J|A|C|S
ARTICLES
Published on Web 03/18/2004

Tryptophan Chemical Shift in Peptides and Proteins: A Solid State Carbon-13 Nuclear Magnetic Resonance Spectroscopic and Quantum Chemical Investigation

Haihong Sun[†] and Eric Oldfield*[,†,‡]

Contribution from the Departments of Biophysics and Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received November 4, 2003; Revised Manuscript Received January 20, 2004; E-mail: eo@chad.scs.uiuc.edu

Abstract: We have obtained the carbon-13 nuclear magnetic resonance spectra of a series of tryptophan-containing peptides and model systems, together with their X-ray crystallographic structures, and used quantum chemical methods to predict the ^{13}C NMR shifts (or shieldings) of all nonprotonated aromatic carbons (C^{γ}, C^{δ_2}, and C^{ϵ_2}). Overall, there is generally good accord between theory and experiment. The chemical shifts of Trp C^{γ} in several proteins, hen egg white lysozyme, horse myoglobin, horse heart cytochrome *c*, and four carbonmonoxyhemoglobins, are also well predicted. The overall Trp C^{γ} shift range seen in the peptides and proteins is 11.4 ppm, and individual shifts (or shieldings) are predicted with an rms error of ~1.4 ppm (R value = 0.86). Unlike C^{α} and N^{H} chemical shifts, which are primarily a function of the backbone ϕ,ψ torsion angles, the Trp C^{γ} shifts are shown to be correlated with the side-chain torsion angles χ_1 and χ_2 and appear to arise, at least in part, from γ-gauche interactions with the backbone C′ and N^{H} atoms. This work helps solve the problem of the chemical shift nonequivalences of nonprotonated aromatic carbons in proteins first identified over 30 years ago and opens up the possibility of using aromatic carbon chemical shift information in structure determination.

Introduction

Folding a protein into its native conformation results in large ranges of nuclear magnetic resonance (NMR) chemical shift nonequivalence: about 3 ppm for ^{1}H, 10 ppm for ^{13}C, 35 ppm for ^{15}N, and ~20 ppm for the probe nucleus ^{19}F.[1,2] For most nuclei (aliphatic and ^{13}C′ carbons, ^{15}N, and ^{19}F), the origins of such folding-induced chemical shifts are now relatively well understood. For backbone carbons and nitrogens, ϕ, ψ torsion angles dominate;[1] for ^{13}C′, hydrogen bond geometry effects play a major role[3], while for ^{19}F, electrostatic field effects dominate the shielding response.[1,4] However, for aromatic carbons, such as Trp C^{γ}, the origins of the large chemical shift range first seen over 30 years ago have been unclear.[5−7] In earlier unpublished work, we attempted to predict the observed chemical shifts in hen egg white lysozyme and in a small series of model Trp-containing peptides and other model systems using both Hartree−Fock and density functional theory methods, but with little success. Since these same methods (including charge field perturbation[8]) had been successful in predicting all other chemical shifts, the origins of their poor performance in predicting, in particular, the Trp C^{γ} chemical shifts, had been puzzling, but it seemed likely that the experimental chemical shift range being investigated was simply too small. However, other problems, such as the presence of molecular motion, solvation, and crystal/solution structural differences, might also play a role with the proteins. We have thus synthesized a larger series of Trp-containing model compounds and obtained their solid-state ^{13}C NMR spectra to remove any dynamical and solvation uncertainties, as well as to (potentially) increase the overall shift range, and to provide a series of highly resolved structures. Then, we employ a variety of quantum chemical methods to predict the ^{13}C NMR shifts, utilizing the high-resolution X-ray crystallographic structures obtained on the same samples used to record the ^{13}C NMR spectra. This approach is shown to enable the first predictions of the ^{13}C NMR chemical shifts of the nonprotonated aromatic carbons C^{γ}, C^{δ_2}, and C^{ϵ_2} in Trp model systems. The same methods are then applied to Trp C^{γ} chemical shifts in proteins, using homology modeling predicted structures when the corresponding crystallographic structures are not available. Overall, our results show that, in particular, Trp C^{γ} shifts are strongly related to side-chain χ_1, χ_2 torsion angles (due to γ-gauche interactions with C′ and N^{H}), but not to backbone ϕ,ψ torsion angles. These results can be expected to be of general interest in the context of protein structure prediction and refinement and may also be applicable

[†] Department of Biophysics.
[‡] Department of Chemistry.

(1) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491−1496.
(2) Lian, C. Y.; Le, H. B.; Montez, B.; Patterson, J.; Harrell, S.; Laws, D.; Matsumura, I.; Pearson, J.; Oldfield, E. *Biochemistry* **1994**, *33*, 5238−5245.
(3) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 11485−11488.
(4) Pearson, J. G.; Oldfield, E.; Lee, F. S.; Warshel, A. *J. Am. Chem. Soc.* **1993**, *115*, 6851−6862.
(5) Allerhand, A.; Childers, R. F.; Oldfield, E. *Biochemistry* **1973**, *12*, 1335−1341.
(6) Groff, J. P.; London, R. E.; Cocco, L.; Blakley, R. L. *Biochemistry* **1981**, *20*, 6169−6178.
(7) Oldfield, E.; Allerhand, A. *J. Biol. Chem.* **1975**, *250*, 6403−6407.
(8) de Dios, A. C.; Oldfield, E. *Chem. Phys. Lett.* **1993**, *205*, 108−116.

10.1021/ja030612u CCC: $27.50 **© 2004 American Chemical Society**

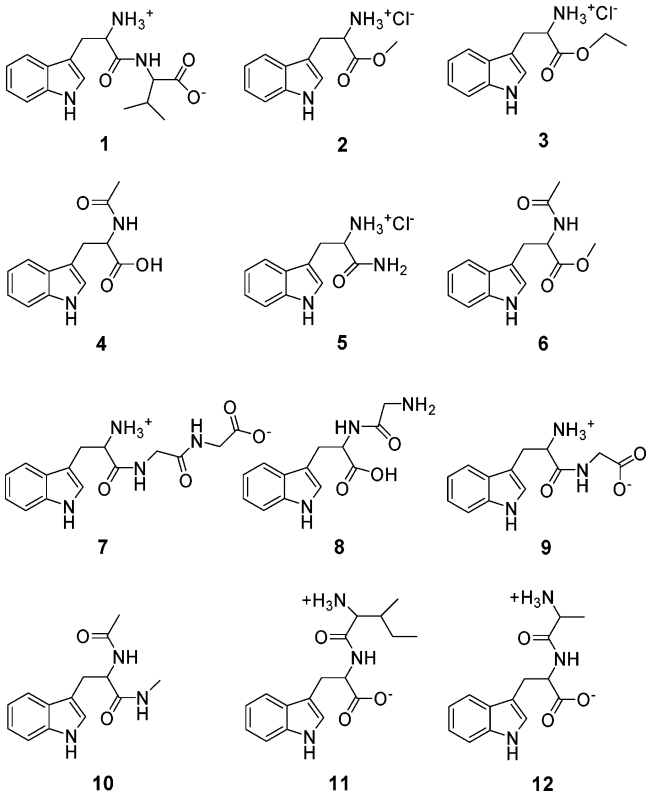


Figure 1. Structures of the 12 Trp-containing peptides and model compounds investigated.

to interpreting chemical shifts in other aromatic species in proteins, such as some drug molecules, as well as in interpreting the ^{13}C NMR chemical shifts of the other aromatic amino acids.

Experimental Section

Crystallographic Aspects. We crystallized 12 Trp-containing species (**1**−**12**) whose structures are shown in Figure 1. **1**, **7**−**9**, **11**, and **12** were obtained from Bachem (King of Prussia, PA), while **2**−**6** were from Sigma-Aldrich (St. Louis, MO). **10** was synthesized by a standard peptide coupling reaction.[9] All compounds were crystallized from water, except for **3**, **5**, and **6**, which were crystallized from methanol. Diffraction data for all 12 compounds were collected at 193 K on a Bruker SMART CCD system. Data reduction and integration were performed with the software package SAINT,[10] and absorption corrections were applied by using the program SADABS.[10] The positions of the non-hydrogen atoms were found by direct methods using the Bruker SHELXTL software package.[10] Crystal data and structure refinement information for **1**, **2**, **4**, and **11** (which have not been reported previously) are shown in Table 1, and more detailed descriptions of these structures (coordinates, geometries, and B factors) are provided in the Supporting Information. The actual structures are shown in Figure 2. The structures of the eight other compounds have been reported previously,[11−16] but we redetermined the structures of each compound crystallographically to ensure that the actual species isolated were as expected. All six χ_1, χ_2 side-chain combinations possible for Trp (χ_1 = 180°, 60°, −60°; χ_2 = ±90°) were represented in these 12 structures, resulting in a very large (11.4 ppm) range in Trp C^{γ} chemical shifts.

Table 1. Crystallographic Results

	1 TrpVal	**2** L-Trp-OMe·Cl	**4** Ac-L-Trp	**11** IleTrp
formula	$C_{16}H_{25}N_3O_5$	$C_{12}H_{15}ClN_2O_2$	$C_{13}H_{14}N_2O_3$	$C_{17}H_{27}N_3O_5$
mol wt	339.39	254.71	246.26	353.42
cryst syst	orthorhombic	triclinic	monoclinic	orthorhombic
space group	$P2_12_12_1$	$P1$	Pc	$P2_12_12_1$
a (Å)	5.7598(15)	5.262(3)	23.418(10)	5.6470(16)
b (Å)	15.998(4)	8.923(4)	14.0129(13)	8.581(3)
c (Å)	18.496(5)	14.004(7)	8.384(4)	36.343(11)
α (deg)	90	77.801(8)	90	90
β (deg)	90	89.771(8)	94.702(9)	90
γ (deg)	90	80.186(8)	90	90
V (Å^3)	1704.4(8)	632.9(5)	1203.8(9)	1761.2(9)
Z	4	2	4	4
ρ_{calcd} (g cm^{-3})	1.323	1.336	1.359	1.333
T (K)	193(2)	193(2)	193(2)	193(2)
μ (mm^{-1})	0.099	0.294	0.098	0.098
R1 (all data)	0.0310	0.0346	0.1008	0.0712
wR2 (all data)	0.0722	0.0800	0.1560	0.0917

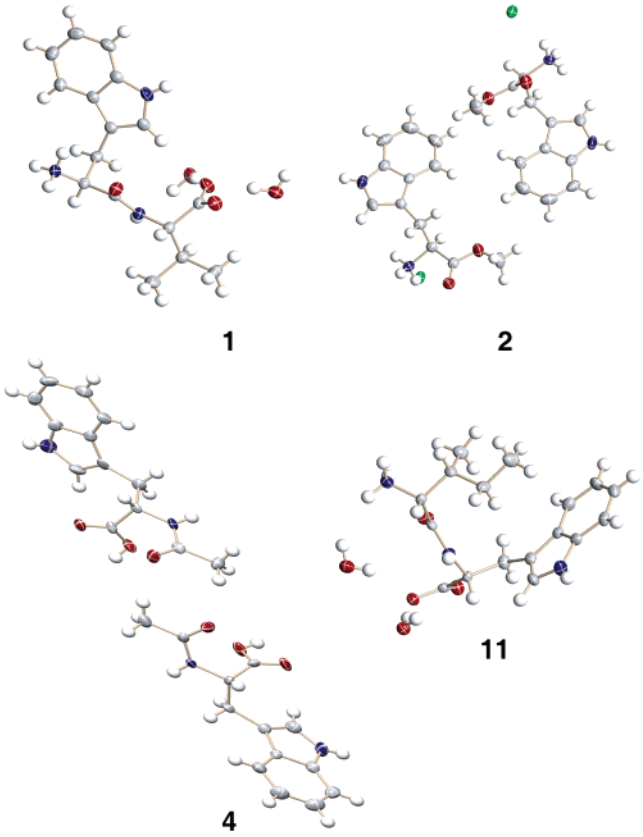


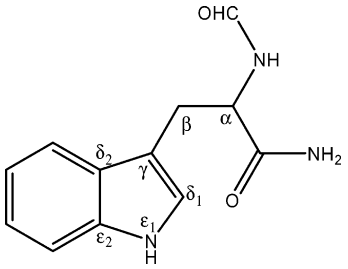
Figure 2. Crystal structures of compounds **1**, **2**, **4**, and **11**.

NMR Spectroscopy. Carbon-13 NMR spectra were obtained by using the cross-polarization magic-angle sample-spinning technique[17,18] with either full proton decoupling[19] or interrupted decoupling[20] (using a dipolar dephasing time of 120 μs) for selection of the nonprotonated Trp aromatic carbons (C^{γ}, C^{δ_2}, C^{ϵ_2}). Spectra were typically recorded

(9) Sheehan, J. C. *J. Am. Chem. Soc.* **1955**, *77*, 1067−1068.
(10) (a) *SAINT*; Bruker AXS, Inc.: Madison, WI, 2001. (b) SADABS; Bruker AXS, Inc.: Madison, WI, 2001. (c) *SHELXTL*; Bruker AXS, Inc.: Madison, WI, 2001.
(11) Emge, T. J.; Agrawal, A.; Dalessio, J. P.; Dukovic, G.; Inghrim, J. A.; Janjua, K.; Macaluso, M.; Robertson, L. L.; Stiglic, T. J.; Volovik, Y.; Georgiadis, M. M. *Acta Crystallogr., Sect. C* **2000**, *56*, E469−E471.
(12) Souhassou, M.; Aubry, A.; Lecomte, C. *Acta Crystallogr., Sect. C* **1990**, *46*, 1303−05.
(13) Vijayalakahmi, B. K.; Srinivasan, R. *Acta Crystallogr., Sect. B* **1975**, *31*, 999−1001.
(14) Wu, S.; Declercq, J. P.; Tinant, B.; Vanmeerssche, M. *Bull. Soc. Chim. Belg.* **1987**, *96*, 581−586.
(15) Cortrait, P. M.; Barrans, Y. *Acta Crystallogr., Sect. B* **1974**, *30*, 510.
(16) Takigawa, T.; Ashida, T.; Sasada, Y. *Bull. Chem. Soc. Jpn.* **1966**, *39*, 2369−2378.
(17) Pines, A.; Gibby, M. G.; Waugh, J. S. *J. Chem. Phys.* **1973**, *59*, 569−590.
(18) Schaefer, J.; Stejskal, E. O. *J. Am. Chem. Soc.* **1976**, *98*, 1031−1032.
(19) Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G. *J. Chem. Phys.* **1995**, *103*, 6951−6958.
(20) Opella, S. J.; Frey, M. H. *J. Am. Chem. Soc.* **1979**, *101*, 5855−5857.

using a 5 s recycle time. The ^{1}H and ^{13}C 90° pulse widths were both 2.75 μs. Chemical shifts were referred to external glycine, setting the α carbon to 43.6 ppm downfield from tetramethylsilane (TMS). All spectra were obtained by using a Varian (Palo Alto, CA) Infinityplus 500 MHz (^{1}H) NMR spectrometer.

Computational Aspects. We used both Hartree−Fock and density functional theory methods as embodied in the Gaussian 98 program[21] to evaluate the Trp chemical shieldings. Most calculations were performed using the continuous set of gauge transformations (CSGT) method.[22,23] Two sets of structures were used. In the first, we used the model system, N-formyl tryptophan amide (FTA):



basically as described previously for amino acid C^α, C^β chemical shielding calculations.[24] Also, consistent with our previous studies, we employed an average Trp geometry, in this case based on an average of the 12 structures determined crystallographically. The actual bond lengths and three atom bond angles are given in the Supporting Information and are very close to those found in standard force fields, for example, 0.004 Å and 1.3° rms deviations from Amber force-field values.[25] Use of this "average" geometry had little effect on the computed shifts or shieldings in the model compounds, but is essential to predict Trp chemical shifts in proteins where, of course, the Trp bond lengths and angles cannot be determined so accurately from protein diffraction results. In a second set of calculations on the model compounds (**1**−**12**), we incorporated the effects of charge field perturbation, basically as described previously.[8,26] In all cases, we used a uniform (6-311++G (2d,2p)) basis set. We also carried out a series of DFT calculations on the FTA model systems, using the B3LYP functional.[27,28] For several heme proteins, diffraction results were not available, and therefore homology models were used, employing the SWISS-MODEL protein modeling program.[29]

Results and Discussion

Experimental Results. We show in Figure 3 the ^{13}C MAS NMR spectra of **1** (TrpVal) and **12** (AlaTrp). Figure 3A shows

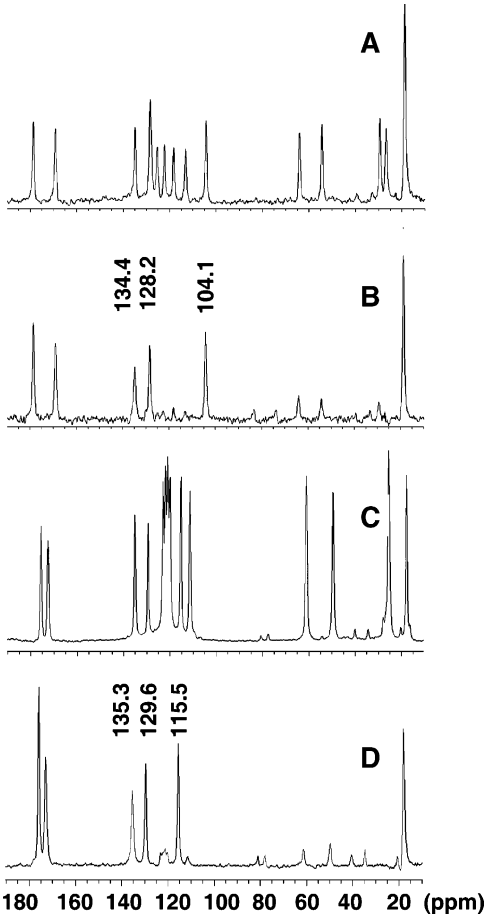


Figure 3. 125 MHz ^{13}C MAS NMR spectra (spin speed is 12 kHz) of **1** and **12**. (A) **1**, fully decoupled and (B) with interrupted decoupling. (C) **12**, fully decoupled and (D) with interrupted decoupling. The Trp C^γ chemical shift range is 11.4 ppm.

the fully proton-decoupled spectrum of **1**, while Figure 3B shows the spectrum of **1** obtained with interrupted decoupling, which results in the rapid dephasing and loss of signal from all of the rigid, protonated carbons. In our previous solution NMR studies of proteins (e.g., refs 5 and 7) essentially the same dephasing effect was accomplished by use of weak, noise-modulated off-resonance proton decoupling.[30] The features of primary interest in Figure 3 are the chemical shifts of the Trp C^γ, C^{δ_2}, and C^{ϵ_2} carbons, which are found to resonate at 104.1, 128.2, and 134.4 ppm downfield from TMS in TrpVal (**1**) and at 115.5, 129.6, and 135.3 ppm in AlaTrp (**12**). The chemical shift of C^γ in TrpVal (**1**) (at 104.2 ppm) is of particular interest since this is the most upfield or shielded value found to date for Trp, either in a protein or in a Trp-containing model compound. On the other hand, in AlaTrp (**12**), the results shown in Figure 3C,D clearly show that in this system, the Trp C^γ is highly deshielded, having an isotropic chemical shift (δ_i) of 115.5 ppm. This is further downfield than any previously reported Trp C^γ chemical shift. For the 10 other compounds (**2**−**11**), the Trp C^γ chemical shifts occur within the 11.4 ppm chemical shift range described by these two model compounds. The actual chemical shifts for each of the 12 peptides and model systems are shown in Table 2 and, graphically, in Figure 4A. These model compound shifts are to be compared with those

(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.
(22) Keith, T. A.; Bader, R. F. W. *Chem. Phys. Lett.* **1992**, *194*, 1−2.
(23) Cheeseman, J. R.; Frisch, M. J.; Trucks, G. W.; Keith, T. A. *J. Chem. Phys.* **1996**, *104*, 5497−5509.
(24) Sun, H.; Sanders, L. K.; Oldfield, E. *J. Am. Chem. Soc.* **2002**, *124*, 5486−5495.
(25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(26) de Dios, A. C.; Laws, D. D.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 7784−7786.
(27) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(28) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(29) (a) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. *Nucleic Acids Res.* **2003**, *31*, 3381−3385. (b) Swiss Model Home Page. http://swissmodel.expasy.org/.
(30) Wenkert, E.; Clouse, A. O.; Cochran, D. W.; Doddrell, D. *J. Am. Chem. Soc.* **1969**, *91*, 6879−6880.

Table 2. Experimental Chemical Shifts and Computed Chemical Shieldings for Nonprotonated Aromatic Carbons in Tryptophan Model Compounds and Proteins

	C^{γ}			C^{δ_2}			C^{ϵ_2}		
compd	exptl[a]	calcd[b]	calcd[c]	exptl[a]	calcd[b]	calcd[c]	exptl[a]	calcd[b]	calcd[c]
1	104.1	82.1	90.8	128.2	60.6	62.5	134.4	51.3	52.5
2	105.4		93	128.7		60.5	137.2		49.7
3	105.5		90.2	127.7		60.5	137		49.9
4	106.9	81.9	82.8	125.8	62.1	64.6	136.3	50.5	50.4
5	107.4		78.6	129		59.5	135.3		50.2
6	108.7	76.8	82.8	126.5	63.4	63.3	136.8	49.8	49.5
7	109.4	75.7	84.8	126.8	62.5	66.1	138.2	49.8	48.4
8	109.7	77.0	74.0	126.5	63.1	61.3	134.3	50.7	51
9	110	78.0	84.6	128.8	60.4	62.8	137.5	49.8	48
10	110.4	77.0	80.9	126.9	62.9	61	136.9	49.6	49.1
11	112.8	73.7	72.9	128.8	62.5	59.1	135.9	49.4	50.6
12	115.5	74.0	69.9	129.6	61.4	57.6	135.3	50.1	49.8
28[d]	108.7	75.8		127.1	64.0		138.3	49.1	
62[d]	110.8	75.2		127.1	63.0		137.7	49.6	
63[d]	110.3	76.2		127.7	63.3		138.8	49.6	
108[d]	112.6	73.8		129.6	61.0		136.3	49.9	
111[d]	108.7	79.5		128.5	61.6		137.9	50.0	
123[d]	111.9	75.4		127.1	63.9		137.9	49.0	
59[e]	110.3	75.1			63.7			49.1	

[a] Experimental shift values in ppm from TMS, from CP-MAS spectra (**1**−**12**) and refs 5 and 31. [b] HF/CSGT computed shieldings using the FTA model and the chloride-free crystallographic structures obtained in this work, together with PDB files 1BWI (hen egg white lysozyme) and 1HRC (horse heart cytochrome *c*). [c] Computed tryptophan derivative (**1**−**12**) shieldings obtained by using charge field perturbation. A uniform 6-311++G (2d,2p) basis was used for the central molecule, while neighboring molecules were represented by point charges, obtained from separate ab initio calculations. [d] Trp residues in 1BWI (hen egg white lysozyme). [e] Trp residue in 1HRC (horse heart cytochrome *c*).

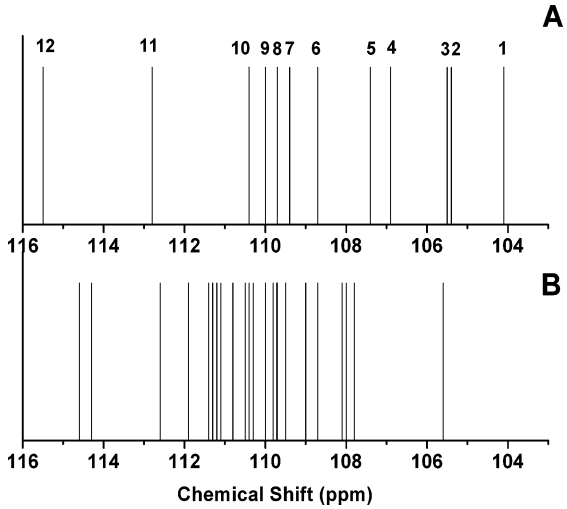


Figure 4. Schematic illustrating the Trp C^{γ} chemical shift nonequivalences observed in (A) model systems (**1**−**12**) and (B) in proteins. Data points for hen egg white lysozyme, horse heart myoglobin, four hemoglobins, horse heart cytochrome *c*, and a dihydrofolate reductase (from *S. faecium*).[5−7,31]

found in proteins, as for example shown in Figure 4B, which includes results for the Trp C^{γ} in hen egg white lysozyme (HEWL), cytochrome *c*, human adult carbonmonoxyhemoglobin, human fetal carbonmonoxyhemoglobin, bovine fetal carbonmonoxyhemoglobin, chicken AII carbonmonoxyhemoglobin, horse carbonmonoxymyoglobin, and the dihydrofolate reductase (DHFR) isozyme 2 from *Streptococcus faecium*.[5−7,31] The model compound shift results are of interest since they considerably extend the known Trp C^{γ} chemical shift range

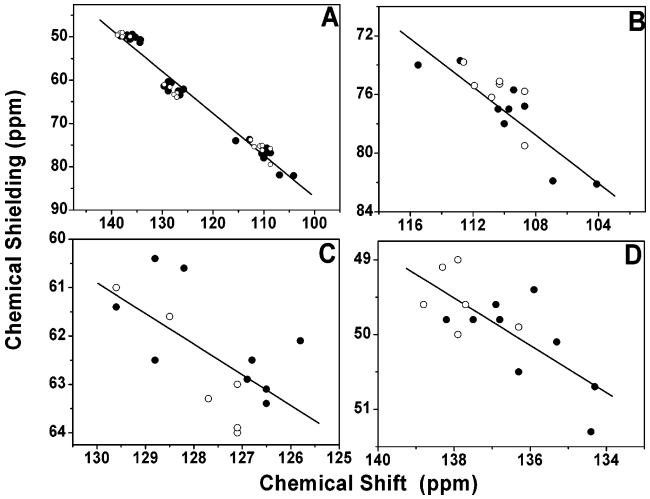


Figure 5. Correlations between experimental chemical shifts and HF/CSGT computed shieldings using FTA models for C^{γ}, C^{δ_2}, and C^{ϵ_2} (nonprotonated aromatic carbon) sites for nine tryptophan derivatives (●) and seven tryptophan residues (○) in proteins.[5,31] (A) All nonprotonated carbons, (B) C^{γ}, (C) C^{δ_2}, (D) C^{ϵ_2}.

Table 3. Statistical Results for Correlations between Experimental Chemical Shifts and Chemical Shieldings for Nonprotonated Aromatic Carbons in Tryptophan Species

atom type	method	*R*	slope	SD (ppm)
C^{γ_2}	FTA[a]	−0.89	−0.82	1.5
	FTA-all[b]	−0.85	−0.86	1.4
	CFP[c]	−0.82	−1.7	4.1
C^{δ_2}	FTA[a]	−0.64	−0.52	0.88
	FTA-all[b]	−0.65	−0.63	0.91
	CFP[c]	−0.66	−1.4	2.1
C^{ϵ_2}	FTA[a]	−0.70	−0.32	0.46
	FTA-all[b]	−0.74	−0.31	0.41
	CFP[c]	−0.87	−0.72	0.54

[a] HF/CSGT method on nine model compounds (without chloride ions): FTA model. [b] HF/CSGT method on nine model compounds (without chloride ions) together with seven Trp residues in proteins (Table 2): FTA model. [c] HF/CSGT CFP results for the 12 model compounds (**1**−**12**) including charge field perturbation.

found in proteins (the assignment of the most deshielded protein Trp C^{γ} shift is not certain),[6] from 6.8 to 11.4 ppm, which can be expected to facilitate testing our ability to predict Trp C^{γ} shifts using quantum chemical methods.

Computational Results. To begin with, we first consider the results of our chemical shielding calculations using the N-formyltryptophan amide model compound, an approach which we have used successfully in the past to predict (primarily) C^{α} chemical shifts (or shieldings) in amino acids and proteins.[1,24,26,32] We focus on the shifts of C^{γ}, C^{δ_2}, and C^{ϵ_2}, the nonprotonated aromatic carbons selected in the interrupted decoupling, and previous noise off-resonance proton decoupling experiments, where assignments can be readily made. As may be seen in Figure 5A and Table 2, there is a good overall correlation between the experimental shifts and the calculated shieldings when using the Hartree−Fock/CSGT method on the chloride free model compounds, with an *R* value of −0.96 and an overall rms deviation from the fitted line of 1.9 ppm. For the C^{γ} carbons alone, the correlation between theory and experiment is also good, as shown in Table 3, with an *R* value

(31) Oldfield, E.; Allerhand, A. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 3531−3535.

(32) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *J. Am. Chem. Soc.* **1993**, *115*, 9768−9773.

of −0.89, a slope of −0.82 (to be compared with ideal values of −1.0) and an rms error of 1.45 ppm over the entire 11.4 ppm chemical shift range. We next used the same approach to predict the nonprotonated aromatic carbon ^{13}C NMR chemical shifts in HEWL and cytochrome *c*, and these results are also shown in Figure 5 (open circles) and in Table 2. Addition of these protein chemical shifts has essentially no effect on the statistics (Table 3). There is thus a generally good correlation between the experimental shifts and the predicted chemical shieldings, especially for C^{γ}; thus, these results indicate that it is now possible to predict Trp C^{γ} chemical shifts in proteins with quite good accuracy. However, in the absence of the new solid-state NMR results on model compounds, which almost double the experimental C^{γ} chemical shift range, the correlation would be less convincing. For C^{δ_2} and C^{ϵ_2}, the presence of only an ∼4−5 ppm range in isotropic chemical shift illustrates this problem, as shown in Figure 5C,D, where we now find $|R|$ values of only ∼0.6−0.7 and poorer slopes (Table 3). The use of density functional theory (DFT) methods offers no improvements, with the HF and DFT/B3LYP results being highly correlated ($R = -0.96$, with a standard deviation of 0.4 ppm). Thus, both Hartree−Fock and DFT methods may be used to predict Trp chemical shifts, with the best results being found for C^{γ}, which has the largest range in isotropic chemical shift.

These results raise the question as to the actual origins of the chemical shifts for C^{γ}, C^{δ_2}, and C^{ϵ_2}: can they be correlated with any obvious structural parameters? For example, it might be that the large experimental chemical shift range for C^{γ} originates from essentially a conformational effect, since C^{γ} is obviously closer to the backbone C^{α} than are C^{δ_2} and C^{ϵ_2}; therefore, perhaps ϕ,ψ or χ_1,χ_2 effects might dominate shielding. However, for C^{ϵ_2}, it might be that electrostatic field effects would be important, due to its proximity to N^{ϵ_1}, and the omission of electrostatic field effects might be the cause of the very small slopes seen for C^{ϵ_2} (Table 3). We thus next investigated the effects of such electrostatic field effects on the computed shieldings by using the charge field perturbation method used previously to model electrostatic interactions in model systems.[1,8,26]

We show in Figure 6 the results of such CFP−CSGT calculations. The overall correlation is quite good (Figure 6A), but clearly the slope for C^{γ} (−1.7) is much worse than that found in the absence of CFP (−0.89), although use of the CFP method does enable us to include results for the three chloride salts (**2**, **3**, and **5**) in the correlation (Figure 6). Also, the slope for C^{ϵ_2} improves, from −0.31 to −0.72, as does its *R* value, suggesting the importance of electrostatic field effects for this site.

Structural Correlations and Shielding Tensors. The results we have described above show that the large (11.4 ppm) Trp C^{γ} chemical shift range seen in peptides and proteins can now be described by use of quantum chemical calculations and that these large Trp C^{γ} shifts (or shieldings) are those which can be the most precisely predicted, in terms of *R* values and slopes. We thus next raise the questions: how do the Trp C^{γ} chemical shieldings change as a function of geometry? Can we make simple correlations between the observed shifts (or computed shieldings) and structure? Can we predict Trp C^{γ} chemical shifts in other proteins, where crystallographic structures are unknown?

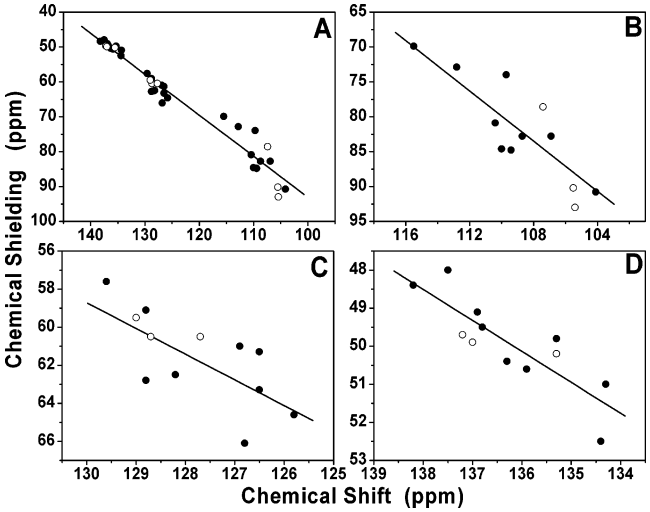


Figure 6. Correlations between experimental chemical shifts and HF/CSGT/CFP computed shieldings for C^{γ}, C^{δ_2}, and C^{ϵ_2} in model systems **1**−**12**. (A) All nonprotonated carbons, (B), C^{γ}, (C) C^{δ_2}, (D) C^{ϵ_2}. Symbols: (●), no chloride present; (○), chloride present.

And how do the shielding tensor magnitudes and orientations change with structure?

To begin with, we consider how the isotropic chemical shifts or shieldings vary with structure. The nine model compound and seven protein systems we have investigated thus far all originate from structures which occupy one of the six allowed χ_1, χ_2 regions of conformational space, as shown in Figure 7A. The least populated regions, based on known conformer libraries,[33,34] occur for $\chi_1 \approx 60°$, $\chi_2 \approx 90°$ and $\chi_1 \approx 60°$, $\chi_2 \approx -90°$. As shown in Figure 7, these species have the highest energies (in our model N-formyl tryptophan amide calculations) for both helical or sheet backbone geometries (Figure 7B,C), as can also be seen from the average energy results shown in Figure 7D (in which the lowest energy helical or sheet conformers were assigned energies of 0 kcal). The most populated region seen in Figure 7A ($\chi_1 \approx -60°$, $\chi_2 \approx 90°$) has an energy ∼1−2 kcal higher than the lowest value found in the model calculations, suggesting either a small deficiency in the calculations or the importance of interresidue interactions in real systems or both. In any case, the two high-energy conformers have $\chi_1 = 60°$. Now, from our chemical shift perspective, it is interesting to note that the $\chi_1 = 60°$, $\chi_2 = 90°$ conformer, which is quite rare, also happens to correspond to the most shielded Trp C^{γ} chemical shift, as shown in the experimental results in Figure 8A, the helical and sheet FTA model shift calculation results shown in Figure 8B,C, and the "averaged" (helix, sheet) chemical shift results shown in Figure 8D. For convenience, we have converted here the computed shieldings to computed shifts by using a value of 189.0 ppm, the calculated shielding we find for TMS (using the HF/CSGT method and a 6-311++G(2d,2p) basis set). The experimental result shown in Figure 8A for $\chi_1 = 60°$, $\chi_2 = 90°$ is for compound **1**, having a Trp C^{γ} shift of 104.1 ppm. Such shifts are only infrequently seen in proteins, consistent with the high energy of this conformer (Figure 7D). A peak at 105.6 ppm was reported for one of the four Trp residues in a *S. faecium* DHFR,[6] and although this peak was not unambiguously assigned, it is of

(33) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775−791.
(34) Lovell, S. C.; Word, M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389−408.

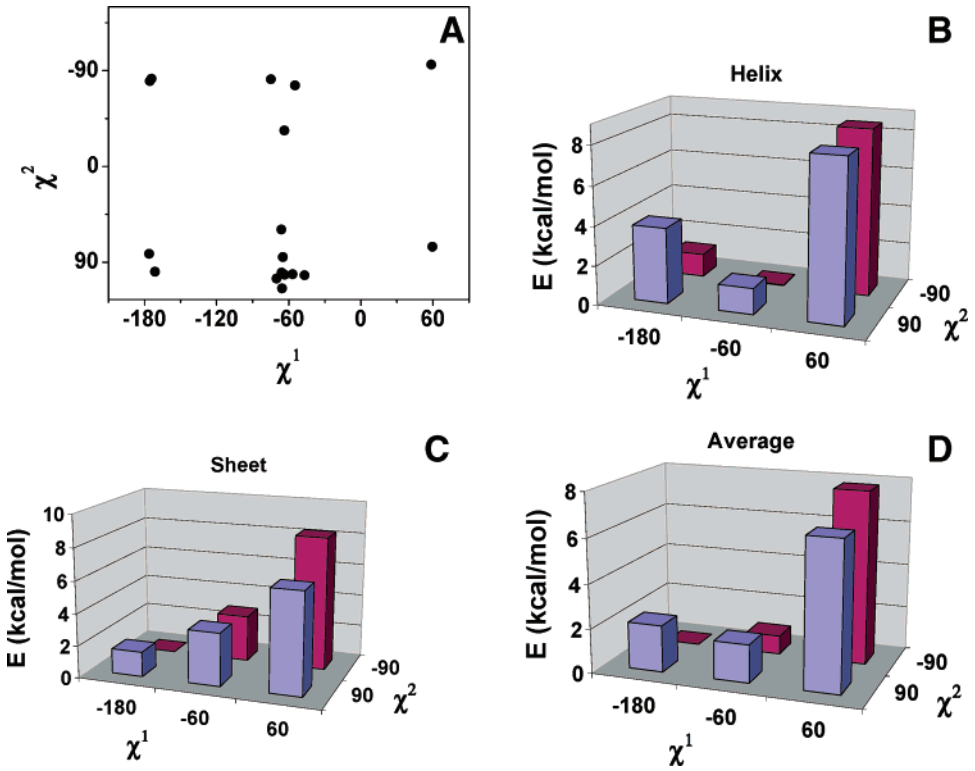


Figure 7. (A) Regions of χ_1, χ_2 space occupied by Trp residues in model systems, HEWL, and cytochrome *c*. (B) Calculated (HF/CSGT) energies for helical FTA models. (C) Calculated energies for sheet FTA residues. The lowest energies are set to zero. (D) Average energy differences, due to χ_1, χ_2 effects.

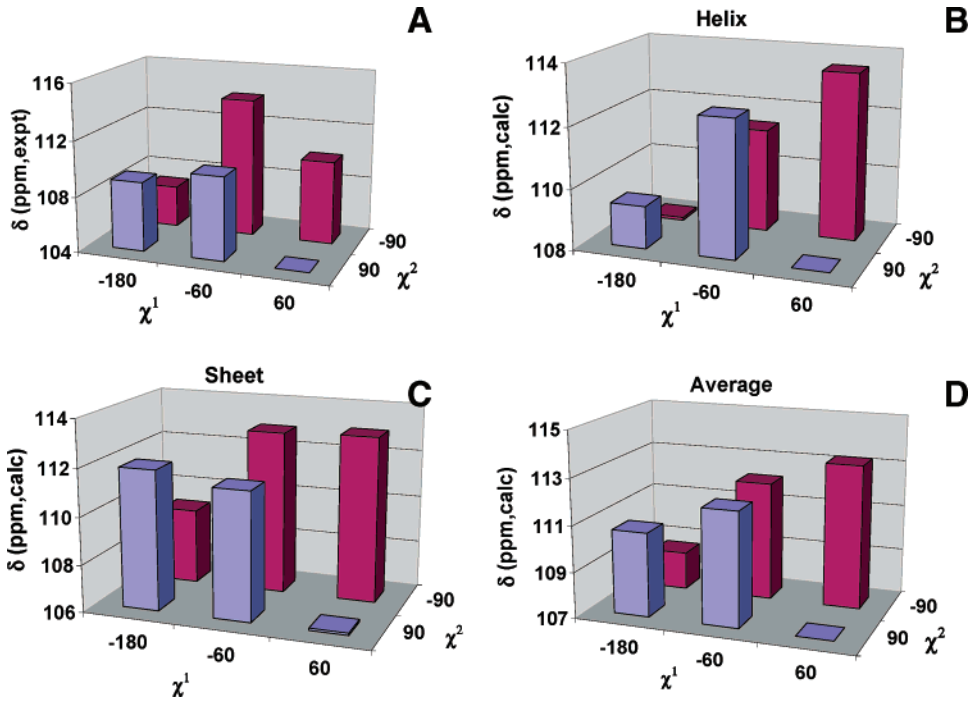


Figure 8. Experimental and computed Trp C$^\gamma$ chemical shifts as a function of χ_1 and χ_2. (A) Experimental chemical shifts for **1**−**12** and the seven Trp C$^\gamma$ chemical shifts in HEWL and cytochrome *c*. (B) Computed (HF/CSGT) chemical shifts for helical FTA species. (C) Computed chemical shifts for sheet FTA species. (D) Averaged helical, sheet chemical shifts, as a function of χ_1 and χ_2. The averaging is done here to facilitate comparison with the experimental results, which cover both helix and sheet geometries.

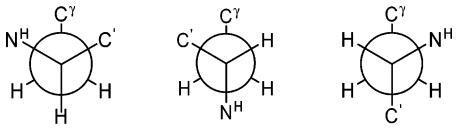
interest to note that at least two of the four Trps in DHFRs having known crystallographic structures have $\chi_1 = 60°$, $\chi_2 = 90°$.[35]

In the experimental results shown in Figure 8A, the next most shielded values are found for $\chi_1 = 180°$, $\chi_2 = -90°$, and this result is again reflected in the computed shifts, shown for example in Figure 8D. Likewise, the four most deshielded values found experimentally are in generally good (but not exact)

(35) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650−13662.

agreement with the average shift predictions shown in Figure 8D. The Trp C^{γ} chemical shifts show no obvious dependence on the backbone ϕ,ψ torsion angles (Figure 8B,C), but are strongly dependent on the side-chain χ_1, χ_2 torsion angles (Figure 8A,D), which implies that such correlations could have utility in structure determination or refinement. But what are the structural origins of these effects? Can they be related to well-known effects of structure on shielding?

One of the most well-known empirical correlations between structure and shielding is the so-called "γ-gauche" effect, discovered some 40 years ago by Grant and Lippmaa,[36,37] in which the presence of γ-gauche steric effects, the presence of large groups gamma to the site of interest and having torsion angles of $\pm 60°$, contribute to increased shielding:



In the case of tryphtophan, γ-gauche effects are expected with the heavy atoms C′ and N^{H}, but not with a γ-H. For the most shielded residues, a χ_1 torsion angle of 60° corresponds to a θ_1 torsion angle ($C^{\gamma}-C^{\beta}-C^{\alpha}-N^{H}$) of 60°, and it follows from symmetry that the other torsion angle (θ_2, $C^{\gamma}-C^{\beta}-C^{\alpha}-C'$) will be $-60°$. The average chemical shift for the residues investigated having these two γ-gauche interactions is 106.3 ± 2.6 ppm, and these two γ-gauche interactions are shown schematically above in **a**. For the next most shielded set of residues, $\theta_1 = \chi_1 = 180°$, and by symmetry $\theta_2 = 60°$, with the C′ group being in the gauche position, as shown in **b** above. For these residues, we find that $\delta_i = 108.0 \pm 1.3$ ppm. And finally, for the remaining 12 residues, $\theta_1 = \chi_1 = -60°$, $\theta_2 = 180°$, and $\delta_i = 111.6 \pm 2.1$ ppm. In this case, it is the N^{H} group which is gauche to C^{γ}, again as shown above in **c**. These results suggest that much of the chemical shift nonequivalence seen with Trp C^{γ} is due to the presence of γ-gauche interactions. The most shielded residues have two γ-gauche interactions (with C′, N^{H}), and the next most shielded has C′ in the γ-gauche position. However, these effects make up only about 40% of the observed chemical shift range and obviously ignore the effects of χ_2 (which correspond to a 180° indole ring flip). The presence of a significant χ_2 effect on shielding is of interest since this effect will be absent for Phe and Tyr, which may help explain their smaller C^{γ} chemical shift range in proteins,[38] although the increased mobility of these smaller residues may also be a factor.

Next, we consider the potential use of homology modeling, combined with quantum chemistry, to predict Trp C^{γ} chemical shifts. In early work,[7] we reported the Trp C^{γ} shifts of a series of carbonmonoxyheme proteins: human adult hemoglobin (HbCOA), human fetal hemoglobin (HbCOF), chicken AII hemoglobin (cHbCOAII), bovine fetal hemoglobin (bHbCOF), and horse myoglobin (hMbCO). There were a total of 15 Trp C^{γ} shifts observed, and these were all assigned based on

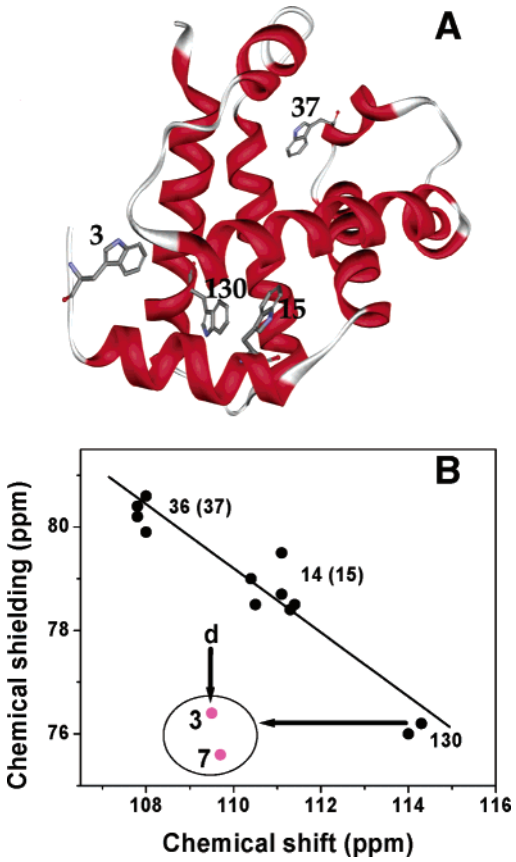


Figure 9. Homology modeling approach to Trp C^{γ} chemical shift prediction. (A) Chicken AII hemoglobin homology model[29] showing Trps 3β, 15β, 37β, and 130β. (B) Experimental shifts versus computed (HF/CSGT; FTA model) shieldings for the 15 Trp C^{γ} sites in the carbonmonoxyheme proteins discussed in the text. The "outliers" are thought to arise from the two terminal Trps (Trp^{3} in chicken, Trp^{7} in myoglobin) which appear at about the "random coil" shift value.

comparisons of the chemical shifts of these five proteins. For example, the C^{γ} of Trp residues 37β, 37γ, 37β, and 36β in HbCOA, HbCOF, cHbCOAII, and bHbCOF resonated at 107.8, 108.0, 108.1, and 107.8 ppm, while Trp 130, seen only in the HbCOF γ chains and chicken AII β chains, resonated at 114.3 and 114.6 ppm, respectively, enabling their specific assignments.[7]

We submitted the structures of HbCOF, cHbCOAII, and bHbCOF to the SWISS-MODEL website,[29] and a representative homology model prediction (for chicken AII hemoglobin, β chain) is shown in Figure 9A. We then used known X-ray structures[39,40] together with these homology model structures and the CSGT method on the FTA model to predict each of the 15 Trp C^{γ} chemical shieldings. The results obtained are shown in Table 4 and graphically in Figure 9B. The results are interesting in that all but two of the 15 points fall on the line shown, in which the most shielded residues have $\chi_1 \approx 60°$, $\chi_2 \approx 90°$, while those residues having $\chi_1 \approx 180°$, $\chi_2 \approx -90°$ (Trp^{14} α, Trp^{15} β chain) have intermediate shieldings, as expected based on the results shown in Figure 8. There are, however, two points which obviously fall off of the trend. These have $\chi_1 \approx -60°$, $\chi_2 \approx 90°$ and might be expected to be highly

(36) Lippmaa, E. P.; Past, J. *Eesti. NSV Tead. Akad. Toim. Fuus. Mat.* **1967**, *16*, 345.
(37) Cheney, B. V.; Grant, D. M. *J. Am. Chem. Soc.* **1967**, *89*, 5319−5327.
(38) Allerhand, A.; Norton, R. S.; Childers, R. F. *J. Biol. Chem.* **1977**, *252*, 1786−1794.
(39) Chu, K.; Vojtchovsky, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. *Nature* **2000**, *403*, 921−923.
(40) Park, S. Y.; Tame, J. R. H. Crystal Structure of Human Carbonmonoxy-Haemoglobin at 1.25 Å Resolution. PDB ID: 1IRD. Metalloprotein Database and Browser. http://metallo.scripps.edu/ (2001).

Table 4. Experimental Chemical Shifts and Computed Chemical Shieldings for 15 Trp C^{γ} Sites in Heme Proteins as a Function of Structure

protein	residue	δ_i (exptl)[a] (ppm)	σ_i (calcd)[b] (ppm)	χ_1[c]	χ_2[c]	ϕ[c]	ψ[c]
HbCOA	14α	111.1	79.5	−174.1	−99.5	−59.4	−36.9
	15β	111.1	79.5	−163.7	−102.1	−61.9	−29.6
	37β	107.8	80.6	64.3	86.1	−59.5	−24.1
HbCOF	14α	111.3	78.4	−174.6	−106.7	−72.9	−28.8
	15$^{\beta}$	110.5	78.5	−177.6	−103.8	−69.0	−30.6
	37$^{\beta}$	108.0	79.9	63.9	91.8	−66.0	−24.5
	130γ	114.3	76.2	−66.2	99.9	−97.2	−28.3
cHbCOAII	3β	109.5	76.4	−64.8	101.2	−94.5	109.1
	15β	110.4	79.0	−179.5	−101.7	−58.4	−37.1
	37β	108.1	80.4	63.4	87.9	−58.8	−25.1
	130β	114.6	76.0	−65.9	98.8	−77.5	−37.2
bHbCOF	14α	111.4	78.5	−164.3	−107.7	−64.7	−29.1
	36β	107.8	80.2	53.3	92.0	−58.9	−23.3
MbCO	7	109.7	75.6	−83.4	120.2	−67.5	−36.9
	14	111.1	78.7	−160.5	−106.0	−64.5	−26.7

[a] From ref 7. [b] Computed using the HF/CSGT method with the FTA model. [c] Torsion angle information obtained from X-ray structures (PDB Files 1DWR[39] (MbCO) and 1IRD[40] (HbCOA)) or from homology modeling (HbCOF, cHbCOAII, bHbCOF). In degrees.

deshielded (as with the Trp 130 β,γ sites). While there might be more than one explanation for these results, we note that the actual chemical shifts observed (∼109.3 ppm) are those seen for Trp C^{γ} in denatured or random coil proteins,[5] plus, the residues of interest are Trp^3 in cHbCOAII and Trp^7 in hMbCO. These two observations suggest the possibility that these residues, located near the N termini, may undergo enhanced motional averaging, and indeed as can be seen in Figure 9A, Trp^3 is in an unstructured region, suggesting the likelihood of enhanced mobility (in solution), resulting in substantial chemical shift averaging.

Finally, we consider the C^{γ} shielding tensor magnitudes and orientations. We used the FTA Trp model compound to compute the C^{γ} shielding tensor elements for the 9 model compounds plus 23 Trp residues in 7 proteins (HEWL, cytochrome *c*, hMbCO, and the four hemoglobins, excluding the two outliers shown in Figure 9) using the CSGT method. The results are shown in Table 5. As may be seen there, the σ_{33} shielding component changes by only a small amount (∼4 ppm) as σ_i becomes more deshielded, but in sharp contrast, σ_{11} and σ_{22} become much more deshielded, as may be seen in Figure 10. Some typical tensor orientations are illustrated in Figure 11 for the FTA models of compounds **1** and **12**. In both cases, which represent the most- and least-shielded Trp C^{γ} σ_i values, σ_{33} is

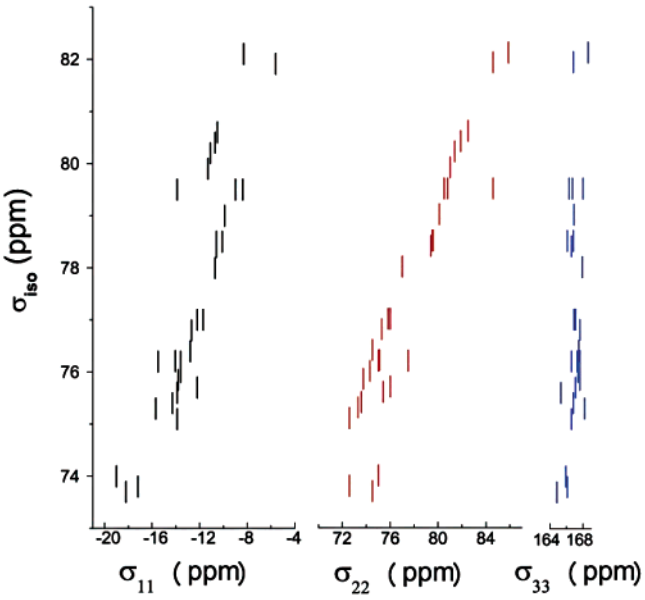


Figure 10. Trp C^{γ} chemical shielding tensor element magnitudes for model compounds and proteins, calculated by using the HF/CSGT method with FTA models.

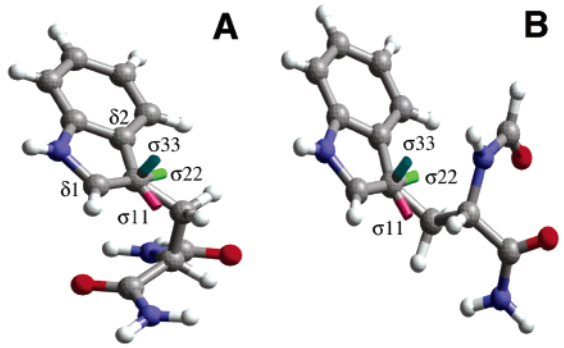


Figure 11. Tensor orientations for FTA models of compounds **1** (A) and **12** (B), calculated by using the HF/CSGT method.

Table 5. Computed Trp C^{γ} Chemical Shielding Tensor Element Magnitudes for Peptides and Proteins

compd	σ_{11}[a]	σ_{22}[a]	σ_{33}[a]	σ_{iso}[b]	δ_{iso}[c]	compd	σ_{11}[a]	σ_{22}[a]	σ_{33}[a]	σ_{iso}[b]	δ_{iso}[c]
1	−8.3	85.9	168.6	82.1	104.1	**10**	−12.2	76	167.1	77	110.4
4	−5.6	84.6	166.8	81.9	106.9	15β[g]	−9.9	80.1	166.9	79.0	110.4
36β[d]	−11.1	81.4	170.1	80.2	107.8	15$^{\beta}$[f]	−10.6	79.5	166.8	78.5	110.5
37β[e]	−10.5	82.5	169.8	80.6	107.8	62[h]	−14.0	75.1	167.6	76.2	110.8
37$^{\gamma}$[f]	−11.3	81	169.9	79.9	108.0	14α[e]	−9	80.8	166.7	79.5	111.1
37β[g]	−10.7	81.9	170	80.4	108.1	15β[e]	−8.4	80.5	166.3	79.5	111.1
28[h]	−13.8	73.7	167.6	75.8	108.7	14[i]	−15.5	77.5	166.6	78.7	111.1
111[h]	−13.9	84.6	168	79.5	108.7	14α[f]	−10.6	79.4	166.6	78.4	111.3
6	−12.7	75.3	167.6	76.8	108.7	14α[d]	−10.1	79.6	166.1	78.5	111.4
7	−12.2	76	167.1	75.7	109.4	123[h]	−14.3	73.6	166.8	75.4	111.9
3β[f]	−12.8	74.5	167.5	76.4	109.5	108[h]	−17.2	72.6	166.1	73.8	112.6
8	−11.7	75.8	166.9	77	109.7	**11**	−18.2	74.5	164.8	73.7	112.8
7[i]	−13.9	75.4	165.3	75.6	109.7	130γ[f]	−13.6	75	167.3	76.2	114.3
9	−10.7	77	167.9	78	110	130β[g]	−13.6	74.3	167.4	76.0	114.6
59[j]	−13.9	72.6	166.6	75.1	110.3	**12**	−19	75	165.9	74	115.5
63[h]	−15.7	73.3	168.2	75.3	110.3						

[a] Shielding tensor element computed by using the HF/CSGT method on an FTA model. In ppm. [b] Isotropic shielding computed by using the HF/CSGT method on an FTA model. [c] Experimental chemical shift (in ppm from TMS) obtained in this work (Table 2) or from refs 5−7 and 31. [d] Residue in homology model of bHbCOF. [e] Residue of HbCOA (PDB file 1IRD[40]). [f] Residue in homology model of HbCOF. [g] Residue in homology model of HbCOAII. [h] Residue in hen egg white lysozyme (PDB file 1BWI). [i] Residue in bMbCO (PDB file 1DWR[39]). [j] Residue in horse heart cytochrome *c* (PDB file 1HRC).

oriented perpendicular to the indole plane, while σ_{22} is approximately along the $C^{\gamma}-C^{\delta_1}$ bond vector and σ_{11} is of course orthogonal to σ_{33} and σ_{22}. Interestingly, even though the σ_{11} and σ_{22} tensor magnitudes change considerably (both by ~14 ppm) when all sites are investigated, there is virtually no change in their orientations between different structures. While this might at first appear somewhat surprising, we noted previously[24] that the C^{β} tensor orientations in most amino acids were essentially independent of the backbone ϕ,ψ torsion angles, even though the orientations of the C^{α} tensors were strongly dependent on ϕ,ψ. Thus, major changes in tensor orientations with structure appear to be observed primarily with C^{α} (which are attached to two peptide groups), but not with side-chain ^{13}C sites. The same trends are also seen in the CFP results for the 12 model systems (Supporting Information). The results shown in Figure 11 also clearly illustrate the χ_1 effects on shielding described above. In the model for **1** (Figure 11A), both C′ and N^{H} adopt the γ-gauche conformation with respect to C^{γ}, corresponding to the most shielded C^{γ} resonance observed to date. On the other hand, in the FTA model for **12** (Figure 11B), the only γ-gauche interaction is with the N^{H} group, which again as discussed above, we find results in the smallest shielding observed. Evidently, the largest effect on C^{γ} shielding arises from the proximity of the C′ carbonyl group. In **1**, the carbonyl oxygen is only ~3.2 Å from the Trp C^{γ}, suggesting the possibility of a π orbital overlap between the indole and CO groups, resulting in large shielding changes perpendicular to $\sigma_{33,}$ i.e., in the indole plane, consistent with the large effects seen for σ_{11} and σ_{22} (Figure 10).

Conclusions

The results we have presented above are of interest for a number of reasons. First, we have obtained the solid-state carbon-13 NMR chemical shifts of a series of Trp-containing peptides and model compounds. Second, we have obtained the single-crystal X-ray structures of four of these systems. Third, we have used quantum chemical methods to predict the ^{13}C NMR shieldings for these peptides, as well as several Trp residues in proteins. Overall, the theory versus experiment correlations for the Trp C^{γ}, C^{δ_2}, and C^{ϵ_2} carbons are good ($R = -0.98$, slope $= -1.17$) when using the HF/CSGT method and a uniform (6-311++G(2d,2p)) basis set. For C^{γ} alone, the correlation is $R = -0.85$ and the slope $= -0.86$. Results for C^{δ_2} (which has only a 3.3 ppm shift range) are worse, but for C^{ϵ_2}, incorporation of CFP again yields quite good correlations with experiment. Fourth, we investigated the relationships between the observed chemical shifts and structure. For C^{γ}, a clear correlation is seen with χ_1, χ_2 backbone torsion angles, an effect which is due at least in part to the presence of γ-gauche effects (with C′ and N^{H}), which contribute to shielding. The importance of χ_2 in Trp C^{γ} shielding may help explain the larger shift range seen experimentally with Trp vs Phe and Tyr $^{13}C^{\gamma}$ shifts. Fifth, we used homology modeling to enable the prediction of Trp C^{γ} shifts in several proteins which lack crystal structures. Sixth, we investigated computationally the C^{γ} shielding tensor element magnitudes and orientations. The changes in isotropic shift or shielding were found to be due to changes in σ_{11} and σ_{22}, not σ_{33}, and the tensor orientations were not found to change appreciably with structure, basically as noted previously for C^{β} in most amino acids (as a function of ϕ, ψ). Overall, the results we have described above give the first insights into the origins of the chemical shift nonequivalences of the nonprotonated aromatic carbons of tryptophan residues in proteins due to folding. Unlike C^{α} or C^{β} shifts, which depend primarily on ϕ, ψ, the large Trp C^{γ} shifts are correlated with χ_1, χ_2 effects. It should now be possible to extend this work to investigating the chemical shifts of other aromatic amino acids in proteins, including their use in structure determination.

Acknowledgment. We thank Scott Wilson for his help with the crystallography, Subhash Ghosh for synthesizing **10**, and Angel C. de Dios and Feng Cheng for their contributions. This work was supported by the United States Public Health Service (NIH Grant GM50694). The X-ray crystallographic facilities were supported by the National Science Foundation (Grant CHE 95-03145). Solid State NMR spectra were obtained in the Varian-Oxford Instruments Center for Excellence in NMR Laboratory. Funding for this instrumentation was obtained from the following sources: School of Chemical Sciences, Department of Chemistry, Department of Chemical Engineering, Department of Biochemistry, Beckman Institute, NIH, Keck Foundation, the Vice Chancellor for Academic Affairs, and the National Science Fundation.

Supporting Information Available: X-ray and shielding tensor data (PDF and CIF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA030612U