PCCP

Determination of Accurate Backbone Chemical Shift Tensors in Microcrystalline Proteins by Integrating MAS NMR and QM/MM

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-01-2018-000647.R1
Article Type:	Paper
Date Submitted by the Author:	05-Mar-2018

Complete List of Authors:	Fritz, Matthew; University of Delaware, Chemistry and Biochemistry Quinn, Caitlin; University of Delaware, Chemistry and Biochemistry Wang, Mingzhang; University of Delaware, Chemistry and Biochemistry Hou, Guangjin; University of Delaware, Chemistry and Biochemistry Lu, Xingyu; University of Delaware, Chemistry & Biochemistry Koharudin, Leonardus; University of Pittsburgh, Structural Biology Struppe, Jochem; Bruker, Case, David; Rutgers University New Brunswick, Chemistry and Chemical Biology Polenova, Tatyana; University of Delaware, Chemistry and Biochemistry Gronenborn, Angela; University of Pittsburgh, Structural Biology

Determination of Accurate Backbone Chemical Shift Tensors in Microcrystalline Proteins by Integrating MAS NMR and QM/MM

Matthew Fritz[1,2], Caitlin M.Quinn[1,2], Mingzhang Wang[1,2,], Guangjin Hou[1], Xingyu Lu[1,2], Leonardus M. I. Koharudin[2,3], Jochem Struppe[4], David A. Case[5], Tatyana Polenova[1,2*], and Angela M. Gronenborn[2,3*]

[1]Department of Chemistry and Biochemistry, University of Delaware, Newark, DE 19716, United States; [2]Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, 1051 Biomedical Science Tower 3, 3501 Fifth Ave., Pittsburgh, PA 15261, United States; [3]Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15261, United States; [4]Bruker Biospin Corporation, 15 Fortune Drive, Billerica, MA, United States; [5]Department of Chemistry and Chemical Biology, Rutgers University, 174 Frelinghuysen Road, Piscataway, NJ 08854-8087, United States

***Corresponding authors:** Angela M. Gronenborn, Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15260, USA, Tel.: (412) 648-9959; Email: amg100@pitt.edu; Tatyana Polenova, Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA, Tel.: (302) 831-1968; Email: tpolenov@udel.edu

ABSTRACT

Chemical shifts are highly sensitive probes of local conformation and overall structure. Both isotropic shifts and chemical shift tensors are readily accessible from NMR experiments but their quantum mechanical calculations remain challenging. In this work, we report and compare accurately measured and calculated $^{15}N^{H}$ and $^{13}C^{\alpha}$ chemical shift tensors in proteins, using the microcrystalline agglutinin from *Oscillatoria agardhii* (OAA). Experimental $^{13}C^{\alpha}$ and $^{15}N^{H}$ chemical tensors were obtained by solid-state NMR spectroscopy, employing tailored recoupling sequences, and for their quantum mechanics/molecular mechanics (QM/MM) calculations different sets of functionals were evaluated. We show that $^{13}C^{\alpha}$ chemical shift tensors are primarily determined by backbone dihedral angles and dynamics, while $^{15}N^{H}$ tensors mainly depend on local electrostatic contributions from solvation and hydrogen bonding. In addition, the influence of including crystallographic waters, the molecular mechanics geometry optimization protocol, and the level of theory on the accuracy of the calculated chemical shift tensors is discussed. Specifically, the power of QM/MM calculations in accurately predicting the unusually upfield shifted $^{1}H^{N}$ G26 and G93 resonances is highlighted. Our integrated approach is expected to benefit structure refinement of proteins and protein assemblies.

INTRODUCTION

Protein structure determination by NMR spectroscopy relies on distance and angular restraints, supplemented by isotropic chemical shifts (CS) for ^{13}C, ^{15}N, and ^{1}H nuclei. NMR chemical shifts are exquisitely sensitive reporters of local geometry and electronic environments and are readily accessible, both in solution and solid-state NMR experiments.[1] In contrast, the orientation-dependent chemical shift anisotropy (CSA) tensors are not directly available in solution because of motional averaging and, until recently, their site-specific measurements in solid-state NMR experiments of proteins were challenging. With the advent of modern recoupling sequences,[2, 3] incorporated into multidimensional magic angle spinning (MAS) solid-state experiments, CSA tensors for heteronuclei (^{13}C and ^{15}N) as well as ^{1}H can now be easily obtained for proteins and protein assemblies.[4, 5]

Experimental ^{1}H, ^{13}C, and ^{15}N CSA tensors provide critical information about protein secondary structure, electrostatic and hydrophobic interactions, protonation states, rotameric states, as well as dynamics,[6-8] and the incorporation of ^{13}C and ^{15}N CSA tensors holds great promise for protein structure refinement,[9, 10] potentially aiding the laborious task of internuclear distance restraint assignment. At present, however, such approaches are not yet widely used, due to the absence of robust computational protocols, particularly for ^{15}N tensors. Most of the current methodologies rely on Density Functional Theory (DFT) calculations, either in the cluster or hybrid quantum mechanics/molecular mechanics (QM/MM) format. In the first approach, typically a 3-6 Å sphere around the residue or region of interest is constructed and only this cluster is subjected to DFT-based geometry optimization followed by NMR parameter calculations.[11, 12] In hybrid QM/MM calculations, the system is partitioned into a 3.5 Å sphere for full atomistic treatment and the remainder of the protein is treated as point charges.[13, 14] The QM/MM treatment of the entire protein results in the calculations not being prohibitively expensive, with the MM minimization typically sufficient for suitably relaxing the protein. QM/MM calculations of NMR isotropic chemical shifts have been benchmarked previously,[15, 16] and hybrid functionals and diffuse basis sets with polarization were shown to result in reasonably accurate agreement with experimental values.[17, 18] Furthermore, when combined with molecular dynamics (MD) simulations, the integrated MD/QM/MM calculations yield satisfactorily accurate CSA tensor predictions for proteins, which exhibit extensive local dynamics, as was shown for HIV-1 CA capsid protein assemblies.[5] However, in general, QM/MM calculations of protein CSA tensors remain far from being routine and accurate predictions of ^{15}N tensors are very challenging.

In order to establish reliable and streamlined QM/MM computational protocols for accurate chemical shift predictions of proteins, it is imperative to examine a sufficiently large number of model benchmark systems. These should encompass representative secondary structure types and tertiary structural topologies, and should comprise proteins for which atomic-resolution structures are available, isotropic chemical shifts and CSA tensors have been measured, and which can be crystallized in the same form for both diffraction and MAS NMR studies. Here we use the 133-residue comprising agglutinin from *Oscillatoria agardhii* (OAA) as the model protein. It exhibits an unusual beta-barrel topology (Figure 1A), created from two sequence repeats in a pseudo two-fold symmetrical arrangement.[19-21] We have solved the X-ray structure of OAA at 1.2 Å resolution (PDB: 3OBL),[21] and extensive NMR characterizations were carried out that provided complete solution chemical shift assignments[22] and, more recently, MAS NMR parameters.[18] Isotropic backbone $^{13}C^{\alpha}$ and $^{5}N^{H}$ chemical shifts were calculated by the QM/MM approach and the influence of hydrogen bonding, crystal contacts, and dynamics on the accuracy of shift predictions were reported.[18]

In this work, we assessed the factors determining the accuracy of backbone $^{13}C^{\alpha}$ and $^{15}N^{H}$ CSA tensor QM/MM calculations. Our results reveal that very good agreement between experiment and theory can be reached for $^{13}C^{\alpha}$ tensors. We also show that $^{13}C^{\alpha}$ tensors are predominantly determined by backbone dihedral angles and dynamics, while the $^{15}N^{H}$ tensors are mainly dependent on local electrostatic contributions from solvation and hydrogen bonding. In addition, we evaluated the influence of the MM minimization scheme, the inclusion of crystallographic waters, the N-H bond length, and the level of theory (functional). The power of the current QM/MM calculations in accurately predicting the unusually high field $^{1}H^{N}$ chemical shifts of G26 and G93 residues, originating from ring current effects of the close-by W90 and W23 side chains is clearly evident. Overall, the integrated approach developed here will be of use for protein structure refinement methodologies that are based on chemical shift anisotropy tensors.

EXPERIMENTAL

Protein expression, purification, and crystallization of OAA were performed as described previously.[21] For all MAS NMR experiments, 30 mg of crystals were packed into 3.2 mm thin-wall Bruker rotors.

MAS NMR experiments were carried out on a 14.1 T narrow bore Bruker AVIII spectrometer outfitted with a 3.2 mm HCN EFree MAS probe. Larmor frequencies were 599.8 MHz (^{1}H), 150.8 MHz (^{13}C), and 60.8 MHz (^{15}N). Several 2D data sets were collected on a 19.96 T narrow bore Bruker AVIII spectrometer using a 3.2 mm HCN EFree MAS probe; Larmor frequencies were 850.4 MHz (^{1}H), 213.9 MHz (^{13}C), and 86.2 MHz (^{15}N). The MAS frequency was set at 10 or 14 kHz for all experiments, and was controlled to within ± 5 Hz by a Bruker MAS III controller. KBr was used as temperature sensor, and the actual temperature of the sample was maintained to within 4 ± 0.1 ^{o}C using the Bruker BCU temperature controller. ^{13}C and ^{15}N chemical shifts were referenced with respect to the external standards adamantane and NH_4Cl, respectively. Typical 90^{o} pulse lengths were 2.8 µs (^{1}H), 4.0 µs (^{13}C), and 4.8 µs (^{15}N), and the contact time of ^{1}H-^{15}N/^{13}C cross polarization (CP) was 2.2/1.8 ms. ^{1}H-^{15}N/^{13}C CP employed a 95-105% linear amplitude ramp on the ^{1}H channel with the center Hartmann-Hahn matched to the first spinning side band. The band-selective magnetization transfer from ^{15}N to $^{13}C^{\alpha}$ was performed through a 4.5 ms SPECIFIC-CP with a tangent amplitude ramp on the ^{15}N channel (49 kHz rf field center) and a constant rf field on the ^{13}C channel (35 kHz). High-power ^{1}H continuous wave (CW) decoupling (89 kHz) was applied during the SPECIFIC-CP period, and SPINAL-64 decoupling (89 kHz) was applied during the direct (t_3) and indirect (t_2) acquisition periods. In ^{13}C and ^{15}N 3D RNCSA experiments,[23] $R10_1^3$ (^{13}C) and $R14_2^5$ (^{15}N)-based symmetry sequences were used to reintroduce the ^{13}C and ^{15}N CSA during the t_1 evolution period, and the phase-alternated rf field irradiation (43 and 50 kHz) was applied on the ^{15}N/^{13}C channel respectively. Simultaneous π pulses were applied on the ^{13}C channel at the center of every two rotor periods to decouple ^{15}N-^{13}C dipolar interactions. ^{13}C and ^{15}N CSA tensors were also recorded using the ROCSA sequence,[24] to determine the absolute sign of the reduced anisotropy and the asymmetry parameter. The MAS frequency was 10 kHz, and the ROCSA rf power was optimized so as to minimize the lineshape distortions.[24]

Processing of NMR data was carried out in NMRpipe[25]; the spectra were analyzed with both Sparky and CCPN.[26, 27] In all 2D and 3D datasets, 30^{o} or 60^{o} shifted sine bell apodization was followed by Lorentz-to-Gaussian transformation. The RNCSA data sets were evaluated as the real-FT of the corresponding indirect dimension zero-filled to 256 points prior to FT.

Numerical simulations of $^{15}N/^{13}C$ CSA lineshapes were simulated using the Minuit package in SIMPSON[28] versions 1.1.2. To produce a powder average, 320 pairs of {α, β} angles were generated according to the REPULSION algorithm, and 16 γ angles (resulting in a total of 5,120 angle triplets) were used for all simulations. NMR parameters in the experiment matched those used during the fitting routine.

In defining the CSA tensor, we used a modified Haeberlen convention below.

$$\delta_{iso} = (\delta_{11} + \delta_{22} + \delta_{33})/3 \quad (1)$$

$$\delta_{\sigma} = \delta_{11} - \delta_{iso} \quad (2)$$

$$\eta = (\delta_{33} - \delta_{22})/(\delta_{11} - \delta_{iso}) \quad (3)$$

$$|\delta_{11} - \delta_{iso}| \geq |\delta_{22} - \delta_{iso}| \geq |\delta_{33} - \delta_{iso}| \quad (4)$$

QM/MM calculations of ^{13}C and ^{15}N chemical shift tensors were carried out in Gaussian09,[29] at the OLYP/TZVP level for the quantum mechanical region, using the scripts generated in AFNMR, and using chain A of PDB ID: 3OBL as initial input. In the QM/MM calculations, the protons are added using the Amber force field libraries followed by the MM minimization of the entire structure. The structure was minimized using the Amber FF99SB molecular mechanics force field and referenced to ubiquitin (PDB ID: 1D3Z) calculated at the same level of theory (^{1}H =32.0 ppm, ^{13}C=182.5, and ^{15}N= 237.8 ppm.). Proton only optimizations were achieved using the MDWeb server (http://mmb.irbbarcelona.org/MDWeb/). All functionals and basis sets were used as implemented in Gaussian 09.

RESULTS AND DISCUSSION

MAS NMR spectra and chemical shift assignments of OAA crystals

2D ^{13}C-detected CORD and NCA spectra of OAA were acquired at 19.96 T (Figure 1B) and backbone connectivities for the stretch of residues from T50 to I59, extracted from 3D NCACX, NCOCX, and CONCA spectra, are shown in Figure 1C. We previously assigned 107 residues on the basis of 2D and 3D ^{13}C-detected spectra acquired at 14.1 T.[18] Here, given the outstanding spectral resolution of the CORD spectra at 19.96 T, additional 15 resonances were resolved. However, even with the significantly improved resolution, there are still a number of residues whose ^{13}C and ^{15}N chemical shifts could not be resolved, given the extensive amino acid similarity of the two sequence repeats. For these residues, solution ^{1}H chemical shifts are distinct, and therefore a ^{1}H-detected (H)NH HETCOR spectrum at 19.96 T and MAS frequency of 60 kHz was acquired. As illustrated in Figure 2A and 2B, the spectral resolution is remarkably high, and this data set in combination with the CORD spectrum permitted the tentative assignment of 16 pairs of residues from the two repeats.

The unusual high field ^{1}H shifts of residues G26 and G93 in the HETCOR spectrum (3.5 and 2.9 ppm, Figure 2B) are in accord with the solution chemical shifts reported previously,[20-22] and caused by ring current effects of the W90 and W23 sidechains (Figure 2C). Remarkably, the ^{1}H amide shifts for these two residues calculated by QM/MM agree well with the experimental values (Table 1), and analysis reveals the contribution of the W90/W23 HOMO orbitals to the ^{1}H magnetic shielding tensors of G26/G93.

Table1. $^{1}H^{N}$ Chemical Shifts for G26 and G93

	MAS NMR	Solution NMR	QM/MM	SHIFTX2	SPARTA+
G26	3.5	3.38	2.93	5.15	8.99
G93	2.9	2.94	2.55	5.15	8.89

Comparison between experimental MAS NMR and QM/MM calculated $^{13}C^{\alpha}$ and $^{15}N^{H}$ chemical shift anisotropy tensors

The experimental $^{13}C^{\alpha}$ and $^{15}N^{H}$ CSA tensors (principal components) were obtained from 3D RNCSA spectra.[4] $^{13}C^{\alpha}$ and $^{15}N^{H}$ RNCSA lineshapes for representative residues are shown in Figure 3A and 4A. Experimental CSA parameters were extracted from fits to the experimental lineshapes and calculated at the QM/MM level in Gausssian09, using the X-ray structure of OAA (PDB ID 3OBL). Since RNCSA lineshapes are not sensitive to the sign of the anisotropy and the asymmetry parameter (see Figure S1 of the Supplementary Information), $^{13}C^{\alpha}$ and $^{15}N^{H}$

CSA tensors were also extracted from ROCSA[24] lineshapes. These are shown for representative residues in Figure S2 of the Supplementary Information. We note that the absolute magnitude of the reduced anisotropy parameters cannot be reliably determined from the ROCSA experiment because of the contribution of ^{13}C-^{13}C homonuclear dipolar coupling to the ROCSA Hamiltonian in uniformly ^{13}C labeled samples, as reported previously[24, 30] and shown in Figure S3 of the Supplementary Information. We also note that the absolute orientations of the CSA tensors in the molecular frame could only be accessible from single-crystal NMR experiments. These are only practical in small molecules, such as amino acids[31], and are not envisioned to be practical in protein studies, because of sensitivity issues, signal overlap, and general inaccessibility of large enough crystals for such experiments.

The experimental and calculated $^{13}C^{\alpha}$ and $^{15}N^{H}$ reduced anisotropy parameters, δ_{σ}, plotted vs. residue number, as well as the correlation between the experimental and calculated principal components of the $^{13}C^{\alpha}$ and $^{15}N^{H}$ CSA tensors, δ_{ii}, are shown in Figure 3B and 4B. The complete listing of experimental and calculated CSA parameters is provided in Table 1S (Supplementary Information).

As can be appreciated from Figures 3 and 4 and Table 1S, the experimental CSA parameters for both $^{13}C^{\alpha}$ and $^{15}N^{H}$ are generally in good agreement with the predicted values, with R^2 values of 0.93 and 0.95, respectively. For the $^{13}C^{\alpha}$, values, most outliers are associated with residues in the loop regions or prolines and glycines (Figure 3C), consistent with previous observations for isotropic shifts.[18] Indeed, deviations between experimental and calculated CSA values for residues in the loop regions are not surprising, since loops exhibit motions which are not well represented when calculations are performed based on a static X-ray structure. To account for the dynamic averaging of the CSA tensors, MD-DFT calculations are needed. Such calculations are time consuming and will be performed in the future. It is also important to note that for a number of residues (magenta bars in Figure 3B), the sign of the reduced anisotropy parameter is undetermined because the asymmetry parameters are close to 1. For several other residues (blue bars in Figure 3B), the sign of the reduced anisotropy extracted from the ROCSA lineshapes, does not agree with the QM/MM calculation. The likely reason for this discrepancy is a possible contribution of homonuclear ^{13}C-^{13}C dipolar coupling, (Figure S3 of the Supplementary Information), to the ROCSA lineshapes.

Inspection of the correlation between the calculated and experimental δ_{ii} of the $^{15}N^{H}$ CSA tensor parameters revealed a systematic offset of -9.07 ppm, which required investigation of the various factors that could affect the accuracy of the computed $^{15}N^{H}$ CSA tensors. As has been noted before, $^{15}N^{H}$ CSA tensor calculations are notoriously difficult,[32-34] and $^{15}N^{H}$ shifts are very

sensitive to the local electronic environment and motions. Below we describe our findings with respect to the dependence of the CSA parameters on the level of theory (functional and basis set), the geometry optimization protocol and inclusion of crystallographic waters in the structure, N-H bond length and hydrogen bonding, and the presence of conformational heterogeneity/dynamics.

Choice of density functional

Six commonly used functionals were evaluated with respect to the accuracy of the calculated $^{15}N^{H}$ CSA tensors: three GGA functionals (OLYP, BLYP, and OPBE), two hybrid functionals (O3LYP and B3LYP), and a meta hybrid GGA functional (M06). For each of the functionals, the calculated principal components of the CSA tensor, δ_{ii}, are displayed vs. the corresponding experimental values (Figure 5) for a test set of representative residues that exhibited varying differences between computed and experimental shifts. Surprisingly, the choice of density functional when paired with the commonly used TZVP basis set had little to no impact on the results, as evidenced by the similar slopes, intercepts, and correlation coefficients. Judging by the RMSD values, OLYP, O3LYP, and B3LYP perform the best, most likely since the TZVP basis set was optimized for the use with LDA exchange functionals.[35] Indeed, other studies showed that quantitative agreement can be reached with hybrid functionals and the TZVP basis set.[5, 16] Therefore, the extra exchange term seems not necessary for accurate $^{15}N^{H}$ tensor predictions.

We also examined whether using meta-GGA functionals with correlation-consistent basis sets improved the predicted $^{15}N^{H}$ CSA tensors. These calculations were carried out using cluster models with and without DFT geometry optimization. This approach was prompted by a recent report on the advantages of using meta-GGA functionals for ^{15}N CSA tensor calculations of small molecules.[36] In contrast to the findings for small molecules, we did not observe that the TPSS functional outperforms OLYP. Interestingly, our results suggest that the M06L functional with the cc-pVTZ basis set yields the highest accuracy for calculated $^{15}N^{H}$ CSA tensors of β-sheet residues. For example, calculations with most functionals and basis sets result in a 5 ppm difference to the experimental values for V6, which is reduced to 90.3 ppm when the M06L/cc-pVTZ combination is used. Similar results were seen for A63 yielding a difference of 5 ppm between the computed and experimental reduced anisotropy parameter. We note that this level of agreement is higher than that reported previously for GB1 and ubiquitin, used as benchmark proteins to test the QM/MM approach for computing chemical shift tensors.[15] For loop residues, the reduced anisotropies are over-estimated, most likely due to the averaging of dynamics on

the nano-microsecond timescale, which would compromise results based on DFT calculations of static structures. Since meta-GGA functionals contain an additional term describing the kinetic energy of the electrons, such shortcomings should be less detrimental. This can be illustrated by considering the geometry of the frontier molecular orbitals. In Figure 5, N-Acetyl valine is shown as an example: the negative electron density of the nitrogen lone pair occupies significant space on the opposite side of the peptide plane and the kinetic term will account for any motion of the lone pair motion.

Optimization protocols and crystallographic waters

The accuracy of predicting NMR parameters by DFT is critically dependent on the atomic coordinates. Even when high-resolution X-ray structures are available for cluster calculations of proteins, DFT geometry optimization is necessary.[17] This is often challenging due to convergence problems, and it is difficult to verify whether a true minimum has been reached, rather than a saddle point on the potential energy surface. For calculating magnetic shielding anisotropy tensors, prior work by us and by others suggests that MM minimization is sufficient.[5, 18] However, several questions still remain, such as: is it necessary to optimize the heavy atom positions? Should one include crystallographic waters? What is the best bond length to use for adding H atoms to the coordinates?

In order to evaluate how different optimization protocols affect the values of the computed CSA tensors, we examined the following four scenarios: (i) optimization of all atoms, (ii) optimization of all atoms, including crystallographic waters, (iii) optimization of H atoms of the protein only, and (iv) optimization of all H atoms of the protein and the crystallographic waters. In (iii) and (iv), the coordinates of the heavy atoms were kept fixed in the X-ray geometry. As illustrated in Figure 6, the overall accuracy of the $^{15}N^{H}$ CSA tensor principal components is largely insensitive to the choice of the MM optimization scheme. In particular, the inclusion of crystallographic waters does not improve the agreement between experimental and calculated values. For example, considering amides that are involved in hydrogen bonds with crystallographic waters, such as L3 and V33, we note that for L3, which is surface exposed and serves as a hydrogen bond donor to a water, worse predictions are obtained. We speculate that this is due to the fact that water molecules on the protein surface are highly dynamic, and that in a particular crystal structure a snapshot of water locations is observed, without knowledge whether this position is always occupied. For V33, the inclusion of the crystallographic waters resulted in improved CSA predictions. V33 is a loop residue, and the amide forms a bridging hydrogen bond with the sidechain hydroxyl of T4, mediated by a water. In this case it may well

be possible that this water molecule is more tightly bound and represents an integral part of the protein structure, compared to the more mobile surface waters. Indeed, the improvement after including this explicit water proton geometry optimization is significant (Figure S4, Supplementary Information). A similar result is observed for N37, whose side chain is hydrated. Fixing the heavy atoms and including explicit waters greatly improves the accuracy of the calculated tensor components. Finally, the water molecules that are outside of the 3.5 Å quantum region, but hydrogen bonded to amides within the QM sphere, are embedded as point charges in the MM region, potentially resulting in increased error.

Overall, our data show that the choice of minimization protocol and inclusion of waters has little to no effect for $^{13}C^{\alpha}$ tensors, while occasionally it can be beneficial for $^{15}N^{H}$ CSA tensors when bridging waters are present. Using the OLYP/TZVP level of theory for the quantum region, it is possible to obtain excellent quantitative agreement for carbons. For nitrogens, on the other hand, the degree of scatter is greater, but can be reduced to some extent by using the more expensive M06L/cc-pVTZ functional/basis set combination.

Conformational heterogeneity

In order to evaluate the influence of conformational heterogeneity, we averaged the results of QM/MM calculations for a rigid (A63) and a dynamic (D80) residue over the 20 lowest-energy solution NMR conformations of OAA (PDB ID 2MWH). The resulting $^{15}N^{H}$ chemical shift anisotropies do not vary significantly within the 20-conformer ensemble, yielding 0.85 ppm for A63 and 2.50 ppm for D80 (Figure S5, Supplementary Information). The larger deviation for D80 vs. A63 is not unexpected: A63 is a rigid residue with an N-H order parameter of 0.92, while D80 is more mobile, with an N-H order parameter of 0.80. Interestingly, the QM/MM calculations over-estimate the magnitude of the reduced anisotropy by 17 ppm for D80 and 8 ppm for A63. Furthermore, for D80 the measured $^{15}N^{H}$ CSA is approximately 0.80 of the QM/MM value, underscoring the fact that it undergoes dynamics on the nano- to microsecond timescale. To account for dynamic averaging of the CSA tensor, an integrated MD-DFT approach will need to be pursued in the future, as noted for HIV-1 CA protein assemblies.[5]

N-H bond lengths and hydrogen bonding

Although MD simulations improve the accuracy of CSA tensor calculations for dynamically averaged residues,[5] nano- to microsecond motions cannot account for the discrepancies between calculated and experimental values for residues, which are static on these timescales. A63 is one such residue, yet the difference between calculated and experimental $^{15}N^{H}$ δ_{σ} is 8

ppm. For other residues, such as V6, I36, I103, F61, F128, and K129 similar differences are observed, namely a 5-8 ppm over-estimation of the reduced anisotropy. All of these residues are located in β-strands. While a 5-8 ppm differences are only of the order of 5-8% of the overall magnitude of δ_σ, for many other amides the agreement between calculation and experiment is much better. For example, residues A15 and L92, which are located in loops, exhibit differences smaller than 2 ppm (this is within the experimental error) between the experimental and computed shifts. Since within the β-sheet structure of OAA extensive hydrogen bonding is present and contributions to the chemical shift tensor have an inverse cubic relationship with distance,[37-40] hydrogen bond donating amides in β-sheets will experience a larger contribution than loop residues. Therefore, if motions were negligible, contributions from hydrogen bonding interactions should dominate the chemical shift tensor. This will be most pronounced for the δ_{11} component, which lies along the N-H bond vector, and it has been previously reported that this component is extremely sensitive to hydrogen bonding interactions.[7, 41]

In addition, even high-resolution X-ray structures of proteins will not contain H atoms and protons are added prior to the geometry minimization step. We therefore examined the effect of systematically varying the N-H bond length on the CSA parameters. As above, we selected two representative residues, A63 and D80, which exhibit distinctive chemical shifts. For A63, increasing the N-H bond length from 1.005 to 1.040 Å results in a linear increase of δ_{iso} from 123.5 to 127.4 ppm and a linear decrease in δ_σ from 103.6 to 100.6 ppm, thus a 1 ppm change for 0.013 Å (Figure 7). The corresponding data for D80 are shown in Figure S6 (Supplementary Information). These changes are too small to account for the observed differences between the experimental and calculated shifts for these two residues. The linear dependence with increasing bond length is consistent with reports for organic molecules.[42]

Table 2. $^{15}N^H$ Experimental and Computed Chemical Shift Parameters, Hydrogen Bond Lengths, and Dipolar Order Parameters for V6, A63, and K129

Residue	**Experimental**	**N-H···O/ Å**	**S_{N-H}**	**OLYP/TZVP**	**B3LYP/6-311++G(2d,2p)**
V6	90.10	1.95	0.88	96.51	90.25
A63	91.32	2.01	0.90	98.20	85.00
K129	95.92	2.01	0.88	102.99	106.95

If the presence of hydrogen bonding were the main cause for the difference in experimental and calculated shifts, then diffuse basis sets should be the most prudent choice. We calculated the $^{15}N^H$ CSA reduced anisotropy parameters at the b3lyp/6-31++G** level for the

three β-sheet residues, V6, A63, and K129, which exhibited > 8 ppm differences between experiment and calculation with the GGA functionals (Table 2). A double zeta basis set with diffuse functionals previously proved effective in determining the hydrogen bond lengths in spider silk[43]. However, disappointingly using the larger basis sets and a hybrid functional, no improvements over OLYP/TZVP is observed, although for both sets the qualitative trends in the anisotropies remain. K129 is always the largest and V6 the smallest.

CONCLUSIONS AND OUTLOOK

Using OAA as a model benchmark protein, we investigated the accuracy of CSA tensor calculations at the QM/MM level. Our study indicates that for $^{13}C^{\alpha}$ and $^{15}N^{H}$ tensors fairly accurate predictions can be made, although for $^{15}N^{H}$ calculations further improvements are necessary to reduce the number of outliers. Nevertheless, to our knowledge, this is the best reported agreement for $^{15}N^{H}$ chemical shift tensors in proteins so far.

Dynamics, hydrogen bonding, and other local interactions modulate the chemical shifts and their effects need to be carefully considered. The current accuracy of calculated CSA tensors appears to be invariant to the MM optimization protocol and the choice of density functional. The effect of including crystallographic waters appears random, without any general trends. Importantly, satisfactory agreement between experiment and calculation can be obtained efficiently with modest basis sets and functionals, although more accurate predictions can be achieved using meta-GGA functionals and contracted basis sets. It is anticipated that reliable incorporation of chemical shift tensors into structure/dynamics characterization protocols will further improve structure characterization of proteins and large macromolecular assemblies by MAS NMR.

CONFLICTS OF INTEREST

There are no conflicts to declare.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health National Institutes of Health (NIH Grants P50 GM082251, Technology Development Project on MAS NMR, and P50 GM103297). We acknowledge the support of the National Science Foundation (NSF Grant CHE0959496) for the acquisition of the 850 MHz NMR spectrometer and of the National Institutes of Health (NIH Grants P30GM103519 and P30GM110758) for the support of core instrumentation infrastructure at the University of Delaware; and of the National Institutes of Health (NIH Grant 1S10OD012213) for the acquisition of the 750 MHz NMR spectrometer at the University of Pittsburgh. CMQ acknowledges the support of the National Institutes of Health Postdoctoral Fellowship grant F32GM113452.

FIGURE 1

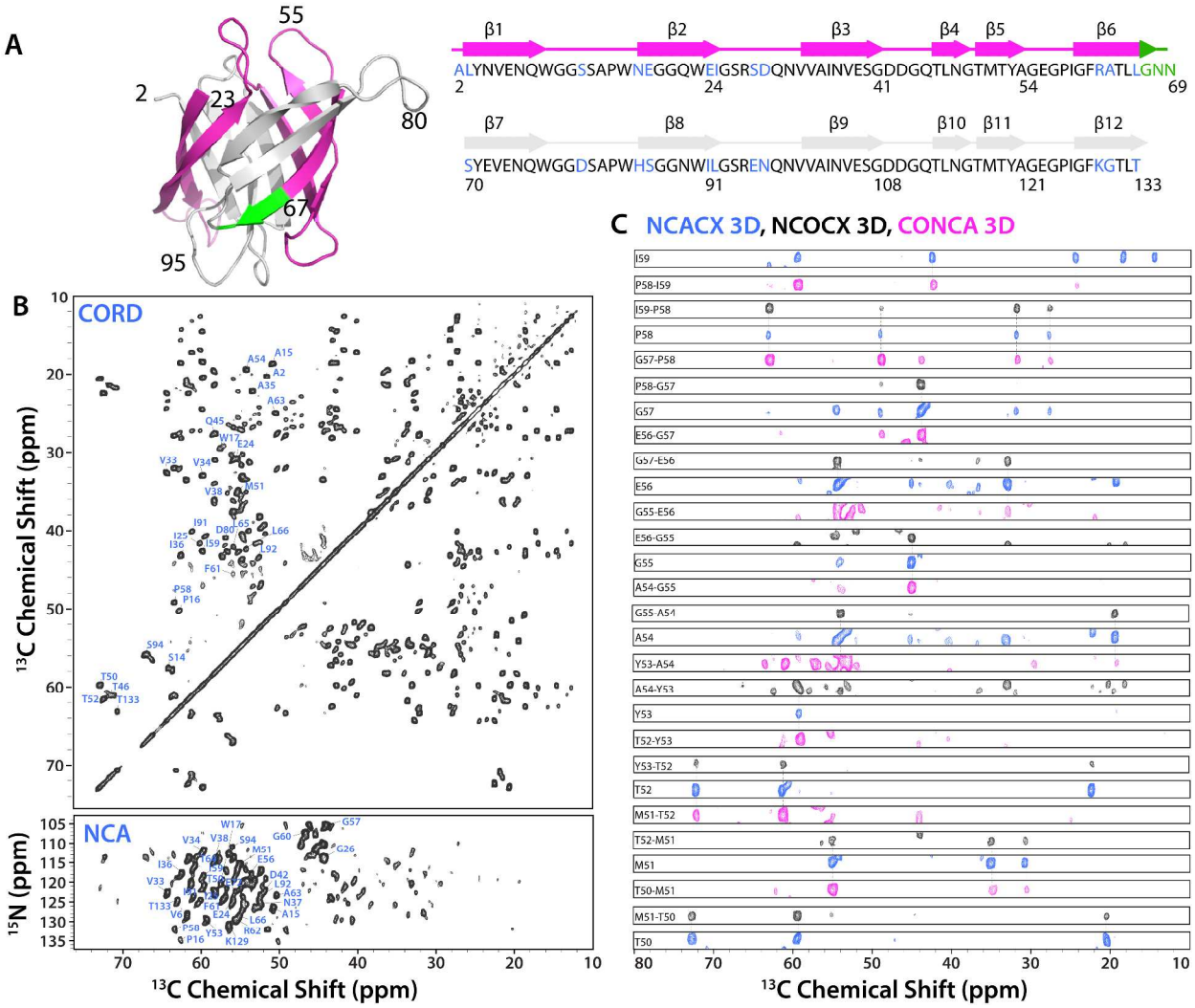


Figure 1. A: Ribbon representation of the X-ray structure (left) and amino acid sequence (right) of *O. agardhii* agglutinin. Identical residues in the sequence repeats are shown in black, linker residues between the repeats are shown in green, β-strands are indicated by arrows above the sequence, and colored magenta and gray for repeat 1 and 2, respectively, both in the structure and sequence. **B:** 2D MAS NMR spectra of OAA microcrystals: CORD (top), NCA (bottom). Assignments for selected resonances are shown and labeled with residue name and number. **C:** Backbone walk for a stretch of residues T50-I59 from 3D heteronuclear NCACX, NCOCX, and CONCA spectra of OAA microcrystals. All 2D and 3D NMR spectra were acquired at 19.96 T and 14.0 T, respectively; the MAS frequency was 14 kHz for all experiments.

FIGURE 2

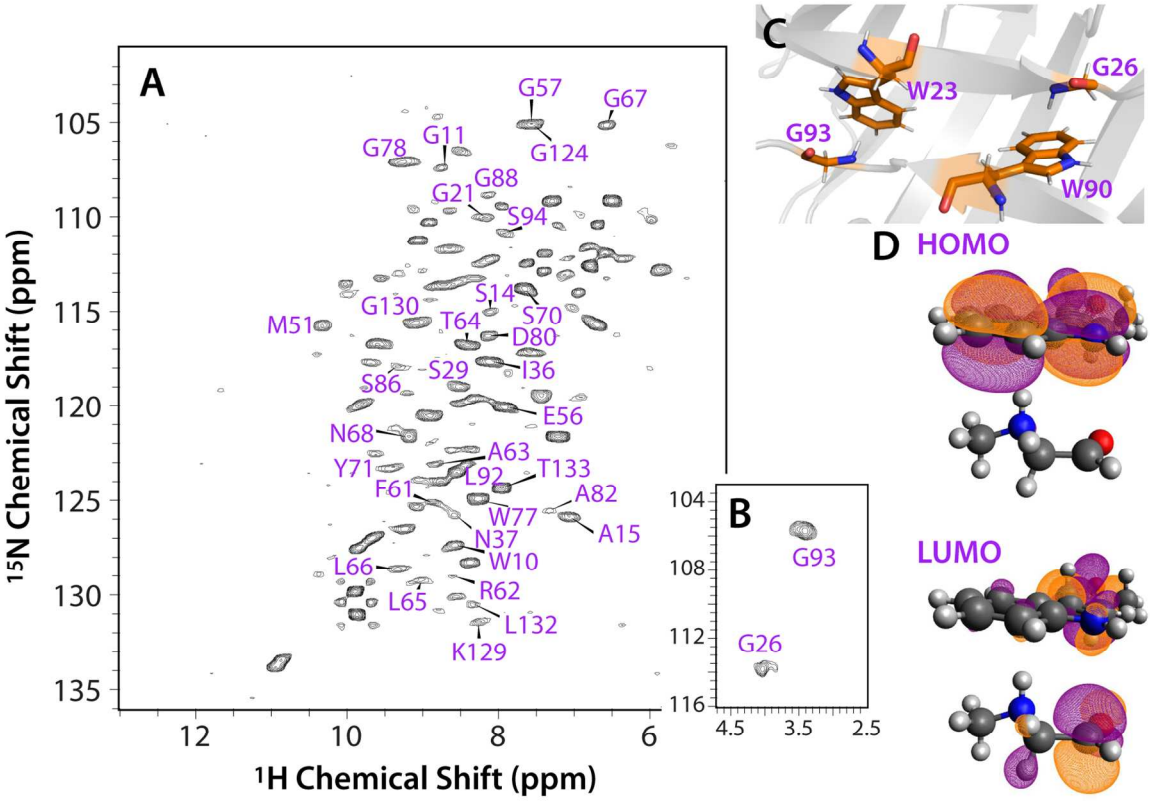


Figure 2. A, B: ^{15}N-^{1}H (H)NH HETCOR spectrum of OAA acquired at 14.0 T and a MAS frequency of 60 kHz, with the region around the G26 and G93 resonances at the bottom right. **C:** Stacking interaction between G26/G93 and the proximal W90/W20 residues in the structure, respectively. This interaction gives rise to strong shielding of the G26/G93 amide protons. **D:** The symmetry of HOMO and LUMO orbitals of G93, illustrating the influence of the W23 orbitals on the magnetic shielding tensor.

FIGURE 3

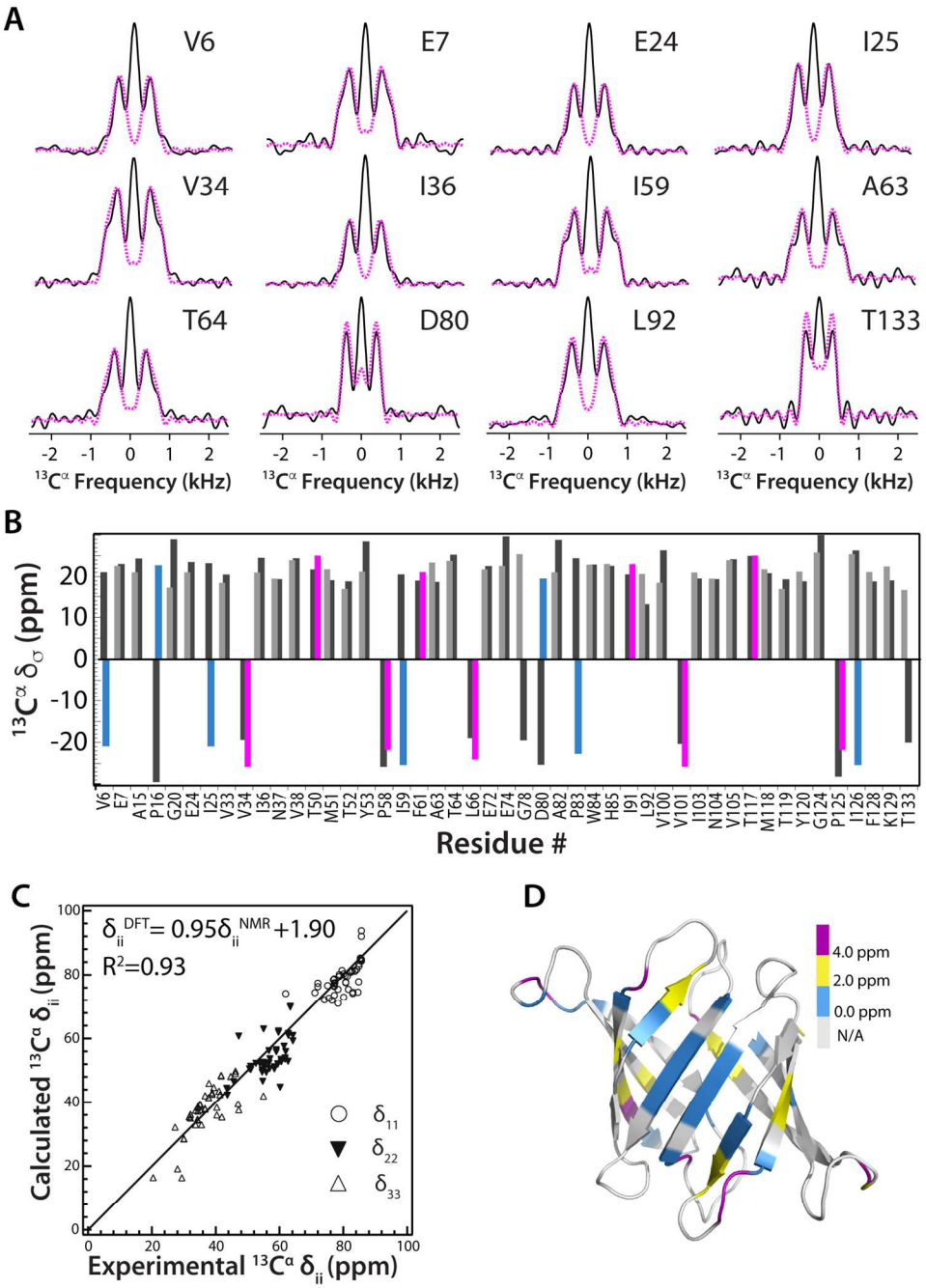


Figure 3. A: Experimental (solid black lines) and simulated (dashed magenta lines) $^{13}C^{\alpha}$ RNCSA lineshapes plotted for selected OAA residues. **B:** The reduced anisotropy parameters, δ_{σ}, of the $^{13}C^{\alpha}$ CSA tensors of OAA, plotted as a function of residue number. The calculated QM(DFT)/MM values are shown as dark grey bars. The experimental δ_{σ} values are shown in magenta if the signs agree with the calculations; in blue if the signs are opposite, and in light grey if the signs are undetermined (asymmetry parameters close to 1). **C:** QM(DFT)/MM calculated $^{13}C^{\alpha}$ CSA tensor principal components δ_{ii}, plotted vs. the corresponding experimental parameters. **D:** Differences between experimental and computed $^{13}C^{\alpha}$ δ_{σ} (ppm) values are mapped onto the OAA structure.

FIGURE 4

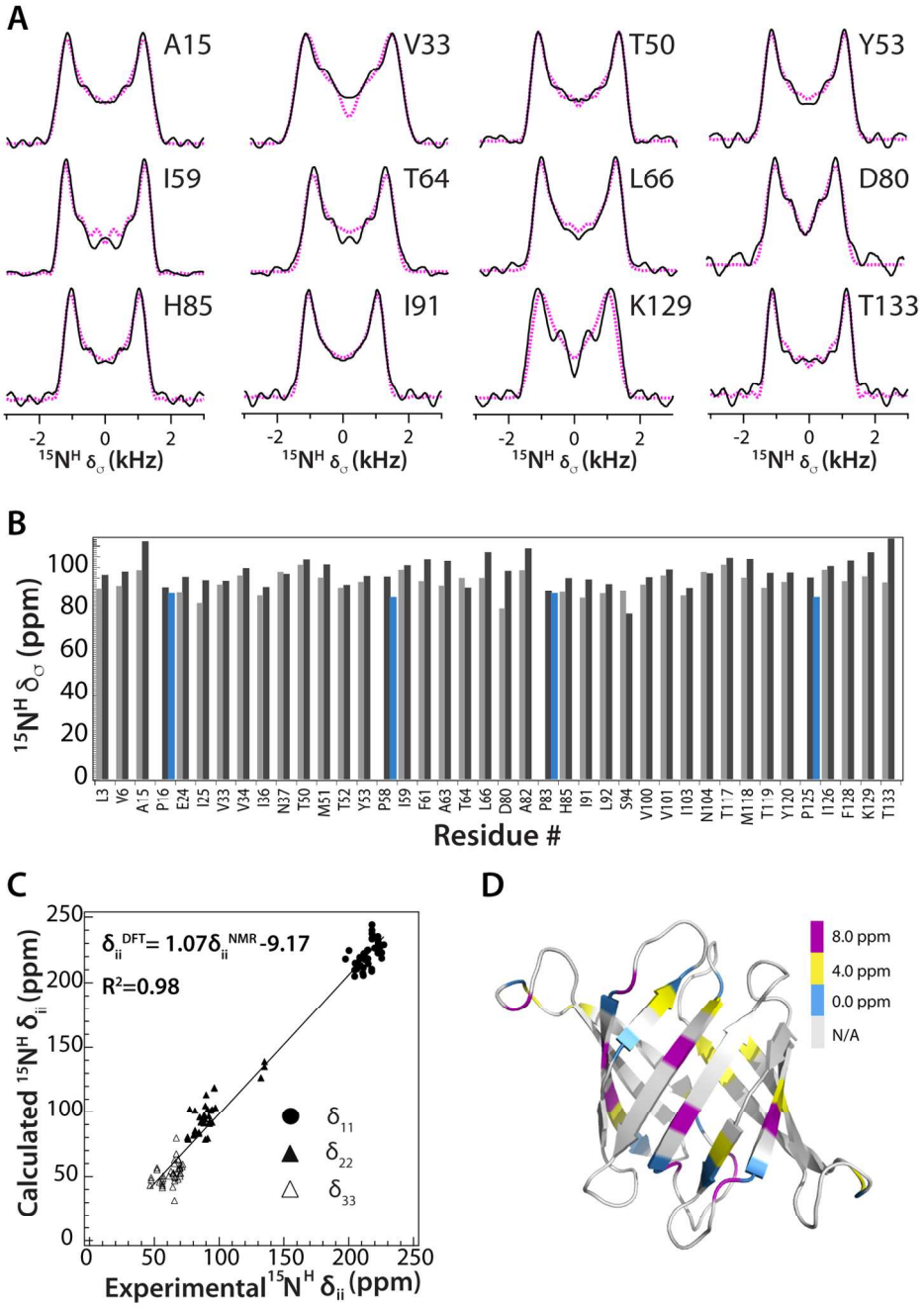


Figure 4. A: Experimental (solid black lines) and simulated (dashed magenta lines) $^{15}N^{H}$ RNCSA lineshapes plotted for selected OAA residues. **B:** The reduced anisotropy parameters of the $^{15}N^{H}$ CSA tensors of OAA, δ_σ, plotted as a function of the residue number. The calculated QM(DFT)/MM values are shown in dark grey bars and the experimental δ_σ values in light grey bars. δ_σ values for the four Pro residues, which are close to 1 and therefore of undetermined sign are shown in blue. **C:** The QM(DFT)/MM calculated $^{15}N^{H}$ CSA tensor principal components δ_{ii}, plotted vs. the corresponding experimental parameters. **D:** Differences between experimental and computed $^{15}N^{H}$ δ_σ values (ppm) are mapped onto the OAA structure.

FIGURE 5

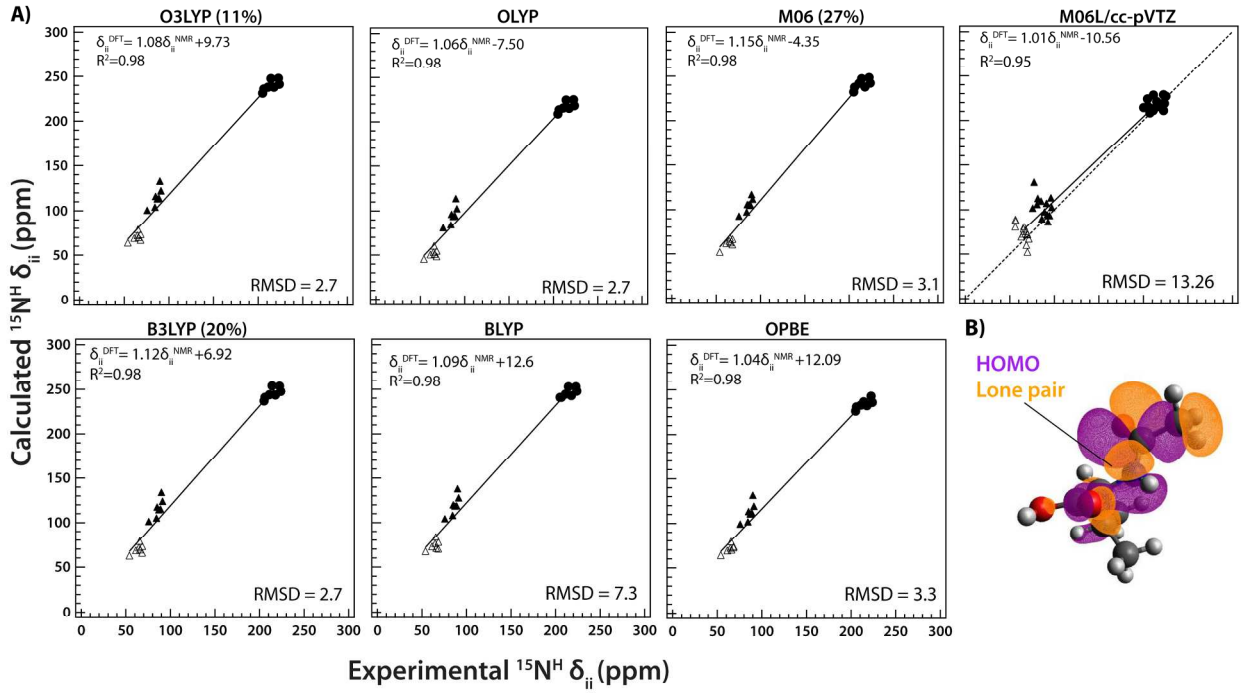


Figure 5. A) Evaluation of the performance of different functionals for the accuracy of QM(DFT)/MM calculated $^{15}N^{H}$ CSA parameters. The tensor principal components δ_{ii}, are plotted vs. the corresponding experimental parameters for O3LYP (11% Hartree-Fock exchange term), OLYP, M06 (27% Hartree-Fock exchange term), M06L (27% Hartree-Fock exchange term), B3LYP (20% Hartree-Fock exchange term), BLYP, and OPBE. The RMSD values for the reduced anisotropy parameter, δ_{σ}, are listed in each plot. Note that for the reduced anisotropy, most accurate predictions are reached with the M06L functional. The basis set was TZVP for all calculations, except for M06L, for which the cc-pVTZ basis set was used. B) HOMO of N-Acetyl valine calculated in Gaussian09 with the PBE0 functional using NBO 3.1 program as implemented in Gaussian09. Note that the negative electron density of the nitrogen lone pair occupies significant space on the opposite side of the peptide plane. Taking the kinetic energy of the electrons into account improves the accuracy of the chemical shift calculations.

FIGURE 6

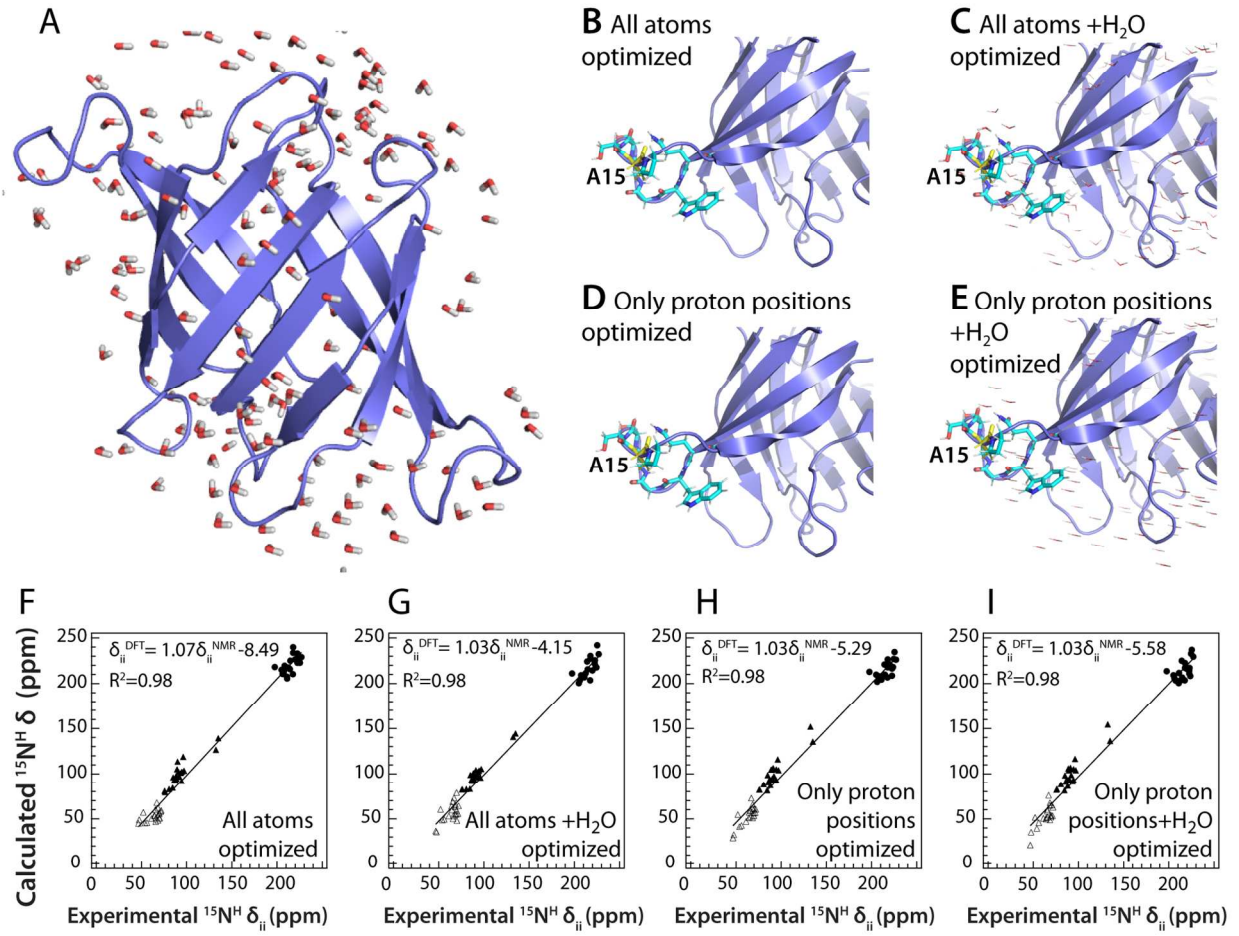


Figure 6. A: Ribbon representation of OAA showing those crystallographic water molecules that were included in the QM(DFT)/MM calculations of NMR chemical shifts. The proton positions were added using the MM force field libraries. **B-E:** Optimized geometries of the QM region for A15 (shown in yellow). **F-I:** QM(DFT)/MM calculated $^{15}N^{H}$ CSA tensor principal components δ_{ii}, plotted vs. the corresponding experimental parameters.

FIGURE 7

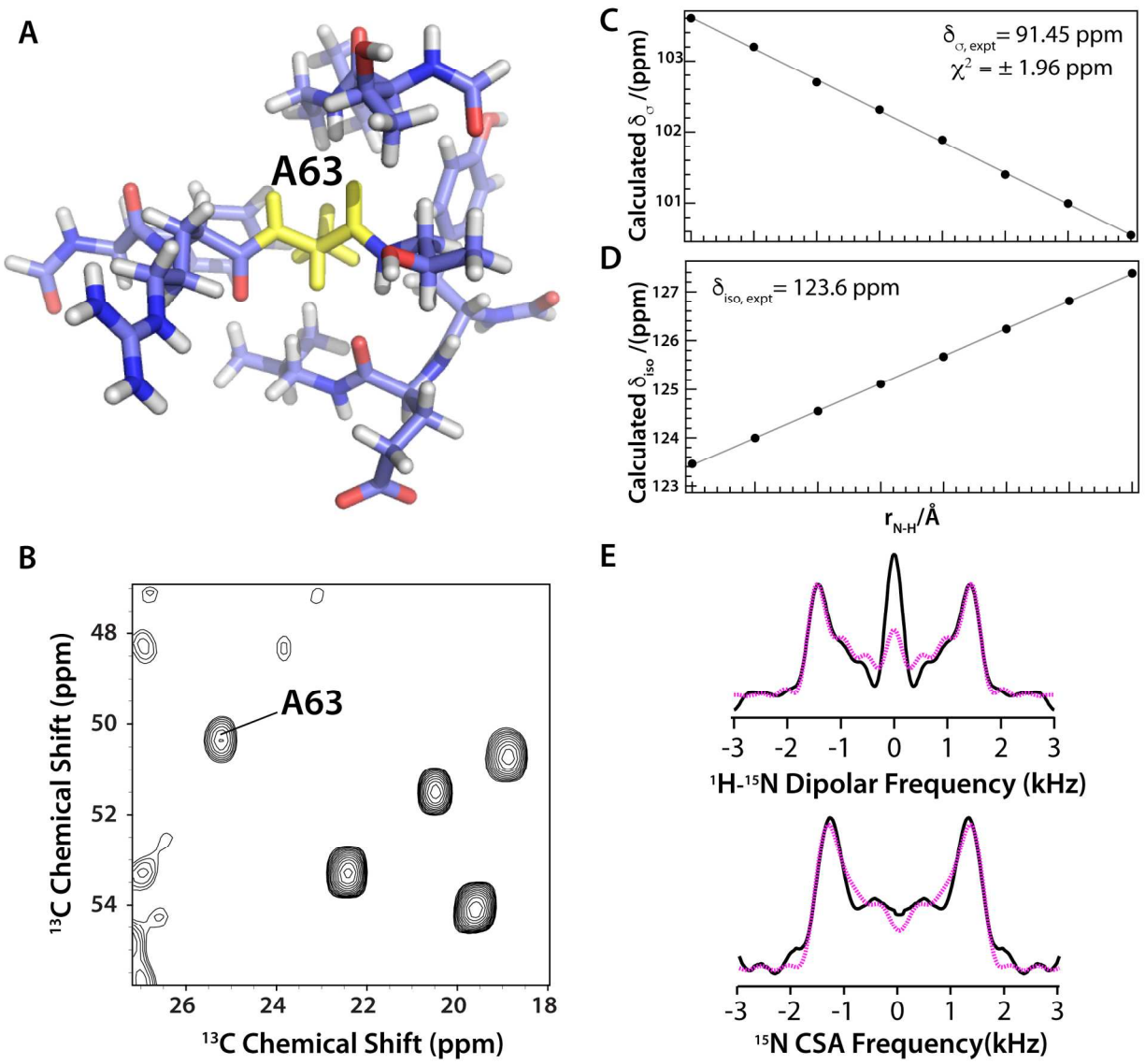


Figure 7. Influence of the NH bond length on the QM(DFT)/MM calculated $^{15}N^H$ CSA parameters for residue A63. **A:** Geometry-optimized QM region around A63. The backbone atoms are shown in yellow. **B:** Left; Expansion around the A63 resonance in the 3D CORD spectrum, illustrating that no resonance overlap is present. **C, D:** The QM(DFT)/MM calculated $^{15}N^H$ CSA parameters δ_σ and δ_{iso}, plotted vs. the N-H bond length. **E:** PARS ^{1}H-^{15}N dipolar (top) and ^{15}N RNCSA (bottom) lineshapes for A63. The experimental lineshapes are shown in black solid lines, the simulated lineshapes are in dotted magenta lines.

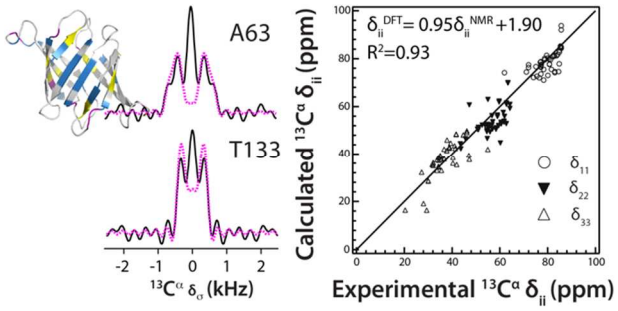
REFERENCES

1. D. A. Case, *Curr. Opin. Struct. Biol.*, 2013, **23**, 172-176.
2. M. Carravetta, M. Eden, X. Zhao, A. Brinkmann and M. H. Levitt, *Chem. Phys. Lett.*, 2000, **321**, 205-215.
3. G. Pileio, M. Concistre, N. McLean, A. Gansmulller, R. C. D. Brown and M. H. Levitt, *J. Magn. Reson.*, 2007, **186**, 65-74.
4. G. Hou, I.-J. L. Byeon, J. Ahn, A. M. Gronenborn and T. Polenova, *J. Chem. Phys.*, 2012, **137**, 134201.
5. H. L. Zhang, G. J. Hou, M. M. Lu, J. Ahn, I. J. L. Byeon, C. J. Langmead, J. R. Perilla, I. Hung, P. L. Gor'kov, Z. H. Gan, W. W. Brey, D. A. Case, K. Schulten, A. M. Gronenborn and T. Polenova, *Journal of the American Chemical Society*, 2016, **138**, 14066-14075.
6. L. J. Mueller and M. F. Dunn, *Acc. Chem. Res.*, 2013, **46**, 2008-2017.
7. J. R. Brender, D. M. Taylor and A. Ramamoorthy, *Journal of the American Chemical Society*, 2001, **123**, 914-922.
8. C. J. Jameson, *Annu. Rev. Phys. Chem.*, 1996, **47**, 135-169.
9. B. J. Wylie, C. D. Schwieters, E. Oldfield and C. M. Rienstra, *Journal of the American Chemical Society*, 2009, **131**, 985-992.
10. Y. Tian, G. J. Lu, F. M. Marassi and S. J. Opella, *J. Biomol. NMR*, 2014, **60**, 67-71.
11. S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Phys.*, 2014, **141**.
12. B. G. Caulkins, B. Bastin, C. Yang, T. J. Neubauer, R. P. Young, E. Hilario, Y. M. M. Huang, C. E. A. Chang, L. Fan, M. F. Dunn, M. J. Marsella and L. J. Mueller, *Journal of the American Chemical Society*, 2014, **136**, 12824-12827.
13. G. Monard and K. M. Merz, *Acc. Chem. Res.*, 1999, **32**, 904-911.
14. K. M. Merz, *Acc. Chem. Res.*, 2014, **47**, 2804-2811.
15. S. Tang and D. A. Case, *J. Biomol. NMR*, 2011, **51**, 303-312.
16. J. Swails, T. Zhu, X. He and D. A. Case, *J. Biomol. NMR*, 2015, **63**, 125-139.
17. J. D. Hartman, T. J. Neubauer, B. G. Caulkins, L. J. Mueller and G. J. O. Beran, *J. Biomol. NMR*, 2015, **62**, 327-340.
18. M. Fritz, C. M. Quinn, M. Z. Wang, G. J. Hou, X. G. Lu, L. M. I. Koharudin, T. Polenova and A. M. Gronenborn, *J. Phys. Chem. B*, 2017, **121**, 3574-3585.
19. L. M. I. Koharudin, S. Kollipara, C. Aiken and A. M. Gronenborn, *J. Biol. Chem.*, 2012, **287**, 33796-33811.
20. L. M. I. Koharudin and A. M. Gronenborn, *Structure*, 2011, **19**, 1170-1181.
21. L. M. I. Koharudin, W. Furey and A. M. Gronenborn, *J. Biol. Chem.*, 2011, **286**, 1588-1597.
22. M. G. Carneiro, L. M. I. Koharudin, C. Griesinger, A. M. Gronenborn and D. Lee, *Biomol. NMR Assign.*, 2015, **9**, 317-319.
23. G. Hou, I.-J. L. Byeon, J. Ahn, A. M. Gronenborn and T. Polenova, *Journal of the American Chemical Society*, 2011, **133**, 18646-18655.
24. J. C. C. Chan and R. Tycko, *J. Chem. Phys.*, 2003, **118**, 8378-8389.
25. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277-293.
26. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, P. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687-696.
27. T. J. Stevens, R. H. Fogh, W. Boucher, V. A. Higman, F. Eisenmenger, B. Bardiaux, B. J. van Rossum, H. Oschkinat and E. D. Laue, *J. Biomol. NMR*, 2011, **51**, 437-447.
28. M. Bak, J. T. Rasmussen and N. C. Nielsen, *J. Magn. Reson.*, 2011, **213**, 366-400.
29. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M.

Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Journal*, 2009, DOI: citeulike-article-id:9096580.

30. B. J. Wylie, T. Franks, D. T. Graesser and C. M. Rienstra, *Journal of the American Chemical Society*, 2005, **127**, 11946-11947.
31. R. Kusumi, F. Kimura, G. Song and T. Kimura, *J. Magn. Reson.*, 2012, **223**, 68-72.
32. E. Oldfield, *Annu. Rev. Phys. Chem.*, 2002, **53**, 349-378.
33. E. Oldfield, *J. Biomol. NMR*, 1995, **5**, 217-225.
34. H. B. Le and E. Oldfield, *J. Phys. Chem.*, 1996, **100**, 16423-16428.
35. A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829-5835.
36. S. T. Holmes, R. J. Iuliucci, K. T. Mueller and C. Dybowski, *J. Chem. Theor.*, 2015, **11**, 5229-5241.
37. C. J. Jameson and A. C. De Dios, in *Nuclear Magnetic Resonance, Vol 42*, eds. K. KamienskaTrela and J. Wojcik, 2013, vol. 42, pp. 45-77.
38. C. J. Jameson and A. C. Dedios, *J. Chem. Phys.*, 1993, **98**, 2208-2217.
39. C. M. Widdifield and R. W. Schurko, *Conc. Magn. Reson.*, 2009, **34A**, 91-123.
40. A. Barszczewicz, M. Jaszunski, T. Helgaker and K. Ruud, *Chem. Phys. Lett.*, 1996, **250**, 1-8.
41. M. K. Pandey and A. Ramamoorthy, *J. Phys. Chem. B*, 2013, **117**, 859-867.
42. D. T. Cui, R. L. Koder, P. L. Dutton and A. F. Miller, *J. Phys. Chem. B*, 2011, **115**, 7788-7798.
43. G. P. M. Holland, Qiushi; Yarger, Jeffery L., *Chem. Commun.*, 2013, **49**, 6680-6682.

TOC entry



Experimental-NMR and calculated-QM/MM $^{13}C^{\alpha}$ and $^{15}N^{H}$ chemical-shift tensors are presented for microcrystalline OAA; factors determining the calculation accuracy are discussed.