local MP2,^{42–45} tighter multipole-based integral estimates,^{46,47} and/or density fitting approaches,⁴³ which will be investigated in the near future. Second, we tried to make the model more realistic. On the one hand, NMR spectra of proteins are measured in aqueous solution. In order to include the polarization effects of the surrounding water, calculations were repeated with an implicit solvent model (IEF-PCM^{64–66}) based on the self-consistent reaction field (SCRF). On the other hand, the parts of the macromolecule, not included in the surroundings, are completely neglected in ADMA calculations. But, at least the polarization of the electron density of the parent molecule by these additional parts can be modeled by including these parts as point charges. This field-adapted ADMA approach (FA-ADMA) gave huge improvements in accuracy of the obtained energies with a given size of the surroundings especially for highly polar molecules.^{33,67} Gasteiger–Marsili charges were created using the molecular modeling software SYBYL X1.2⁶⁸ and placed on all positions of atoms neglected in the individual parent molecule calculations. They were then included in the quantum chemical calculation on their own or in combination with the PCM solvent model. Please note that at the moment the cavity used for the implicit solvent model only accommodates the quantum mechanical atoms (fragment and explicit surroundings). Thus, the high dielectric constant is also taken at the position of the atoms not part of the parent molecules, and the partial charges representing these atoms are actually embedded in the implicit solvents. Additional studies are on their way varying the cavity size and/or using solute cavities around the complete macromolecule to investigate possible interferences between the solvent model and the partial charges. Finally, the use of locally dense basis sets, i.e., the combination of two basis sets where the larger one is used for the nuclei for which the chemical shifts are calculated and the smaller one for all other nuclei, has been shown to be advantageous for NMR chemical shift calculations.^{69–71} We

will test here if this approach could lead to time savings without significantly compromising accuracy. For doing so we defined two distance criteria. The first, smaller one set to 3 Å defines the region for the large basis set. The smaller basis set is assigned to atoms with a distance between the smaller and larger distance criterion. For the larger distance criterion, the same 5 Å as in the calculation with only one basis set was used.

■ RESULTS AND DISCUSSION

a. ^1H NMR Chemical Shifts of the HA2 Domain. We started our study with the 32 amino acid long HA2 domain of the influenza virus glycoprotein hemagglutinin. The structure and the chemical shifts of the first 23 amino acids are available from the Protein Data Bank (PDB code 2kxa⁶⁰) and the Biological Magnetic Resonance Bank (BMRB accession number 16907), respectively. It shows a tight helical hairpin structure, with its N-terminal α helix (Gly1–Gly12) packed tightly against its second α helix (Trp14–Gly23) (see Figure 2). This

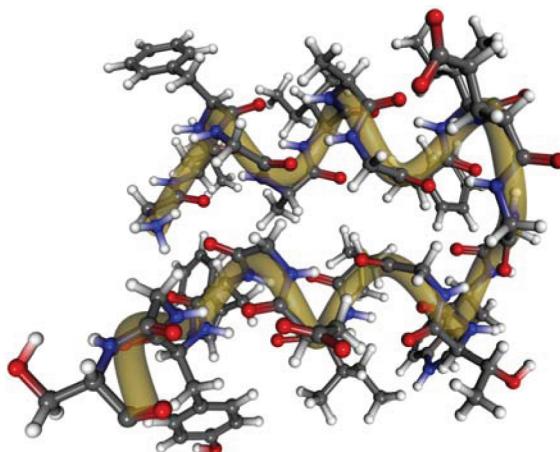


Figure 2. Structure of the HA2 domain of the influenza virus glycoprotein hemagglutinin (PDB code 2kxa).

combination of small size, well-defined structure with strong intramolecular interactions, and the availability of ^1H , ^{13}C , and ^{15}N chemical shifts render it the perfect test system. To avoid possible influences of the missing C-terminus on our comparisons, we additionally excluded residue 23 so that all our results are based only on the residues 1–22.

Results for ^1H chemical shifts (mean absolute error, MAE) using two density functionals and two differently sized basis sets are given in Figure 3 and Table 1 (corresponding values of the mean error and standard deviation of errors for this and the following tables are provided in the Supporting Information). For the larger basis set, the implicit solvent model and partial charges to describe additional parts of the peptide are used as described in the Materials and Methods. As has already been recognized in our previous publication,²¹ the agreement between the calculations and the experiments strongly increases if the polar protons are omitted. This behavior is observed in all combinations of density functionals and basis sets (see also Figure 4). It can be explained by inaccuracies in the description of hydrogen bonds and especially the neglect of explicit solvents molecules forming hydrogen bonds to the solute.^{22,58,72} Therefore, we will exclude these protons in the following. The remaining proton chemical shifts are calculated to a remarkable accuracy when using the valence triple- ζ 6-311g(d) basis set.

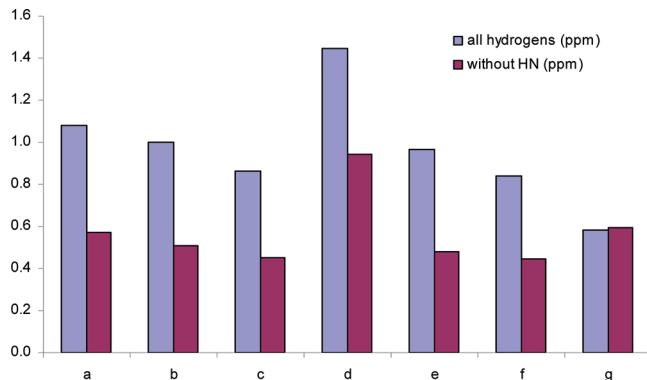


Figure 3. MAE of ^1H NMR chemical shifts of the HA2 domain including and excluding H^{N} calculated with different density functionals and basis sets as well as the empirical SHIFTX+ method.^{1,2} The description of the methods corresponding to the letters on the X-axis is given in Table 1.

The mPW1PW91/6-311g(d) combination gives the best results (MAE = 0.44 ppm) considerably better than the structure-based component (SHIFTX+) of the empirical SHIFTX2 method^{1,2} (MAE = 0.61 ppm, the sequence-based component SHIFTY+^{1,3} was deactivated since some of the proteins used in this study are part of RefDB,⁷³ which is used for sequence matching and alignment, so that the exact chemical shifts of these protein are available to SHIFTY+). This superior behavior can also be seen in the correlation in Figure 4 ($R^2 = 0.95$ vs 0.91). The deviation of the slope from its optimal value 1 can

be explained by the aromatic protons, which are predicted to be too low-field shifted, while the aliphatic ones are too high-field shifted. This fact was also noticed in the publication of Sarotti and Pellegrinet⁷⁴ and was in their study accounted for by the use of two different standards for aromatic and aliphatic protons. In Figure 5 the spatial distribution of errors is presented. SHIFTX+ generally shows larger errors (SHIFTX+, maximum error = 1.83 ppm; mPW1PW91/6-311g(d), maximum error = 1.46 ppm) distributed throughout the molecule, even if more flexible, solvent-exposed regions seem to be more problematic. The largest errors of the quantum chemical calculation are concentrated in one region in the upper left part of the figure, suggesting possible structural uncertainties in this region. These could originate from lower conformational definition (i.e., molecular flexibility) or even slight errors in the 3D structure.

b. ^{13}C NMR Chemical Shifts of the HA2 Domain.

Motivated by the very encouraging results for ^1H chemical shifts, a systematic investigation of all the different parameters described in the Materials and Methods (level of theory, basis set, inclusion of solvent effects, and additional parts as point charges) was performed on ^{13}C chemical shifts. An overview of the MAEs of ^{13}C NMR chemical shifts from the experimental values calculated with different density functionals and basis set combinations is given in Figure 6 and Table 2. Additionally, the results of the two empirical methods, SPARTA+ (version 1.12)^{4,5} and SHIFTX+ (from SHIFTX2, version 1.02),^{1,2} are listed. SPARTA+ is a sequence-based method like SHIFTY+^{1,3} and is only parametrized for backbone and C^{β} atoms. To compare to this method, results for this subset are also shown.

Table 1. MAE of ^1H NMR Chemical Shifts of the HA2 Domain Including and Excluding H^{N} Calculated with Different Density Functionals and Basis Sets as Well as the Empirical SHIFTX+ Method^{1,2} ^a

	method	basis set	implicit solvent	charges	all hydrogens (ppm)	without H^{N} (ppm)
a	B3LYP	6-31g(d)	–	–	1.08	0.57
b		6-311g(d)	–	–	1.00	0.51
c			+	+	0.86	0.45
d	mPW1PW91	6-31g(d)	–	–	1.45	0.94
e		6-311g(d)	–	–	0.96	0.48
f			+	+	0.84	0.45
g	SHIFTX+				0.58	0.59

^aCalculations with implicit solvents and/or charges to describe additional surroundings are marked with a “+” in the corresponding column.

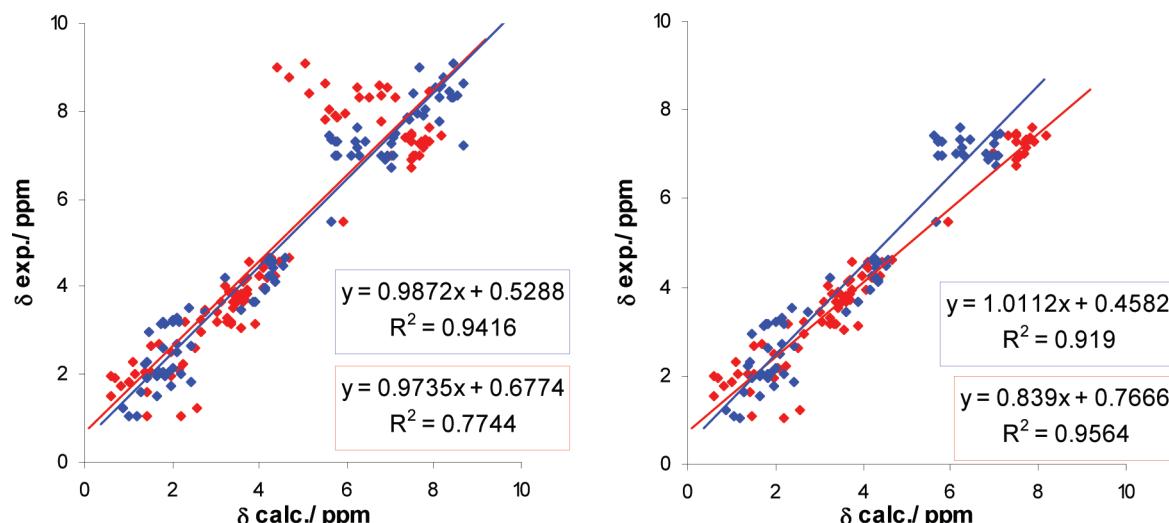


Figure 4. Correlation of ^1H NMR chemical shifts of the HA2 domain: blue, SHIFTX+; red, mPW1PW91/6-311g(d) with implicit solvents and partial charges; left, including H^{N} ; right, excluding H^{N} .

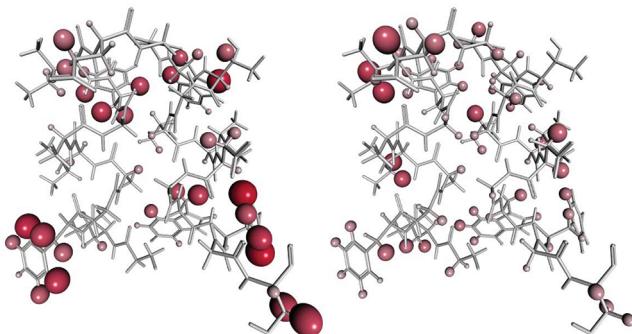


Figure 5. Spatial distribution of the errors in the ^1H chemical shifts of the HA2 domain: The values are coded by the color and size of the spheres representing the atoms (darker red color and larger spheres correspond to larger error); left, SHIFTX+ (maximum error = 1.83 ppm); right, mPW1PW91/6-311g(d) with implicit solvent and charges for additional surroundings (maximum error = 1.46 ppm).

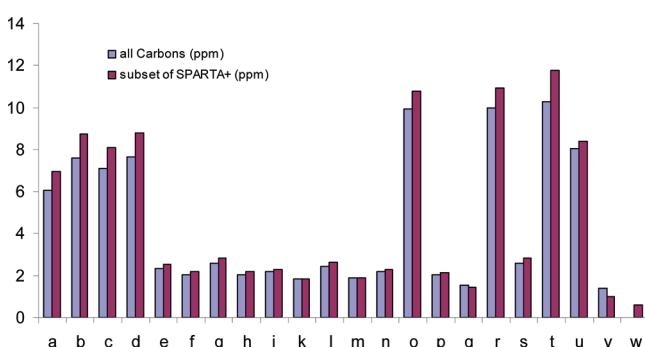


Figure 6. MAE of ^{13}C NMR chemical shifts of the HA2 domain calculated with different density functionals and basis sets as well as two empirical methods. Additionally to the results for all carbons, for which experimental values are available, results for the SPARTA+ subset (backbone and some C^β) are given. The description of the methods corresponding to the letters on the X-axis is given in Table 2.

For the small basis sets [6-31g and 6-31g(d)] and the B3LYP functional relatively large MAEs are obtained. But as can be seen in Figure 7, the calculated values nevertheless correlate reasonably well with experiment. In comparison to the empirical SHIFTX+ method, it becomes evident that all values and especially the carbonyl carbons are high-field shifted. This was already described in our previous publication.²¹ Taking the carbonyl carbon atoms out of the correlation, the slope of the correlation line is approaching the optimal value of 1. The combination of two different basis sets, with the smaller 3-21g(d) basis used for the surroundings further than 3 Å away from the fragment, only insignificantly worsened the results. This supports the usefulness of locally dense basis sets for NMR calculations, even if the overall accuracy for these small basis sets is probably not sufficient. The implicit solvent model improved the results slightly to an MAE of 7.11 ppm (without solvents, MAE = 7.59) mainly caused by a slight downfield shift of the carbonyl carbon atoms compared to the calculations without solvents.

It is surprising that for the smaller 6-31g basis set the MAE (6.07 ppm) is more than 1 ppm lower than for 6-31g(d) (MAE = 7.59 ppm). This can be only explained by a favorable cancellation of errors. But it is a well-known fact^{48,49,75} that relatively large basis sets are needed for accurate chemical shift calculations.

As just shown, enlarging the basis set by polarization functions (6-31g vs 6-31g(d) with one set of d polarization functions on all second-row elements) is at least not very helpful and in this one case even detrimental. Therefore, improving the basis set by moving from a valence double- ζ to a valence triple- ζ basis set (decontraction) was applied, which was also key for the improvements in the ^1H chemical shifts. With the 6-311g(d) basis set, the preferred effect with an almost 4-fold decrease in the MAE (2.18 ppm) was reached. This improvement results mainly from the much better description of the carbonyl carbon atoms, which now show the same level of accuracy as the other atoms (see Figure 8). Here, the inclusion of polarization functions is beneficial in comparison to the 6-311g basis set [6-311g, MAE = 2.35 ppm vs 6-311g(d), MAE = 2.18 ppm], even if the improvement is not that pronounced as upon decontraction of the basis set. Using a locally dense basis set [6-311g(d) for the fragment and surroundings up to 3 Å and 6-31g(d) for surroundings up to 5 Å] led again to an insignificant increase in the errors, which is very satisfactory since much computer power can be saved by such mixed basis set calculations.

The treatment of solvent effects by the PCM continuum model improved the accuracy even further, resulting in an MAE of less than 2 ppm. In contrast, including additional parts of the system as partial charges according to FA-ADMA was not beneficial. FA-ADMA performed best in comparison to the original ADMA for highly polar molecules.^{33,67} The HA2 domain only contains two formally charged amino acids and can therefore be classified as only weakly polar, and accordingly, combining partial charges and the continuum solvent model shows a very similar result as solvent only. Nevertheless, since partial charges will be favorable for polar molecules, we will continue to use the combination for the rest of this publication.

After 6-311g(d) has been identified as well-suited, we studied now the influence of the level of theory. Besides B3LYP used so far, three other density functionals were tried since they have been used already in comparison studies of small molecules:^{48,49,75} HCTH,^{76–78} VSXC,⁷⁹ and mPW1PW91.⁸⁰ HTCH and especially VSXC gave consistently considerably worse results. This is surprising since VSXC was explicitly mentioned to perform extremely well in a recent study comparing density functionals for the calculation of small molecule NMR chemical shifts.⁴⁹ mPW1PW91 shows the best performance over all calculations performed so far with an MAE of 1.53 ppm when combined with the 6-311g(d) basis set, charges for additional surroundings, and the implicit solvent model. In this calculation, no nucleus has an error larger than 5 ppm compared to four nuclei with errors larger than 5 ppm when using B3LYP. Only the offset of the correlation line is a little bit higher (0.75 ppm) than in the corresponding B3LYP calculation (0.41 ppm) suggesting a slightly larger systematic error (see Figure 9).

This accuracy is now very close to the empirical methods. Even if SHIFTX+ and SPARTA+ still outperform our approach, if one concentrates on the SPARTA+ subset of nuclei (backbone + C^β), the MAE is only 0.15 ppm larger in ADMA than in SHIFTX+, if all carbon nuclei for which experimental chemical shifts are available are considered. This shows that the quantum chemical calculations have a very uniform accuracy over all nuclei while the empirical methods perform better on the backbone than for the side chains. This can be explained by the larger variability of side-chain conformations and the statistical problems arising from these during the empirical parametrization. Finding a representative data set for other biomolecules like DNA and also small-molecule ligands will be even harder, which gives us

Table 2. MAE of ^{13}C NMR Chemical Shifts of the HA2 Domain Calculated with Different Density Functionals and Basis Sets as Well as Two Empirical Methods^{a,b}

	method	basis set	implicit solvent	charges	MAE (ppm)	
					all carbons	subset of SPARTA+
a	B3LYP	6-31g	—	—	6.07	6.95
b		6-31g(d)	—	—	7.59	8.73
c			+	—	7.11	8.08
d		6-31g(d)/3-21g(d)	—	—	7.65	8.79
e		6-311g	—	—	2.35	2.53
f			+	—	2.06	2.17
g			—	+	2.57	2.81
h			+	+	2.05	2.17
i		6-311g(d)	—	—	2.18	2.29
jk			+	—	1.85	1.85
l			—	+	2.44	2.63
m			+	+	1.87	1.88
n		6-311g(d)/6-31g(d)	—	—	2.17	2.28
o	mPW1PW91	6-31g(d)	—	—	9.91	10.76
p		6-311g(d)	—	—	2.05	2.13
q			+	+	1.53	1.44
r	HCTH	6-311g(d)	—	—	9.98	10.94
s			+	+	2.60	2.84
t	VSXC	6-31g(d)	—	—	13.90	11.77
u		6-311g(d)	+	+	8.01	8.41
v	SHIFTX+ ^c				1.38	0.98
w	SPARTA+					0.59

^aAdditionally to the results for all carbons, for which experimental values are available, results for the SPARTA+ subset (backbone and some C^β) are given. ^bCalculations with implicit solvents and/or charges to describe additional surroundings are marked with a “+” in the corresponding column.

^cInteresting to note is the fact that for an older version of SHIFTX+ considerably worse results are obtained, documenting the improvements obtained by new parametrizations based on more experimental data (data not shown).

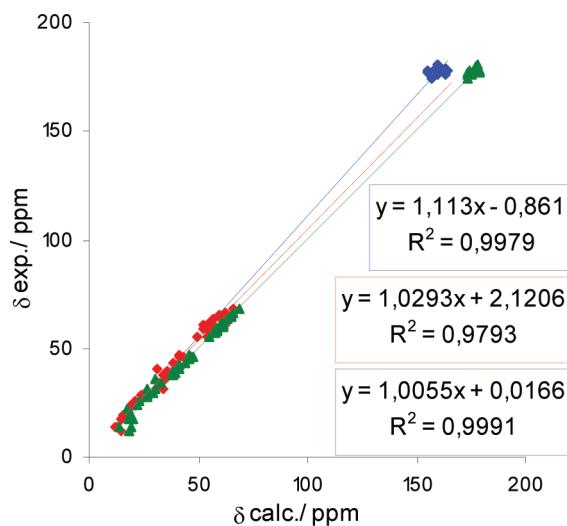


Figure 7. Correlation of ^{13}C NMR chemical shifts of the HA2 domain: blue, complete protein with B3LYP/6-31g(d); red, without carbonyl carbon atoms with B3LYP/6-31g(d); green, complete protein with SHIFTX+.

confidence that especially for these kinds of molecules and molecular complexes our method could be extremely valuable.

Fragments with parent molecules up to a size of around 100 atoms could be calculated in a reasonable amount of time using MP2 and the small 6-31g(d) basis set as well as the combination of 6-311g(d) for the fragment and 3-21g for the surroundings (computation time was on the order of 60 days for Gly12 of the HA2 domain. Calculations with identical basis

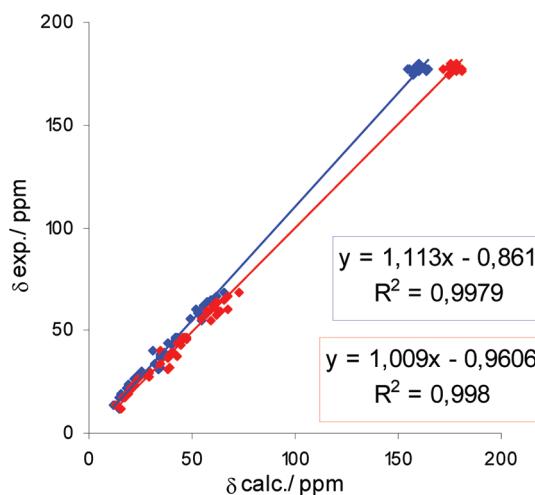


Figure 8. Correlation of ^{13}C NMR chemical shifts of the HA2 domain: blue, B3LYP/6-31g(d); red, B3LYP/6-311g(d).

sets and surroundings as in our DFT calculations would have required at least 100-fold processor time not to mention memory size). Therefore, only results for 23 out of 36 fragments with a small surrounding of 3 Å can be presented here and are given in Figure 10 and Table 3. For comparison also DFT and SHIFTX+ results are shown for this subset. While for the small basis set B3LYP (MAE = 6.88 ppm) gives slightly worse results than MP2 (MAE = 6.56), B3LYP and mPW1PW91 with the larger basis sets and SHIFTX+ perform much better with MAEs around 2 ppm or smaller. Using the large basis set with MP2 also leads to an improvement but not as strong as for the DFT

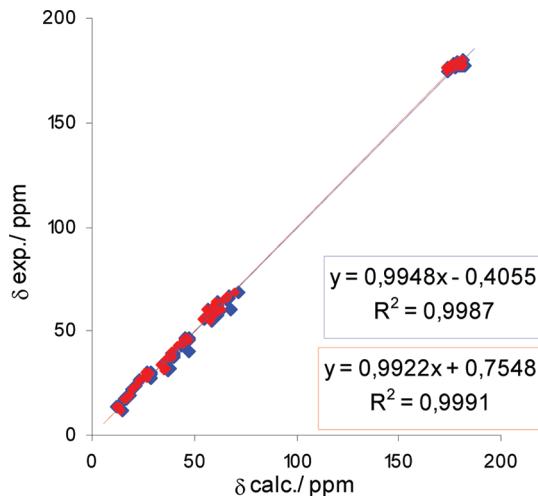


Figure 9. Correlation of ^{13}C NMR chemical shifts of the HA2 domain with implicit solvent and charges for additional surroundings: blue, B3LYP/6-311g(d); red, mPW1PW91/6-311g(d).

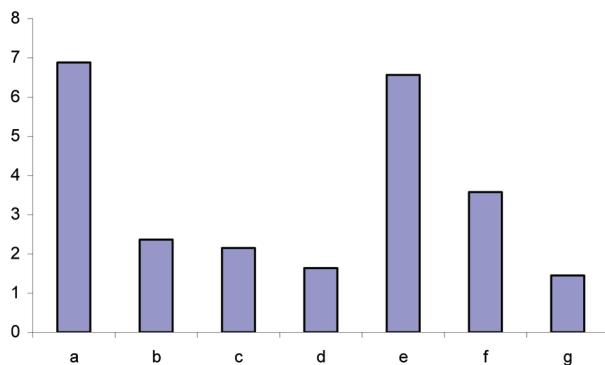


Figure 10. MAE of ^{13}C NMR chemical shifts for the subset of fragments of the HA2 domain, which could be calculated up to the MP2 level of theory. The description of the methods corresponding to the letters on the X-axis is given in Table 3.

calculations. This demonstrates that for ^{13}C chemical shifts the use of a large basis set is more important than the use of higher levels of theory. But it should be kept in mind that a mixed basis set was used in the MP2 calculations and even more important the parent molecules of the MP2 calculations are much smaller than the ones in the DFT calculations (3 and 5 Å surroundings, respectively); thus, long-range interactions are neglected, contributing to the errors seen in the MP2 results.

Concluding this first part, the results of the best three methods, SHIFTX+, B3LYP/6-311g(d) with implicit solvent and partial charges, and mPW1PW91/6-311g(d) with implicit solvent and partial charges, are visualized on the molecular structure in Figure 11. SHIFTX+ shows the largest number of nuclei with small errors almost not visible in the graphical representation. But for some of the nuclei considerably larger errors are observed. B3LYP shows a very similar distribution even if all the errors are slightly higher than for SHIFTX+. Finally, mPW1PW91 gives the narrowest distribution of errors. There are no deviations of more than 5 ppm. At the current state SHIFTX+ still slightly outperforms the quantum chemical calculations if the MAE is used as a measure due to the very small number of relatively big failures, but the quantum chemical methods are undoubtedly catching up.

Table 3. MAE of ^{13}C NMR Chemical Shifts for the Subset of Fragments of the HA2 Domain, Which Could Be Calculated up to the MP2 Level of Theory^a

	method	basis set	implicit solvent	charges	subset of MP2 (ppm)
a	B3LYP	6-31g(d)	—	—	6.88
b		6-311g(d)	—	—	2.36
c			+	+	2.15
d	mPW1PW91	6-311g(d)	+	+	1.64
e	MP2	6-31g(d)	—	—	6.56
f		6-311g(d)/3-21g	—	—	3.57
g	SHIFTX+				1.45

^aCalculations with implicit solvents and/or charges to describe additional surroundings are marked with a “+” in the corresponding column.

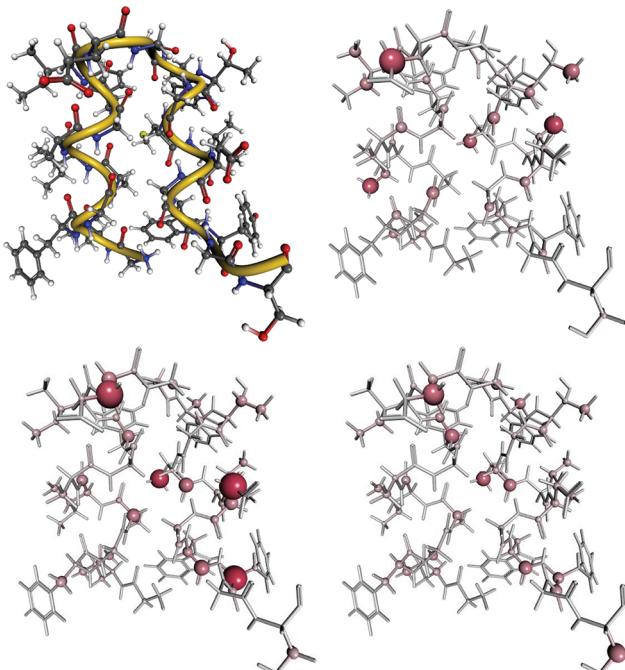


Figure 11. Spatial distribution of ^{13}C chemical shift errors of the HA2 domain; the values are coded by the color and size of the spheres representing the atoms (darker red color and larger spheres correspond to larger errors): upper left, structure of the HA2 domain; upper right, SHIFTX+ (maximum error = 6.54 ppm); lower left, B3LYP/6-311g(d) with implicit solvent and charges for additional surroundings (maximum error = 7.40 ppm); lower right, mPW1PW91/6-311g(d) with implicit solvent and charges for additional surroundings (maximum error = 4.98 ppm).

For none of the methods a clear conclusion can be drawn regarding the spatial distribution of errors. They are distributed throughout the molecule and not limited to some specific parts as solvent-exposed or highly flexible regions. There is also only a weak correlation between the deviations in the two calculations using different density functions (data not shown). Only the large errors in the upper left part of Figure 11 are reoccurring in all calculations. This region also shows the largest deviation of ^1H chemical shifts supporting the assumption of a structural cause of these errors.

c. ^{13}C NMR Chemical Shifts of p63 DNA-Binding Domain and the Complex of Tfb1 with a p53 Fragment. To verify the results obtained so far, we performed additional calculations on two larger systems (see Figure 12 and Table 4): p63

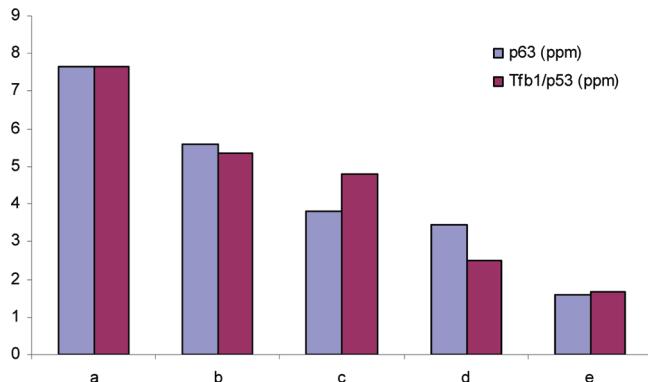


Figure 12. MAE of ^{13}C NMR chemical shifts of p63 and the complex of Tfb1 with a p53 fragment calculated with different density functionals and basis sets as well the empirical SHIFTX+ method.^{1,2} The description of the methods corresponding to the letters on the X-axis is given in Table 4.

Table 4. MAE of ^{13}C NMR Chemical Shifts of p63 and the Complex of Tfb1 with a p53 Fragment Calculated with Different Density Functionals and Basis Sets as Well as the Empirical SHIFTX+ Method^{1,2,a}

	method	basis set	implicit solvent	charges	p63 (ppm)	Tfb1/p53 (ppm)
a	B3LYP	6-31g(d)/3-21g(d)	—	—	7.66 ^b	7.65 ^b
b		6-311g(d)	—	—	5.58	5.34
c			+	+	3.81	4.79
d	mPW1PW91	6-311g(d)	+	+	3.44	2.51
e	SHIFTX+				1.57	1.67

^aCalculations with implicit solvents and/or charges to describe additional surroundings are marked with a “+” in the corresponding column. ^bResults are taken from Frank et al.²¹

DNA-binding domain (PDB code 2rmn;⁸¹ chemical shifts were provided by Horst Kessler, private communication) and the complex of Tfb1 with a p53 fragment (PDB code 2gs0,⁸² BMRB accession number 6225). These show the same trend observed for the HA2 domain in that a valence triple- ζ basis set is needed for acceptable results, especially for carbonyl carbon atoms. Additionally, the mMPW1PW91 functional again gives the best results almost halving the error in the p63 example in comparison to B3LYP. But even the best results give an MAE of 3.4 and 2.5 ppm for p63 and Tfb1/p53, respectively, while SHIFTX+ performs very similar to the small test example (MAE between 1.5 and 1.7 ppm). Thus, if only MAE is considered, one would conclude that the empirical method is outperforming the quantum chemical calculations, but it becomes evident when looking at the spatial distribution of the errors that all large errors in the quantum chemical calculations especially for TFB1/p53 are located in specific regions (see Figure 13). Due to the size of the proteins, the structure determination is more complex than for the small HA2 domain, and some parts of the structures are less well-defined, as documented by variations in the NMR ensemble given in the PDB files. Since the quantum chemical calculations are more sensitive to conformational changes, it could also be argued that these regions indicate experimental uncertainties. Only the first structure of the ensemble was taken for the study presented here, but later structures could give better results at least for some regions of the proteins. In the p63 case, the main errors

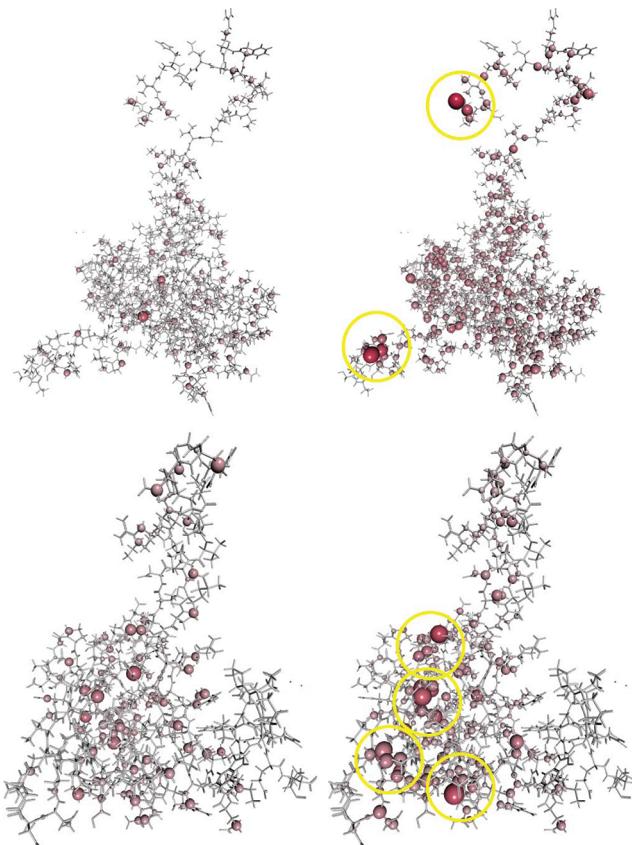


Figure 13. Spatial distribution of errors in the ^{13}C chemical shifts; the values are coded by the color and size of the spheres representing the atoms (darker red color and larger spheres correspond to larger errors), and regions with large deviations are marked with yellow circles: upper left, p63, SHIFTX+ (maximum error = 11.02 ppm); upper right, p63, mPW1PW91/6-311g(d) with implicit solvent and partial charges (maximum error = 15.99 ppm); lower left, Tfb1/p53, SHIFTX+ (maximum error = 9.38 ppm); lower right, Tfb1/p53, mPW1PW91/6-311g(d) with implicit solvent and partial charge (maximum error = 12.60 ppm).

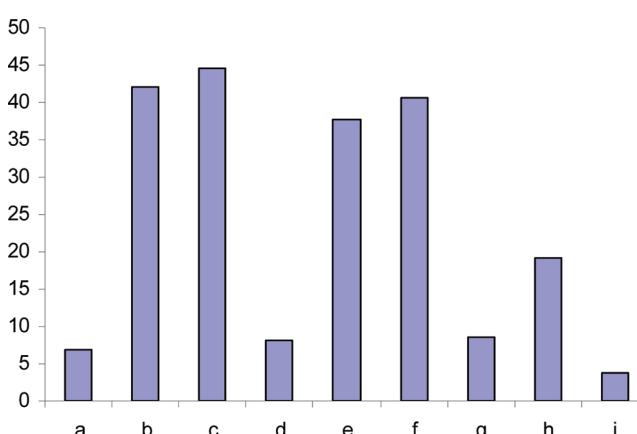


Figure 14. MAE of ^{15}N NMR chemical shifts of the HA2 domain calculated with different density functionals and basis sets as well the empirical SHIFTX+ method.^{1,2} The description of the methods corresponding to the letters on the X-axis is given in Table 5.

are located in the flexible C- and N-terminus, and for these regions, the arguments just given also apply. But, additionally, it seems that the carbonyl carbons are still predicted

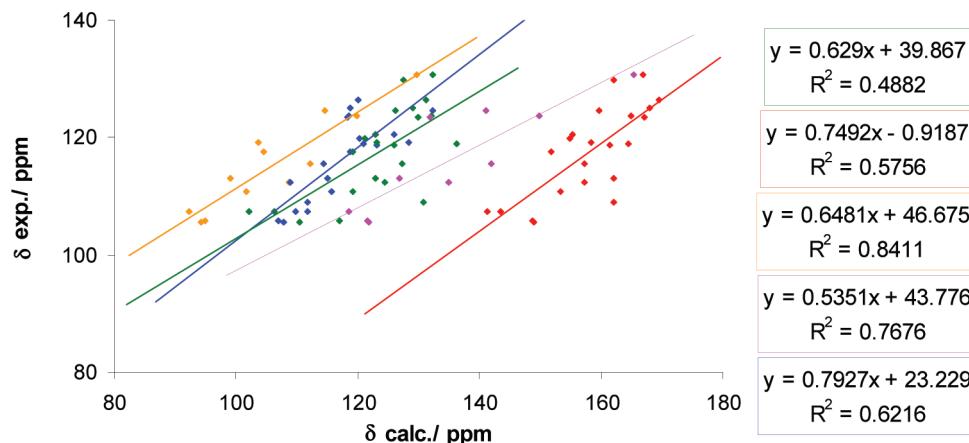


Figure 15. Correlation of ^{15}N NMR chemical shifts of the HA2 domain: green, B3LYP/6-31g(d); red, mPW1PW91/6-311g(d) with implicit solvents and partial charges; orange, MP2/6-31g(d); magenta, MP2/6-311g(d)/3-21g; blue, SHIFTX+.

too high-field shifted for this example (even if not as strong as with the small basis set) while the other carbon atoms are too low-field shifted. Additional investigations will show if this can be improved by the use of different standards, as described above.

d. ^{15}N NMR Chemical Shifts of the HA2 Domain.

Finally, we looked at the ^{15}N chemical shifts (see Figure 14 and Table 4). B3LYP with the small basis set [6-31g(d)] gives a very comparable MAE as for ^{13}C chemical shifts. But here the decontraction of the basis set did not improve the results. On the contrary, extreme deviations to the experiment are obtained with the valence triple- ζ basis set [6-311g(d)]. These can partly be explained by a large systematic error (see Figure 15), which extremely shifts all values to lower field quantified by the mean error (ME; see Supporting Information). While it is very small for all ^1H and ^{13}C chemical shift calculation with the 6-311g(d) basis set ($-1 \text{ ppm} < \text{ME} < 1 \text{ ppm}$), in the case of ^{15}N the ME (around -40 ppm and below) is in the same range as the MAE pointing toward a significant systematic error. In principle, this systematic error can be removed by using an internal standard. Obviously, even the standard used here (NH_3) caused a large share of the systematic errors observed in our calculations. While for 6-31g(d) an isotropic chemical shielding of 267.38 ppm is calculated for ammonia, 6-311g(d) gives a value of 285.27 ppm. The chemical shielding of the nuclei of the protein is much more similar in the two basis sets than this almost 20 ppm difference, and shifting of all chemical shifts by this value would highly decrease the MAE. But also the correlation is worse for the larger basis set than for the smaller one or SHIFTX+, even if also the latter results are not convincing. The MP2 calculations with the 6-31g(d) basis set show exactly the opposite behavior compared to the valence triple- ζ basis set in DFT. Here, all values are systematically shifted to higher field (MAE = 8.49 ppm, ME = 7.85 ppm). Even if the MP2 results are only based on a smaller number of ^{15}N chemical shifts (subset of 23 out of 36 fragments, which could be calculated at the MP2 level of theory) and the MAE for SHIFTX+ is almost 50% lower, we would like to note that the obtained correlation is the best over all calculations including the empirical SHIFTX+. Increasing the size of the basis set [mixed basis set with 6-311g(d) for the fragment] again deteriorates the accuracy. This has multiple possible reasons: First, the very small surroundings used are not sufficient for accurate approximations. Second, the use of the mixed basis set may lead to basis set superposition

errors. Third, ^{15}N is more sensitive to changes in the molecular structure and conformational averaging has to be considered, which is the scope of ongoing investigations. Taking all these results together, the favorable cancellation of errors, letting B3LYP/6-31g(d) appear to be one of the best quantum chemical methods, can be explained by the compensation of the high-field and low-field shift tendency of the small basis set and density functional theory, respectively. Both these effects can perhaps be circumvented by using MP2 with at least a valence triple- ζ basis set, sufficient surroundings, and conformational averaging as well as solvent effects taken into account. But such calculations are currently too demanding with respect to computer power but also main memory and disk space to be used in a standard approach.

Table 5. MAE of ^{15}N NMR Chemical Shifts of the HA2 Domain Calculated with Different Density Functionals and Basis Sets as Well as the Empirical SHIFTX+ Method.^{1,2 a}

	method	basis set	implicit solvent	charges	all nitrogens (ppm)
a	B3LYP	6-31g(d)	—	—	6.77
b		6-311g(d)	—	—	42.14
c			+	+	44.60
d	mPW1PW91	6-31g(d)	—	—	8.09
e		6-311g(d)	—	—	37.60
f			+	+	40.582
g	MP2	6-31g(d)	—	—	8.49 ^b
h		6-311g(d)/3-21g	—	—	19.14 ^b
i	SHIFTX+				3.84

^aCalculations with implicit solvents and/or charges to describe additional surroundings are marked with a “+” in the corresponding column. ^bMAE was calculated for the subset of 23 out of 36 fragments, which could be calculated with MP2 only.

CONCLUSION

The fragment-based calculations using the ADMA approach presented here are for the first time able to very accurately predict NMR chemical shifts of proteins on a quantum chemical basis. Mean absolute errors as low as 0.44 ppm for hydrogens excluding H^{N} and between 1.53 and 3.44 ppm for carbon atoms compare very favorable to the corresponding

values of 0.60 and around 1.5 for the empirical program SHIFTX+. These results were obtained with the mPW1PW91 density functional, the 6-311g(d) valence triple- ζ basis set, the PCM implicit solvent model, and partial charges to describe additional parts of the macromolecule in the fragment calculations. Especially the triple- ζ basis set helped to circumvent the problem of high-field shifted carbonyl carbons seen in our previous publication.²¹ These excellent results are especially remarkable when compared to studies of small molecules. Auer et al.⁵⁰ state that there is no alternative to CCSD(T) calculations paired with a very large basis set and consideration of vibrational effects to get accurate chemical shifts for very small molecules in the gas phase. Even with this method the deviations are still in the range of 1–2 ppm for carbon atoms. Sefzik et al.⁴⁸ compared 102 ¹³C chemical shift tensors in single crystals from a series of aromatic and saccharide molecules and obtained root-mean-square deviations (rmsd) around 5 ppm for similar functional/basis set combinations as used here. Finally, we mention here the study of Kupka et al.,⁴⁹ which compared a large number of density functionals. Interestingly, VSXC was identified as one of the best functionals for ¹³C chemical shifts in this study (rmsd of around 5 ppm) much better than B3LYP (rmsd of around 15 ppm). This is in large contrast to the results presented here.

The prediction of ¹H and nitrogen atoms needs additional improvements. Points not analyzed here in detail but a topic of additional, ongoing research are conformational averaging including zero-point vibrational effects (zpv)^{23,26,50–52,54–57} and explicit interactions with molecules of the first solvent shell.^{52,55,57,58} Auer et al.⁵⁰ have shown that such zpv effects cannot be neglected and decrease the error for small molecules compared to their gas-phase experimental values by more than 2 ppm to a mean deviation of only 1.6 ppm using CCSD(T) with a very large basis set. The main reason for the errors in the polar hydrogen atoms is the neglect of explicit solvent molecules, which cannot be included so easily due to the need of conformational sampling. ¹H chemical shifts are exquisitely sensitive to the degree and geometry of hydrogen bonds formed within the protein and with water.^{22,58,72} Additional research is on its way to quantify this effect by including explicit solvent molecules and conformational averaging. For nitrogens, density functional theory in combination with larger basis sets shows very high systematic but also high stochastic errors independent of the density functional used. This is in accordance with the literature^{25–29} and statements therein that ¹⁵N chemical shifts are much harder to predict than those of ¹³C due to their stronger dependence on structural features and electrostatics of neighboring groups as well as solvent effects. The relatively good results of the small basis set can only be explained by a very favorable cancellation of errors. Due to the semiempirical nature of the development of density functionals (see e.g. ref 83) it is very hard to find reasons for the cancellation of errors also leading to the very small mean absolute deviations for ¹H and ¹³C chemical shifts. We can only speculate that the limited amount of different functional groups in proteins leads to comparable systematic errors made by the density functionals in the calculation of the different nuclei in the target molecule as well as in the standard, which then cancel in the calculation of the chemical shifts as the difference of the chemical shieldings. For ¹⁵N, problems in the functionals are manifested more strongly if larger basis sets are used, probably caused by different behaviors in the calculations of the protein and the standard (NH_3) potentially due to the different size of

the molecules. Post-Hartree–Fock methods like MP2 are methods of choice if DFT fails. But MP2 with large basis sets drastically increases the computational demand and are only possible for very small surroundings with our actual setup not leading to the accuracy we hoped for. As already described in the Introduction, large improvements are made in the field of linear-scaling quantum chemical approaches^{37,42–47,84} enabling calculations on system sizes of a few hundreds of atoms even with very high levels of theory,^{45–47,84} which is in the same range as the parent molecules using surroundings of 5–6 Å of the DFT calculations presented here. Modifications to the ADMA procedure are on their way to use other programs as the quantum chemical engine for the parent molecules allowing such advanced calculations. Additionally, density functionals (e.g., the functionals by Kean and Tozer⁷⁵) combined with basis sets (e.g., the pcS and pcJ hierarchies of Jensen^{85,86}) explicitly developed for chemical shift calculation emerging in the recent literature appear as an interesting alternative.^{83,87}

As just described, one complication of the quantum chemical calculations is the need to use isotropic chemical shielding values of the standard (tetramethylsilane and ammonia) to derive the chemical shifts from the chemical shielding tensors (see eq 2). As seen in the case of nitrogen with a large basis set, this procedure can result in a very large offset even if the correlation between experimental and theoretical chemical shifts is reasonable. But in principle the correlation could also be used to obtain the value of the standard as intersection of the correlation line and the Y-axis, at least if the slope of the correlation line is reasonably close to 1. By referencing calculated chemical shifts in this way, offset errors might be circumvented. This could even be extended to multiple standards for different atoms, like aromatic and aliphatic carbon atoms, as done by Sarotti and Pellegrinet.⁷⁴ For some of the correlations shown (especially for hydrogen atoms; see Figure 4) such a procedure would be advantageous even if it renders the method more empirical.

Despite the many possibilities pursuable for further improvements just discussed, the current state of accuracy will already be useful for a number of applications especially if no other method is available. For standard proteins, empirical methods are privileged by their much lower demand on computational resources combined with comparable or slightly higher accuracy. But if interactions with other molecules like DNA/RNA, carbohydrates, or other organic molecules are important, quantum chemical calculations are the only possibility. To combine the advantages of all prediction methods, subsequent work will try to use the most appropriate method for specific parts of the proteins. Starting with sequence-based methods for parts with high homology over empirical structure-based methods for standard parts with low homology to ab initio methods for nonstandard parts, the highest achievable accuracy, speed, and flexibility can be obtained. Due to the fragment-based nature of ADMA, such a combination can easily be achieved.

ASSOCIATED CONTENT

S Supporting Information

Additional information as mentioned in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49-7531-88-2015. Fax: +49-7531-88-3587. E-mail: thomas.exner@uni-konstanz.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Konstanz Research School Chemical Biology (KoRS-CB), the Zukunftskolleg and the Young Scholar Fund of the Universität Konstanz, and the Juniorprofessoren-Programm of the state Baden-Württemberg. Additionally, we thank the Common Ulm Stuttgart Server (CUSS) and the Baden-Württemberg grid (bwGRID), which is part of the D-Grid system, for providing the computer resources, making the computations possible. Prof. Horst Kessler is gratefully acknowledged for providing the chemical shifts of p63.

REFERENCES

- (1) Han, B.; Liu, Y.; Ginzinger, S.; Wishart, D. SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR* **2011**, *50*, 43–57.
- (2) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and accurate calculation of protein ^1H , ^{13}C and ^{15}N chemical shifts. *J. Biomol. NMR* **2003**, *26*, 215–240.
- (3) Wishart, D. S.; Watson, M. S.; Boyko, R. F.; Sykes, B. D. Automated ^1H and ^{13}C chemical shift prediction using the BioMagResBank. *J. Biomol. NMR* **1997**, *10*, 329–336.
- (4) Shen, Y.; Bax, A. SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **2010**, *48*, 13–22.
- (5) Shen, Y.; Bax, A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* **2007**, *38*, 289–302.
- (6) Iwadate, M.; Asakura, T.; Williamson, M. C $^\alpha$ and C $^\beta$ carbon-13 chemical shifts in proteins from an empirical database. *J. Biomol. NMR* **1999**, *13*, 199–211.
- (7) Meiler, J. PROSHIFT: Protein chemical shift prediction using artificial neural networks. *J. Biomol. NMR* **2003**, *26*, 25–37.
- (8) Oldfield, E. Chemical shifts in amino acids, peptides, and proteins: From quantum chemistry to drug design. *Annu. Rev. Phys. Chem.* **2002**, *53*, 349–378.
- (9) Mulder, F. A. A.; Filatov, M. NMR chemical shift data and ab initio shielding calculations: Emerging tools for protein structure determination. *Chem. Soc. Rev.* **2010**, *39*.
- (10) Casabianca, L. B.; de Dios, A. C. Ab initio calculations of NMR chemical shifts. *J. Chem. Phys.* **2008**, *128*, 052201–052210.
- (11) Vila, J. A.; Arnautova, Y. A.; Martin, O. A.; Scheraga, H. A. Quantum-mechanics-derived ^{13}C chemical shift server (CheShift) for protein structure validation. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 16972–16977.
- (12) Sun, H.; Oldfield, E. Tryptophan chemical shift in peptides and proteins: A solid state carbon-13 nuclear magnetic resonance spectroscopic and quantum chemical investigation. *J. Am. Chem. Soc.* **2004**, *126*, 4726–4734.
- (13) Vila, J. A.; Aramini, J. M.; Rossi, P.; Kuzin, A.; Su, M.; Seetharaman, J.; Xiao, R.; Tong, L.; Montelino, G. T.; Scheraga, H. A. Quantum chemical C $^\alpha$ chemical shift calculations for protein NMR structure determination, refinement, and validation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14389–14394.
- (14) Vila, J. A.; Scheraga, H. A. Factors affecting the use of C $^\alpha$ chemical shifts to determine, refine, and validate protein structures. *Proteins* **2008**, *71*, 641–654.
- (15) Vila, J. A.; Arnautova, Y. A.; Scheraga, H. A. Use of C $^\alpha$ chemical shifts for accurate determination of β -sheet structures in solution. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1891–1896.
- (16) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. Use of C $^\alpha$ chemical shifts in protein structure determination. *J. Phys. Chem. B* **2007**, *111*, 6577–6585.
- (17) Jacob, C. R.; Visscher, L. Calculation of nuclear magnetic resonance shieldings using frozen-density embedding. *J. Chem. Phys.* **2006**, *125*, 194104.
- (18) Lee, A. M.; Bettens, R. P. A. First principles NMR calculations by fragmentation. *J. Phys. Chem. A* **2007**, *111*, 5111–5115.
- (19) Johnson, E. R.; DiLabio, G. A. Convergence of calculated nuclear magnetic resonance chemical shifts in a protein with respect to quantum mechanical model size. *J. Mol. Struct. THEOCHEM* **2009**, *898*, 56–61.
- (20) He, X.; Wang, B.; Merz, K. M. Jr. Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B* **2009**, *113*, 10380–10388.
- (21) Frank, A.; Onila, I.; Möller, H. M.; Exner, T. E. Toward the quantum chemical calculation of nuclear magnetic resonance chemical shifts of proteins. *Proteins* **2011**, *79*, 2189–2202.
- (22) Hori, S.; Yamauchi, K.; Kuroki, S.; Ando, I. Proton NMR chemical shift behavior of hydrogen-bonded amide proton of glycine-containing peptides and polypeptides as studied by ab initio MO calculation. *Int. J. Mol. Sci.* **2002**, *3*, 907–913.
- (23) Tang, S.; Case, D. Vibrational averaging of chemical shift anisotropies in model peptides. *J. Biomol. NMR* **2007**, *38*, 255–266.
- (24) Tang, S.; Case, D. Calculation of chemical shift anisotropy in proteins. *J. Biomol. NMR* **2011**, *51*, 303–312.
- (25) Xu, X. P.; Case, D. A. Probing multiple effects on ^{15}N , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^{13}\text{C}'$ chemical shifts in peptides using density functional theory. *Biopolymers* **2002**, *65*, 408–423.
- (26) Manalo, M. N.; de Dios, A. C. CSGT-DFT calculation of ^{13}C and ^{15}N NMR shielding of the backbone amide group, C $^\alpha$, and C $^\beta$ in ω -conotoxin GVIA. *J. Mol. Struct. THEOCHEM* **2004**, *675*, 1–8.
- (27) Cai, L.; Fushman, D.; Kosov, D. Density functional calculations of chemical shielding of backbone ^{15}N in helical residues of protein G. *J. Biomol. NMR* **2009**, *45*, 245–253.
- (28) Cai, L.; Kosov, D.; Fushman, D. Density functional calculations of backbone ^{15}N shielding tensors in beta-sheet and turn residues of protein G. *J. Biomol. NMR* **2011**, *50*, 19–33.
- (29) Cai, L.; Fushman, D.; Kosov, D. Density functional calculations of ^{15}N chemical shifts in solvated dipeptides. *J. Biomol. NMR* **2008**, *41*, 77–88.
- (30) Mezey, P. G. Quantum chemistry of macromolecular shape. *Int. Rev. Phys. Chem.* **1997**, *16*, 361–388.
- (31) Mezey, P. G. Macromolecular density matrices and electron densities with adjustable nuclear geometries. *J. Math. Chem.* **1995**, *18*, 141–168.
- (32) Exner, T. E.; Mezey, P. G. Ab initio quality electrostatic potentials for proteins: An application of the ADMA approach. *J. Phys. Chem. A* **2002**, *106*, 11791–11800.
- (33) Exner, T. E.; Mezey, P. G. The field-adapted ADMA approach: Introducing point charges. *J. Phys. Chem. A* **2004**, *108*, 4301–4309.
- (34) Vila, J. A.; Villegas, M. E.; Baldoni, H. A.; Scheraga, H. A. Predicting C $^\alpha$ chemical shifts for validation of protein structures. *J. Biomol. NMR* **2007**, *38*, 221–235.
- (35) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (36) Mezey, P. G. Functional groups in quantum chemistry. *Adv. Quantum Chem.* **1996**, *27*, 163–222.
- (37) *Linear-Scaling Techniques in Computational Chemistry and Physics Methods and Applications*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011.
- (38) Nagata, T.; Fedorov, D. G.; Kitaura, K. Mathematical formulation of the fragment molecular orbital method. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011; pp 17–64.
- (39) Kobayashi, M.; Nakai, H. Divide-and-conquer approaches to quantum chemistry: Theory and implementation. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011; pp 97–127.

- (40) Rahalkar, A. P.; Yeole, S. D.; Ganesh, V.; Gadre, S. R. Molecular tailoring: An art of the possible for ab initio treatment of large molecules and molecular clusters. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011; pp 199–225.
- (41) Eckard, S. M.; Frank, A.; Onila, I.; Exner, T. E. Approximations of long-range interactions in fragment-based quantum chemical approaches. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011; pp 157–173.
- (42) Saebo, S. Linear scaling second-order Moller Plesset perturbation theory. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011; pp 65–82.
- (43) Korona, T.; Kats, D.; Schütz, M.; Adler, T. B.; Liu, Y.; Werner, H. J. Local approximations for an efficient and accurate treatment of electron correlation and electron excitations in molecules. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, Netherlands, 2011; pp 345–407.
- (44) Kaminsky, J.; Mata, R. A.; Werner, H. J.; Jensen, F. The accuracy of local MP2 methods for conformational energies. *Mol. Phys.* **2008**, *106*, 1899–1906.
- (45) Maschio, L. Local MP2 with density fitting for periodic systems: A parallel implementation. *J. Chem. Theory Comput.* **2011**, *7*, 2818–2830.
- (46) Doser, B.; Lambrecht, D. S.; Ochsenfeld, C. Tighter multipole-based integral estimates and parallel implementation of linear-scaling AO-MP2 theory. *Phys. Chem. Chem. Phys.* **2008**, *10*.
- (47) Schweizer, S.; Doser, B.; Ochsenfeld, C. An atomic orbital-based reformulation of energy gradients in second-order Moller-Plesset perturbation theory. *J. Chem. Phys.* **2008**, *128*, 154101–154108.
- (48) Sefzik, T. H.; Turco, D.; Iuliucci, R. J.; Facelli, J. C. Modeling NMR chemical shift: A survey of density functional theory approaches for calculating tensor properties. *J. Phys. Chem. A* **2005**, *109*, 1180–1187.
- (49) Kupka, T.; Stachow, M.; Nieradka, M.; Kaminsky, J.; Pluta, T. Convergence of nuclear magnetic shieldings in the Kohn–Sham limit for several small molecules. *J. Chem. Theory Comput.* **2010**, *6*, 1580–1589.
- (50) Auer, A.; Gauss, J.; Stanton, J. F. Quantitative prediction of gas-phase ^{13}C nuclear magnetic shielding constants. *J. Chem. Phys.* **2003**, *118*, 10407.
- (51) Prochnow, E.; Auer, A. A. Quantitative prediction of gas-phase ^{15}N and ^{31}P nuclear magnetic shielding constants. *J. Chem. Phys.* **2010**, *132*, 064109–7.
- (52) Moon, S.; Case, D. A. A comparison of quantum chemical models for calculating NMR shielding parameters in peptides: Mixed basis set and ONIOM methods combined with a complete basis set extrapolation. *J. Comput. Chem.* **2006**, *27*, 825–836.
- (53) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. A comparison of models for calculating nuclear magnetic resonance shielding tensors. *J. Chem. Phys.* **1996**, *104*, 5497–5509.
- (54) Woodford, J. N.; Harbison, G. S. Effects of zero-point and thermal vibrational averaging on computed NMR properties of a model compound for purine nucleosides. *J. Chem. Theory Comput.* **2006**, *2*, 1464–1475.
- (55) Eriksen, J. J.; Olsen, J. M.; Aidas, K.; Agren, H.; Mikkelsen, K. V.; Kongsted, J. Computational protocols for prediction of solute NMR relative chemical shifts. A case study of L-tryptophan in aqueous solution. *J. Comput. Chem.* **2011**, *32*, 2853–2864.
- (56) Dumez, J. N.; Pickard, C. J. Calculation of NMR chemical shifts in organic solids: Accounting for motional effects. *J. Chem. Phys.* **2009**, *130*, 104701–104708.
- (57) Dracinsky, M.; Kaminsky, J.; Bour, P. Structure of the alanine hydration shell as probed by NMR chemical shifts and indirect spin–spin coupling. *J. Phys. Chem. B* **2009**, *113*, 14698–14707.
- (58) van Mourik, T. Density functional theory reveals an increase in the amino ^1H chemical shift in guanine due to hydrogen bonding with water. *J. Chem. Phys.* **2006**, *125*, 191101–191104.
- (59) Exner, T. E.; Mezey, P. G. Ab initio quality properties for macromolecules using the ADMA approach. *J. Comput. Chem.* **2003**, *24*, 1980–1986.
- (60) Lorieau, J. L.; Louis, J. M.; Bax, A. The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 11341–11346.
- (61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Ciosowski, J.; Fox, D. J. *Gaussian 09, Revision B.01*; Gaussian, Inc.: Wallingford, CT, 2009.
- (62) Eckard, S.; Exner, T. E. Generalized hybrid orbitals in the FA-ADMA method. *Z. Phys. Chem.* **2006**, *220*, 927–944.
- (63) Eckard, S.; Exner, T. E. Improvements in the generalized hybrid orbital method. *Int. J. Quantum Chem.* **2009**, *109*, 1451–1463.
- (64) Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
- (65) Mennucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries. *J. Chem. Phys.* **1997**, *106*, 5151–5158.
- (66) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. Ab initio study of ionic solutions by a polarizable continuum dielectric model. *Chem. Phys. Lett.* **1998**, *286*, 253–260.
- (67) Exner, T. E.; Mezey, P. G. Evaluation of the field-adapted ADMA approach: Absolute and relative energies of crambin. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061–4069.
- (68) SYBYL-X 1.2. 2011; Tripos International, 1699 South Hanley Rd., St. Louis, MO 63144, USA.
- (69) Chesnut, D. B.; Rusiloski, B. E.; Moore, K. D.; Egolf, D. A. Use of locally dense basis sets for nuclear magnetic resonance shielding calculations. *J. Comput. Chem.* **1993**, *14*, 1364–1375.
- (70) Chesnut, D. B.; Byrd, E. F. C. The use of locally dense basis sets in correlated NMR chemical shielding calculations. *Chem. Phys.* **1996**, *213*, 153–158.
- (71) Provaci, P. F.; Aucar, G. A.; Sauer, S. P. A. The use of locally dense basis sets in the calculation of indirect nuclear spin–spin coupling constants. *J. Chem. Phys.* **2000**, *112*, 6201.
- (72) Cordier, F.; Nisius, L.; Dingley, A. J.; Grzesiek, S. Direct detection of N–H···O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nat. Protocols* **2008**, *3*, 235–241.
- (73) Zhang, H.; Neal, S.; Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **2003**, *25*, 173–195.
- (74) Sarotti, A. M.; Pellegrinet, S. C. A multi-standard approach for GIAO ^{13}C NMR calculations. *J. Org. Chem.* **2009**, *74*, 7254–7260.
- (75) Allen, M. J.; Keal, T. W.; Tozer, D. J. Improved NMR chemical shifts in density functional theory. *Chem. Phys. Lett.* **2003**, *380*, 70–77.
- (76) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.* **1998**, *109*, 6264–6271.

- (77) Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik, M. New generalized gradient approximation functionals. *J. Chem. Phys.* **2000**, *112*, 1670–1678.
- (78) Boese, A. D.; Handy, N. C. A new parametrization of exchange–correlation generalized gradient approximation functionals. *J. Chem. Phys.* **2001**, *114*, 5497–5503.
- (79) Van Voorhis, T.; Scuseria, G. E. A novel form for the exchange–correlation energy functional. *J. Chem. Phys.* **1998**, *109*, 400–410.
- (80) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998**, *108*, 664–675.
- (81) Enthart, A.; Furrer, J.; Dehner, A.; Kessler, H. Solution structure and binding studies of the p63 DNA binding domain. Manuscript in preparation.
- (82) Di Lello, P.; Jenkins, L. M.; Jones, T. N.; Nguyen, B. D.; Hara, T.; Yamaguchi, H.; Dikeakos, J. D.; Appella, E.; Lagault, P.; Omichinski, J. G. Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol. Cell* **2006**, *22*, 731–740.
- (83) Keal, T. W.; Tozer, D. J. A semiempirical generalized gradient approximation exchange–correlation functional. *J. Chem. Phys.* **2004**, *121*, 5654–5660.
- (84) Schutz, M.; Werner, H. J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.
- (85) Jensen, F. Basis set convergence of nuclear magnetic shielding constants calculated by density functional methods. *J. Chem. Theory Comput.* **2008**, *4*, 719–727.
- (86) Jensen, F. The basis set convergence of spin–spin coupling constants calculated by density functional methods. *J. Chem. Theory Comput.* **2006**, *2*, 1360–1369.
- (87) Teale, A. M.; Tozer, D. J. Exchange representations in Kohn–Sham NMR shielding calculations. *Chem. Phys. Lett.* **2004**, *383*, 109–114.