J|A|C|S
ARTICLES
Published on Web 11/01/2002

Investigation of the Neighboring Residue Effects on Protein Chemical Shifts

Yunjun Wang and Oleg Jardetzky*

Contribution from the Department of Molecular Pharmacology, Stanford University, Stanford, California 94305-5174

Received May 7, 2002. Revised Manuscript Received August 16, 2002

Abstract: In this study, we report nearest neighbor residue effects statistically determined from a chemical shift database. For an amino acid sequence XYZ, we define two correction factors, $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$, representing the effects on Y's chemical shifts from the preceding residue (X) and the following residue (Z), respectively, where X, Y, and Z are any of the 20 naturally occurring amino acids, n stands for $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ nuclei, and s represents the three secondary structural types β-strand, random coil, and α-helix. A total of ~14400 $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$, representing nearly all combinations of X, Y, Z, n, and s, have been quantitatively determined. Our approach overcomes the limits of earlier experimental methods using short model peptides, and the resulting correction factors have important applications such as chemical shift prediction for the folded proteins. More importantly, we have found, for the first time, a linear correlation between the $\Delta(^{X}Y)n,s$ ($n = ^{15}N$) and the $^{13}C^{\alpha}$ chemical shifts of the preceding residue X. Since $^{13}C^{\alpha}$ chemical shifts of the 20 amino acids, which span a wide range of 40−70 ppm, are largely dominated by one property, the electron density of the side chain, the correlation indicates that the same property is responsible for the effect on the following residue. The influence of the secondary structure on both the chemical shifts and the nearest neighbor residue effect are also investigated.

Introduction

Chemical shifts of amino acid groups in proteins reflect the primary, secondary, and tertiary structure of the protein. The structural effects permit the identification of chemically identical, but positionally nonequivalent, amino acid residues and thus form the basis of all structure determination by nuclear magnetic resonance (NMR).[1] Although the structural effects on the chemical shift have been known for a long time,[2−6] the theoretical interpretation of chemical shifts has never been sufficiently accurate to allow the information contained in the shift itself to be used for purposes of structure determination. Only in recent years, with the accumulation of large chemical shift databases, has it become possible to begin to decipher this information using statistical analysis and to attempt to use it for refinement of NMR structures.[7−10]

Attempts to define individual contributions to the overall chemical shifts and to introduce appropriate correction factors for structural effects have thus far met with limited success. Thus, ^{15}N chemical shifts, which cover a range of 40 ppm, can only be predicted with an error of about 3 ppm.[11,12] The separate identification of secondary structure and nearest neighbor effects has proved to be particularly difficult. Reasonable values for the effects of secondary structure on the chemical shift have been obtained by statistical analysis of a large number of structures[13] and the use of an empirical shielding surface.[14] Contributions of specific structural features, such as the helix capping box[15] and the β-hairpin,[16] have also been identified, and a correlation of the chemical shift with ϕ and ψ angles has yielded useful limits to Ramachandran angle constraints that can be derived from chemical shift data.[17,18] Less successful have been efforts to correct for nearest neighbor effects. These are typically estimated from studies of short peptides (e.g., AcGGXGG-HN_2 under denaturing experimental conditions).[19−23]

* Address correspondence to this author. E-mail: jardetzky@stanford.edu.

(1) Roberts, G. C.; Jardetzky, O. *Adv. Protein Chem.* **1970**, *24*, 447−545.
(2) Nakamura, A.; Jardetzky, O. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 2212−2219.
(3) Nakamura, A.; Jardetzky, O. *Biochemistry* **1968**, *7*, 1226−1230.
(4) Markley, J. L.; Meadows, D. H.; Jardetzky, O. *J. Mol. Biol.* **1967**, *27*, 25−40.
(5) McDonald, C. C.; Phillips, W. D. *J. Am. Chem. Soc.* **1967**, *89*, 6332−6341.
(6) Cohen, J. S.; Jardetzky, O. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *60*, 92−99.
(7) Szilagyi, L.; Jardetzky, O. *J. Magn. Reson.* **1989**, *83*, 441−449.
(8) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647−1651.
(9) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171−180.
(10) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1995**, *B 107*, 293−297.
(11) Le, H.; Oldfield, E. *J. Biomol. NMR* **1994**, *4*, 341−348.
(12) Wishart, D. S.; Nip, A. M. *Biochem. Cell Biol.* **1998**, *76*, 1−10.
(13) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311−333.
(14) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490−5492.
(15) Gronenborn, A. M.; Clore, G. M. *J. Biomol. NMR* **1994**, *4*, 455−458.
(16) Santiveri, C. M.; Rico, M.; Jimenez, M. A. *J. Biomol. NMR* **2001**, *19*, 331−345.
(17) Beger, R. D.; Bolton, P. H. *J. Biomol. NMR* **1997**, *10*, 129−142.
(18) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289−302.
(19) Kricheldorf, H. R. *Org. Magn. Reson.* **1981**, *15*, 162−177.
(20) Glushka, J.; Lee, M.; Coffin, S.; Cowburn, D. *J. Am. Chem. Soc.* **1989**, *111*, 7716−7722.
(21) Braun, D.; Wider, G.; Wuthrich, K. *J. Am. Chem. Soc.* **1994**, *116*, 8466−8469.
(22) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67−81.
(23) Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J. Am. Chem. Soc.* **2001**, *123*, 2970−2978.

Such studies suffer from a number of serious drawbacks: a relatively small database, since an experimental study of all the required 8000 tripeptide sequences is prohibitive; the use of denaturing solvents, which are likely to have selective effects on the chemical shift; and the limitation to the random coil, which neglects possible variation of the nearest neighbor effect with secondary structure. Only one study of the nearest neighbor effect, using an empirical chemical shift database, has thus far been reported, and it is limited to $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts.[24]

In the present study we have established a large empirical database for $^{1}H^{\alpha}$, ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts. In the preceding paper[25] we have reported analysis of secondary structure effects. In the present paper we present the analysis of nearest neighbor effects for all 20 amino acids in different secondary structures.

Methods

Nomenclature and Definitions. For an amino acid sequence XYZ, we propose that the observed chemical shift of *Y*, $\delta n,s$, is composed of the following components:

$$\delta n,s = \delta n,coil + \Delta\delta n,s + \Delta(^{X}Y)n,s + \Delta(Y^{Z})n,s \quad (1)$$

where n represents $^{1}H^{\alpha}$, ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ nuclei, s stands for the three secondary structural types, β-strand, random coil, and α-helix (hereinafter they are abbreviated as beta, coil, and helix, respectively), and X, Y, and Z are any of the 20 common amino acids. δn,coil is the chemical shift of Y at random coil state; $\Delta\delta$n,s is the secondary structural effect on Y's chemical shift, which is zero by definition when s = coil; $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s are the effects from the preceding residue (X) and the following residue (Z), respectively. We further express $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s as:

$$\Delta(^{X}Y)n,s = \Delta(^{X}Y)n,coil + \Delta\Delta(^{X}Y)n,s \quad (2)$$

$$\Delta(Y^{Z})n,s = \Delta(Y^{Z})n,coil + \Delta\Delta(Y^{Z})n,s \quad (3)$$

where $\Delta(^{X}Y)$n,coil and $\Delta(Y^{Z})$n,coil represent the neighbor residue effects on the random coil state, and $\Delta\Delta(^{X}Y)$n,s and $\Delta\Delta(Y^{Z})$n,s represent the variations of the nearest neighbor effects with secondary structure. Again, by definition $\Delta\Delta(^{X}Y)$n,s and $\Delta\Delta(Y^{Z})$n,s equal to zero when s = coil. Thus the observed chemical shift, δn,s, can be expressed as:

$$\delta n,s = \delta n,coil + \Delta(^{X}Y)n,coil + \Delta(Y^{Z})n,coil + \Delta\delta n,s + \Delta\Delta(^{X}Y)n,s + \Delta\Delta(Y^{Z})n,s \quad (4)$$

Preparation of Chemical Shift Database. A database which contains $^{1}H^{\alpha}$, ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts, backbone (ϕ, ψ, and ω) torsion angles, backbone hydrogen bonds, and secondary structure was created. The chemical shifts were downloaded from BioMagResBank (http:// www.bmrb-.wisc.edu/bmrb), and those meeting the following criteria were selected: (1) the length of protein sequence >60; (2) the most commonly used materials, DSS, TMS, TSP, and liquid NH_3 were used as either direct or indirect reference for ^{1}H, ^{15}N, and ^{13}C chemical shifts;[26] (3) for a small fraction of proteins, H_2O was used as reference for ^{1}H chemical shift. For each of these proteins, the average $^{1}H^{\alpha}$ chemical shifts in either α-helix, β-strand, or both were calculated and compared with that derived from the bulk of the data. If no notable discrepancy (e.g., >0.05 ppm) was found, the data were included. When several BMRB entries were available for the same protein, priority was given to the one with the most complete assignments. Abnormal chemical shift assignments were thoroughly checked; many of these were found to be obvious typing errors (e.g., 8.7 ppm for an ^{15}N shift). A total of 112 such unusual assignments were identified and removed from the database. The assignments of the very first N-terminal and last C-terminal residues of each protein were also excluded to avoid possible terminal effects on chemical shifts. In total, 232 530 chemical shifts (42 815, 28 946, 43 785, 28 754, 44 171, and 44 059 for ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{N}$, and $^{1}H^{\alpha}$, respectively) derived from 415 distinct, nonparamagnetic proteins were collected into our database. The three-dimensional (3D) coordinates of 326 out of 415 proteins were found to be available and were downloaded from Protein Data Bank (http://www.rcsb.org/pdb). As described in our preceding study,[25] the secondary structures were determined from the 3D coordinates using the program DSSP[27] and VADAR.[28] The backbone dihedral angles and hydrogen bond were determined using VADAR. For the remaining 89 proteins without 3d coordinates, the secondary structures were identified from combined chemical shift data using our program PSSI.[25]

Determination of $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s. For each of the nuclei n, chemical shifts were categorized into three separate groups: β-strand, random coil, and α-helix, based on the secondary structure of Y. $\Delta(^{X}Y)$n,s for each of the three secondary structural types was calculated for each of the 20 amino acids of X and Y:

$$\Delta(^{X}Y)n,s = \langle\delta n,s(X)\rangle - \langle\delta n,s(w/o\ X)\rangle \quad (5)$$

where $\langle\delta$n,s(X)$\rangle$ and $\langle\delta$n,s(w/o X)$\rangle$ represent the averaged chemical shift of amino acid Y with and without amino acid X at the preceding position, respectively. $\Delta(Y^{Z})$n,s was determined a similar way.

$$\Delta(Y^{Z})n,s = \langle\delta n,s(Z)\rangle - \langle\delta n,s(w/o\ Z)\rangle \quad (6)$$

In this study, we have investigated the neighbor residue effects for each of the 20 amino acids (both the neighboring residue and the residue under study) and for each of the three secondary structural types. To describe them, 1200 correction factors $\Delta(^{X}Y)$n,s (same for the $\Delta(Y^{Z})$n,s) are needed for each of the six nuclei. Critical to the present approach are the following items: First, a sufficient number of chemical shifts must be available for statistical analysis. On average, the sample size in this study is statistically adequate. For example, for $^{13}C'$ nuclei, which has the least number of chemical shifts in our database, the averaged numbers of chemical shifts in the data sets of $\langle\delta$n,s(X)$\rangle$ and $\langle\delta$n,s(w/o X)$\rangle$ are 25 and 456, respectively. For certain amino acids such as Trp and Cys with low natural occurrence, the number of available chemical shifts is relatively small. Second, the difference between sets of chemical shifts $\langle\delta$n,s(X)$\rangle$ and $\langle\delta$n,s(w/o X)$\rangle$ must be large enough to be statistically

(24) Iwadate, M.; Asakura, T.; Williamson, M. P. *J. Biomol. NMR* **1999**, *13*, 199−211.
(25) Wang, Y.; Jardetzky, O. *Protein Sci.* **2002**, *11*, 852−861.
(26) Wishart, D. S.; Bigam, C. G.; Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., Sykes, B. D. *J. Biomol. NMR* **1995**, *6*, 135−140.
(27) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(28) Wishart, D. S.; Willard, L.; Sykes, B. D. VADAR 1.1 University of Alberta, 1995.

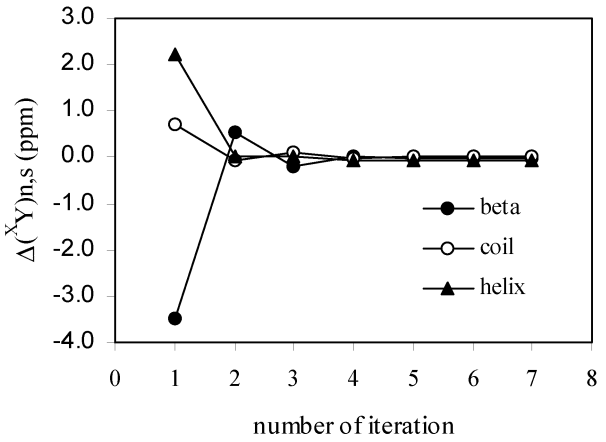

Figure 1. $\Delta(^{X}Y)n,s$ (n = ^{15}N, X = Gly, Y = Ile) converge to zero during the iterations.

significant. The minimum value of $\langle\delta n,s(X)\rangle - \langle\delta n,s(w/o\ X)\rangle$ and $\langle\delta n,s(Z)\rangle - \langle\delta n,s(w/o\ Z)\rangle$ assessed by a Student's *t*-test at level of $p = 0.05$ are 1.20, 0.60, 0.50, 0.60, 0.20, and 0.10 ppm for ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{N}$, and $^{1}H^{\alpha}$, respectively. As we will show later, the preceding residue effect on ^{15}N shift, which spans a range of nearly 6 ppm, and the effect on $^{13}C'$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ shifts from the following Pro exceed the above criteria by far. Finally, and most importantly, the interference of effects other than those from the neighboring residue can be effectively limited. For influences other than that of the neighboring residues, we assume that they would have similar distributions on the two sets of data, $\langle\delta n,s(X)\rangle$ and $\langle\delta n,s(w/o\ X)\rangle$, and could be effectively reduced by canceling each other during the calculation. To test this assumption, the backbone dihedral angle distribution and the percentage of the hydrogen bond were investigated for several pairs of arbitrarily selected $\langle\delta n,s(X)\rangle$ and $\langle\delta n,s(w/o\ X)\rangle$ data sets. Both the dihedral angle distribution and the percentage of hydrogen bond are indeed very close (e.g., the differences in the averaged ϕ and ψ dihedral angles and the percentage of the hydrogen bond are less than 5%) for each pair of data (data not shown). In addition, separation of the chemical shifts based on the three secondary structural types helps to efficiently limit conformational effects during the calculation of $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$. Specifically, the α-helix has a very narrow range of backbone dihedral angles and a very regular hydrogen bond pattern. As a consequence, the backbone dihedral angle distribution and the percentage of the hydrogen bond in the two data sets, $\langle\delta n,helix\ (X)\rangle$ and $\langle\delta n,helix\ (w/o\ X)\rangle$, show a very high degree of similarity.

Iterative calculations were performed to eliminate the interference between the effects from the preceding and the following residues during the calculation. During the iterations, each of the observed chemical shifts was corrected using the $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ obtained from the previous step using the following equation:

$$\delta n,s(corrected) = \delta n,s(observed) - \Delta(^{X}Y)n,s - \Delta(Y^{Z})n,s \quad (7)$$

A new set of correction factors, $\Delta(^{X}Y)*n,s$ and $\Delta(Y^{Z})*n,s$, were then recalculated from the corrected chemical shifts, $\delta n,s(corrected)$. In total, seven iterations were carried out. As shown in Figure 1, during the iterations, each of the $\Delta(^{X}Y)*n,s$ and $\Delta(Y^{Z})*n,s$ values quickly converged to zero, and thus the corrected chemical shift, $\delta n,s(corrected)$, approached to be a constant value (data not shown). The $\Delta(^{X}Y)*n,s$ and $\Delta(Y^{Z})*n,s$ obtained in each iteration step were added together as final results of $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$. Without the iterations, on average the $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ values can change by ~20%. In total, 14 400 correction factors, $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ were calculated and tabulated together as a "dictionary", which is presented as Supporting Information. A JAVA user interface program, which automatically calculates the neighboring residue effects $\Delta(^{X}Y)n,s + \Delta(Y^{Z})n,s$ from protein sequence, has been developed. This program is free and can be obtained by sending email to one of the authors (yunjunwang@yahoo.com).

Determination of $\langle\Delta(^{X}Y)n,s\rangle$ and $\langle\Delta(Y^{Z})n,s\rangle$. The weighted average of $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$ over all the 20 amino acids of Y, defined as $\langle\Delta(^{X}Y)n,s\rangle$ and $\langle\Delta(Y^{Z})n,s\rangle$ respectively, were calculated for each amino acid type of the neighboring residue (X and Z) by:

$$\langle\Delta(^{X}Y)n,s\rangle = \sum N(X)*\Delta(^{X}Y)n,s/\sum N(X)$$

where $N(x)$ is the number of chemical shifts of Y preceded by X, $\langle\Delta(Y^{Z})n,s\rangle$ was calculated in a similar way.

Determination of $\langle\Delta\delta n,s\rangle$, $\langle\Delta\Delta(^{X}Y)n,s\rangle$, and $\langle\Delta\Delta(Y^{Z})n,s\rangle$. The averaged secondary structural effect, $\langle\Delta\delta n,s\rangle$, was obtained by calculating *Y*'s chemical shift difference between the β-strand (or α-helix) and the random coil, i.e.,

$$\langle\Delta\delta n,beta\rangle = \langle\delta n,beta(corrected)\rangle - \langle\delta n,coil(corrected)\rangle$$

The averaged variation of the neighbor residue effect with secondary structures, defined as $\langle\Delta\Delta(^{X}Y)n,s\rangle$, was obtained by calculating the difference between $\langle\Delta(^{X}Y)n,s\rangle$ and $\langle\Delta(^{X}Y)n,coil\rangle$, i.e.,

$$\langle\Delta\Delta(^{X}Y)n,beta\rangle = \langle\Delta(^{X}Y)n,beta\rangle - \langle\Delta(^{X}Y)n,coil\rangle$$

All the calculations and data manipulations were accomplished using a series of JAVA programs coded by one of the authors (Y. J. Wang).

Results and Discussion

$\langle\delta n,coil\rangle$ and $\langle\Delta\delta n,s\rangle$. Since the random coil chemical shifts are usually used as reference to identify the conformational change, they are of special interest to NMR spectroscopists. The statistics of the corrected random coil chemical shifts (averaged value, $\langle\delta n,coil\rangle$, together with standard deviation) for the 20 amino acids are listed in Table 1. The comparison of the present random coil chemical shifts with those reported in our earlier study[25] shows an excellent agreement between the two sets of data. Slight differences for certain amino acids are simply due to the different chemical shift database used. In this study, the overall averaged chemical shifts, which were used as reference during the calculation of $\Delta(^{X}Y)n,s$ and $\Delta(Y^{Z})n,s$, remained constant after correction for the neighbor residue effect. However, the correction does bring about two important changes. First, after the correction, on average, the standard deviations of chemical shift distribution drop by 0.55, 0.22, 0.25, 0.11, 0.11, and 0.07 ppm for ^{15}N, ^{13}C, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{N}$, and $^{1}H^{\alpha}$, respectively. Second, and more importantly, after correction for the neighbor residue effect, the secondary structural effects on chemical shifts were clearly revealed. It is well-known that chemical shifts are influenced by secondary structures. However,

Table 1. Averaged Random Coil Chemical Shift,[a] ⟨δn,Coil⟩, and the Secondary Structure Factors, ⟨Δδn,beta⟩ and ⟨Δδn,helix⟩[b]

amino acids		^{15}N			^{13}C′			^{13}Cα		
		⟨$\delta_{n,coil}$⟩	⟨$\Delta\delta_{beta}$⟩	⟨$\Delta\delta_{helix}$⟩	⟨$\delta_{n,coil}$⟩	⟨$\Delta\delta_{beta}$⟩	⟨$\Delta\delta_{helix}$⟩	⟨$\delta_{n,coil}$⟩	⟨$\Delta\delta_{beta}$⟩	⟨$\Delta\delta_{helix}$⟩
I	I	120.58(4.45)	2.90(4.07)	0.21(2.32)	175.52(1.34)	−0.71(1.42)	2.03(1.27)	60.79(1.58)	−0.99(1.40)	3.77(1.61)
	V	119.91(4.91)	2.61(4.40)	0.47(2.48)	175.66(1.35)	−0.96(1.45)	1.90(1.35)	62.00(1.93)	−1.43(1.55)	4.22(1.35)
II	D	120.37(3.77)	2.86(3.74)	−0.99(2.20)	176.00(1.23)	−0.53(1.36)	2.16(1.25)	54.00(1.49)	−0.41(1.44)	2.78(1.28)
	N	118.50(4.06)	3.24(3.67)	−0.81(2.22)	174.84(1.40)	−0.68(1.35)	2.09(1.40)	53.00(1.41)	−0.69(1.24)	2.41(1.28)
III	F	119.72(4.03)	1.83(3.99)	−0.13(2.74)	175.46(1.62)	−1.18(1.71)	1.60(1.29)	57.46(1.84)	−1.18(1.36)	3.52(1.69)
	H	118.92(3.44)	2.19(4.14)	−0.79(2.39)	174.78(1.49)	−0.70(1.33)	2.21(1.16)	55.74(1.57)	−1.01(1.60)	2.99(1.40)
	W	120.99(3.55)	2.24(3.99)	−0.57(2.03)	175.87(1.04)	−0.57(1.46)	1.93(1.14)	57.54(1.56)	−1.34(1.43)	2.39(1.34)
	Y	119.37(4.06)	2.84(4.12)	−0.19(2.60)	175.29(1.36)	−0.83(1.57)	1.85(1.32)	57.64(1.87)	−1.07(1.41)	3.45(1.43)
IV	K	121.10(3.69)	1.86(3.76)	−1.63(2.33)	176.15(1.34)	−1.10(1.29)	2.17(1.44)	56.29(1.59)	−1.27(1.25)	2.49(1.40)
	L	121.57(3.77)	3.36(3.51)	−1.74(2.32)	176.70(1.45)	−1.26(1.33)	1.80(1.33)	54.77(1.54)	−1.06(1.22)	2.60(1.17)
	M	120.14(3.54)	1.90(3.23)	−1.87(2.11)	175.94(1.23)	−1.26(1.27)	1.81(1.28)	55.43(1.49)	−1.28(1.06)	2.57(1.62)
	Q	119.82(3.46)	2.16(3.26)	−1.45(2.27)	175.75(1.23)	−1.09(1.03)	2.60(1.29)	55.89(1.52)	−1.28(1.17)	2.60(1.19)
	R	120.75(3.86)	1.75(3.61)	−1.72(2.27)	176.01(1.30)	−1.03(1.31)	2.22(1.34)	56.18(1.70)	−1.80(1.48)	2.82(1.33)
	E	120.62(3.43)	2.02(3.49)	−1.38(2.44)	176.32(1.28)	−1.20(1.23)	2.37(1.19)	56.66(1.57)	−1.49(1.33)	2.39(1.16)
V	T	113.88(4.53)	3.62(4.52)	1.39(3.46)	174.78(1.44)	−1.44(1.36)	1.44(1.18)	61.30(1.77)	−0.39(1.53)	4.48(1.80)
	C	118.10(3.22)	3.00(3.53)	0.57(2.90)	175.11(1.22)	−0.95(1.54)	1.85(0.97)	58.24(1.99)	−1.21(1.99)	3.96(1.94)
	S	116.00(3.62)	1.30(3.53)	−0.75(2.31)	174.41(1.32)	−0.91(1.52)	1.61(1.26)	58.20(1.64)	−0.99(1.35)	2.77(1.36)
others	A	123.82(3.22)	1.19(3.79)	−2.04(2.12)	177.28(1.36)	−1.52(1.31)	1.95(1.47)	52.46(1.40)	−1.52(1.27)	2.18(1.05)
	G	109.48(3.27)	0.27(3.50)	−2.03(2.24)	174.01(1.58)	−1.29(2.02)	1.85(1.30)	45.28(1.10)	−0.44(1.18)	1.57(0.93)
	P				176.62(1.39)	−0.80(1.32)	1.70(1.02)	63.24(1.24)	−0.80(1.06)	2.61(0.94)

amino acids		^{13}Cβ			^{1}H^{N}			^{1}Hα		
		<$\delta_{n,coil}$>	<$\Delta\delta_{beta}$⟩	<$\Delta\delta_{helix}$⟩	<$\delta_{n,coil}$>	<$\Delta\delta_{beta}$⟩	<$\Delta\delta_{helix}$⟩	<$\delta_{n,coil}$>	<$\Delta\delta_{beta}$⟩	<$\Delta\delta_{helix}$⟩
I	I	38.43(1.67)	1.29(1.96)	−0.73(1.09)	7.94(0.65)	0.74(0.61)	0.10(0.50)	4.18(0.33)	0.50(0.40)	−0.51(0.26)
	V	32.35(1.38)	1.37(1.71)	−0.93(0.79)	7.98(0.62)	0.67(0.57)	0.04(0.53)	4.13(0.36)	0.47(0.45)	−0.57(0.28)
II	D	40.78(1.32)	1.20(1.64)	−0.41(1.19)	8.31(0.56)	0.24(0.54)	−0.11(0.50)	4.62(0.24)	0.30(0.34)	−0.20(0.18)
	N	38.43(1.48)	1.66(1.87)	0.00(1.08)	8.35(0.64)	0.25(0.55)	−0.16(0.52)	4.66(0.30)	0.51(0.40)	−0.14(0.20)
III	F	39.41(1.78)	2.11(1.85)	−0.51(1.26)	8.09(0.68)	0.67(0.62)	0.14(0.58)	4.59(0.37)	0.50(0.45)	−0.45(0.36)
	H	29.50(1.77)	2.50(2.03)	0.09(1.48)	8.18(0.65)	0.46(0.62)	−0.16(0.47)	4.60(0.31)	0.46(0.43)	−0.17(0.25)
	W	29.60(1.16)	1.65(1.25)	−0.69(0.96)	7.97(0.63)	0.70(0.67)	0.18(0.54)	4.60(0.27)	0.46(0.30)	−0.31(0.28)
	Y	38.78(1.71)	1.95(1.78)	−0.56(1.11)	7.99(0.64)	0.82(0.66)	0.12(0.56)	4.56(0.39)	0.50(0.46)	−0.39(0.28)
IV	K	32.53(1.47)	1.81(1.63)	−0.40(0.97)	8.17(0.54)	0.29(0.58)	−0.21(0.51)	4.28(0.27)	0.44(0.39)	−0.27(0.22)
	L	42.14(1.45)	1.72(1.87)	−0.56(1.07)	8.06(0.62)	0.60(0.61)	0.02(0.51)	4.36(0.31)	0.48(0.41)	−0.36(0.25)
	M	32.92(2.04)	1.64(2.19)	−0.81(1.51)	8.22(0.53)	0.32(0.53)	−0.14(0.47)	4.47(0.31)	0.49(0.37)	−0.37(0.30)
	Q	29.01(1.52)	2.22(1.69)	−0.74(0.92)	8.20(0.58)	0.32(0.56)	−0.05(0.46)	4.29(0.29)	0.50(0.41)	−0.26(0.22)
	R	30.36(1.53)	1.97(1.74)	−0.54(0.91)	8.21(0.59)	0.33(0.56)	−0.13(0.47)	4.26(0.30)	0.53(0.44)	−0.27(0.25)
	E	29.87(1.38)	2.14(1.89)	−0.68(0.97)	8.36(0.53)	0.23(0.53)	−0.12(0.55)	4.28(0.26)	0.44(0.37)	−0.30(0.19)
V	T	68.92(0.90)	0.59(0.75)	−0.39(0.76)	8.16(0.61)	0.37(0.53)	−0.14(0.42)	4.44(0.32)	0.42(0.40)	−0.44(0.25)
	C	29.54(1.94)	0.40(1.60)	−1.76(1.49)	8.10(0.59)	0.70(0.51)	−0.05(0.58)	4.59(0.34)	0.40(0.45)	−0.48(0.26)
	S	63.75(1.32)	1.18(1.52)	−0.90(0.92)	8.22(0.61)	0.32(0.57)	−0.07(0.43)	4.45(0.28)	0.46(0.44)	−0.21(0.22)
others	A	18.98(1.20)	1.98(1.83)	−0.86(0.96)	8.09(0.54)	0.46(0.62)	−0.02(0.50)	4.31(0.28)	0.49(0.43)	−0.25(0.24)
	G				8.37(0.61)	0.04(0.73)	−0.13(0.55)	3.97(0.20)	0.00(0.31)	−0.10(0.17)
	P	31.81(0.97)	0.23(1.17)	−0.51(0.79)				4.41(0.25)	0.19(0.41)	−0.23(0.25)

[a] After correction for the neighboring residue effects. [b] Standard deviation is in parentheses.

due to the interference of the neighboring residue effect, the secondary structural effects usually remain ambiguous for the observed ^{15}N chemical shifts. As a representative example, plotted in Figure 2 are the averaged net structural chemical shifts of Ala preceded by each of the 20 amino acids. As shown Figure 2a, before neighboring residue effect correction, the net secondary structural shifts (⟨δn,beta⟩ − ⟨δn,coil⟩ and ⟨δn,helix⟩ − ⟨δn,coil⟩) vary greatly with the preceding residue. After correction for the neighbor residue effect, they become constant (totally independent of the preceding residue).

As shown in Table 1, five groups of amino acids can be distinguished among the 20 amino acids based on the similarity of their side chain structures. Group I contains Val and Ile, which both have branched side chains at their C$^{\beta}$ atom. Group II contains Asp and Asn, which both have similar functional groups (COOH and CONH$_2$) connected the C$^{\beta}$ atom. Group III contains His, Phe, Trp, and Tyr, all of which have an aromatic ring as the side chain. Group IV contains Leu, Met, Arg, Lys, Gln, and Glu, whose C$^{\beta}$ atoms are connected to either CH$_2$ or CH groups. Group V contains Thr, Cys, and Ser, all of which have a similar functional group SH or OH connected to the C$^{\beta}$ atom. As also shown in Table 1, in each of the above groups, the ⟨Δδn,s⟩ values are close to each other with few outliers, indicating a correlation between ⟨Δδn,s⟩ and side chain property of Y. As we will discuss later, such correlations also exist between ⟨ΔΔ(^{X}Y)n,s⟩ (n = ^{15}N) and the side chain property of X.

⟨Δ(^{X}Y)n,s⟩ and ⟨Δ(Y^{Z})n,s⟩. To generally evaluate the ability of an amino acid to influence it's neighboring residue's chemical shift, the averaged correction factors, ⟨Δ(^{X}Y)n,s⟩ and ⟨Δ(Y^{Z})n,s⟩, over the 20 amino acids of Y are calculated and listed in Table 2. This table shows the following: (1) The preceding residue has a particularly large effect on ^{15}N chemical shift. On average, the magnitudes of the preceding residue effects are in the order of ^{15}N ≫ ^{13}C′ ∼ ^{13}Cα ∼ ^{13}Cβ > ^{1}HN > ^{1}Hα. (2) The following residue has insignificant effects except for proline, which has a remarkable influence on ^{13}C′, ^{13}Cα, and ^{13}Cβ chemical shifts of the preceding residue. On average, the magnitudes of the following residue effects are in the order of ^{15}N > ^{13}C′ > ^{13}Cα > ^{13}Cβ > ^{1}H^{N} > ^{1}Hα. (3) ⟨Δ(^{X}Y)n,s⟩ and

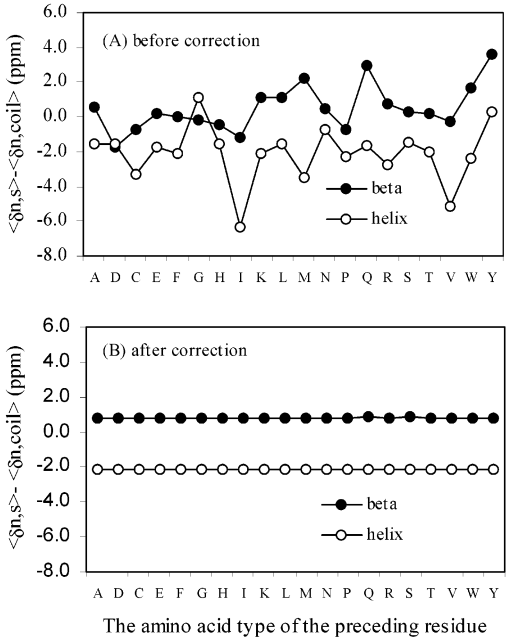

Figure 2. The averaged ^{15}N structural chemical shift, $\langle\delta n,s\rangle - \langle\delta n,coil\rangle$ (s = beta, helix), of Ala when preceded by each of the 20 amino acids before (A) and after (B) the correction for the neighbor residue effects.

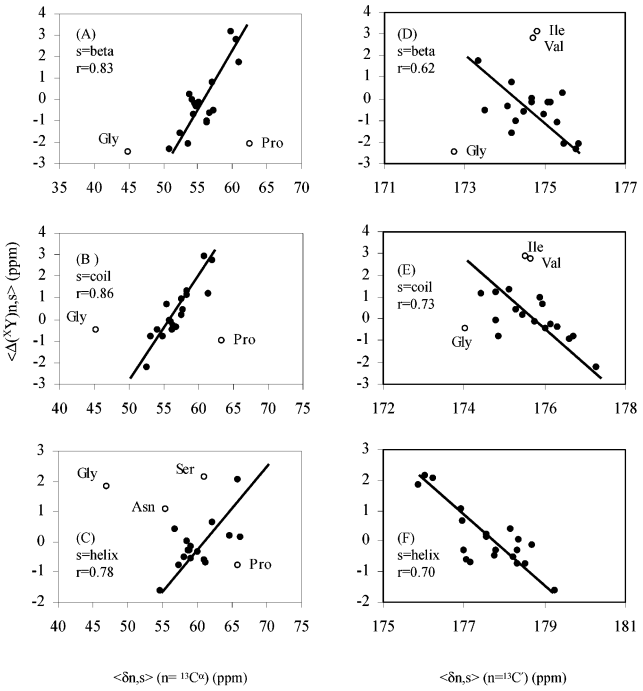

Figure 3. Relationships between $\langle\Delta(^{X}Y)n,s\rangle$ (n = ^{15}N) and $^{13}C^{\alpha}$ (A, B, C), and $^{13}C'$ (D, E, F) chemical shifts of the preceding residue X. Only part of the amino acid types of X are labeled. The secondary structural types and the correlation coefficients are shown in each individual figure.

$\langle\Delta(Y^{Z})n,s\rangle$ vary greatly with the three secondary structural types, and the averaged magnitude of $\langle\Delta(^{X}Y)n,s\rangle$ (n = ^{15}N) is in the order of β-strand >random coil > α-helix.

The large magnitude and wide variability in the ^{15}N chemical shift caused by the preceding residue make it possible to investigate the nature of the neighboring residue effect. Careful study on our data reveal a linear correlation between $\langle\Delta(^{X}Y)n,s\rangle$ (n = ^{15}N) and the $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts of the preceding residue X. As shown in Figure 3, the correlation coefficient of

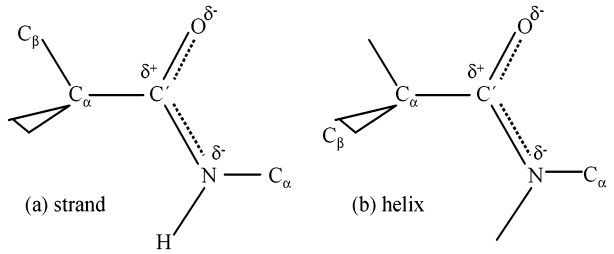

Figure 4. In β-strand the C^{β} atom of the preceding residue is opposite from the N atom. In α-helix they are on the same side.

the linear relation can be as high as 0.86. In Figure 3, parts A and B, there are two outliers, the preceding residue X = Gly and Pro, which can be attributed to the special structural characteristics: the absence of side chain for Gly and the absence of an amide proton and the cyclic nature of the side chain for Pro. Notable exceptions are X = Asp, Asn, Cys, and Ser in Figure 3C (s = helix), X = Val, Ile in Figure 3, parts D and E. We believe these deviations to be due to the conformational preferences of these amino aicds. For example, Asp, Asn, Cys, and Ser have high preference on forming α-helix capping box;[29] Val and Ile have branched side chains at their C^{β} atoms and process the highest β-strand propensities among the 20 amino acid.[30] The poor correlation between $\langle\Delta(^{X}Y)n,s\rangle$ (n = ^{15}N) and the $^{13}C'$ chemical shifts of the preceding residue X (Figure 3 parts D, E, and F) can be attributed to the narrow dispersion of the $^{13}C'$ chemical shifts ($\sim$3 ppm in each secondary structural type).

The 20 amino acids differ from one another in their variable side chain attaching to the C^{α} atom. Spanning a wide range of 40−70 ppm, $^{13}C^{\alpha}$ chemical shift of each of 20 amino acids is in fact dominated by the electron-withdrawing or electron-donating property of the side chain. The effects from the side chain and from the C^{α} atom are transmitted through the peptide bond and extend to the ^{15}N atom of the following residue. In this respect, the side chain of the preceding residue can be regarded as a substituent, and the C−N bond can be regarded as a conductor. The protein C−N peptide bond has some double bond character (40%) due to the resonance that occurs with amides. As a consequence of this resonance, C−N bonds in protein are found to be almost planar, i.e., the six atoms O(i-1), C′(i-1), C^{α}(i-1), N(i), HN(i), and C^{α}(i) are approximately coplanar. On the other hand, the overall electron transfer in the C−N is toward the N atom because of its high electronegativity, leaving the C−N bond with a considerable ionic character; the actual net charge in N atom is negative.[31] This hybrid covalent and ionic bond feature, we believe, may facilitate the effect from the preceding residue. Of great significance is that $\langle\Delta(^{X}Y)n,s\rangle$ (n = ^{15}N) varies notably with secondary structure. This observation supports the notice that the neighboring residue effect is an electron inductive effect. Due to the coplanarity of the peptide bond, the geometry between the N atom and the C^{β} atom of the preceding residue is dependent on the backbone dihedral angle ψ_{i-1}. As shown in Figure 4, in a β-strand (ψ_{i-1} = 130 ± 50°), C^{β}(i−1) atom is also approximate to the six-atom plane and is nearly in a trans position to the N(i) atom. In an α-helix (ψ_{i-1} = −47 ± 10°), the C^{β}−C^{α} bond of the

(29) Aurora, R.; Rose, G. D. *Protein Sci.* **1998**, *7*, 21−38.
(30) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267−270.
(31) Milner-White, E. *J. Protein Sci.* **1997**, *6*, 2477−2482.

Table 2. The Averaged Neighboring Residue Effect Correction Factor, $\langle\Delta(^{X}Y)n,s\rangle$ and $\langle\Delta(Y^{Z})n,s\rangle$, among the 20 Amino Acid Type of Y[a]

$\langle\Delta(^{X}Y)n,s\rangle$

	^{15}N			$^{13}C'$			$^{13}C^{\alpha}$		
X	beta	coil	helix	beta	coil	helix	beta	coil	helix
A	−2.29(0.91)	−2.21(0.43)	−1.60(0.48)	0.05(0.39)	0.14(0.26)	0.11(0.27)	−0.03(0.40)	−0.01(0.31)	−0.13(0.36)
C	0.80(2.01)	1.36(0.95)	0.67(1.37)	0.08(1.58)	−0.00(0.71)	−0.05(0.72)	−0.03(0.65)	0.44(0.56)	0.40(0.81)
D	−2.06(0.96)	−0.43(0.69)	0.43(0.74)	0.01(0.53)	0.07(0.25)	−0.21(0.41)	0.22(0.54)	0.20(0.20)	−0.12(0.25)
E	−0.15(0.91)	−0.36(0.89)	−0.13(0.52)	0.01(0.41)	0.04(0.29)	−0.02(0.32)	0.12(0.36)	0.01(0.31)	−0.23(0.24)
F	−1.01(0.89)	0.20(1.10)	−0.60(0.89)	−0.18(0.51)	−0.14(0.53)	0.05(0.53)	−0.22(0.31)	0.02(0.49)	0.13(0.46)
G	−2.45(0.63)	−0.43(0.56)	1.86(0.62)	−0.07(0.34)	−0.09(0.25)	−0.25(0.65)	0.13(0.36)	−0.26(0.25)	−0.32(0.51)
H	−0.34(1.44)	−0.05(1.30)	−0.27(0.96)	0.03(0.66)	−0.13(0.63)	−0.40(0.54)	0.14(0.61)	0.16(0.36)	0.01(0.56)
I	3.16(0.94)	2.92(1.28)	0.23(0.65)	0.12(0.33)	0.11(0.42)	0.12(0.43)	−0.17(0.24)	−0.15(0.37)	0.43(0.40)
K	−0.12(1.10)	−0.26(0.77)	−0.27(0.47)	0.08(0.42)	−0.09(0.29)	−0.18(0.40)	0.10(0.33)	−0.07(0.25)	−0.26(0.42)
L	0.27(0.70)	−0.76(0.88)	−0.74(0.59)	0.07(0.28)	0.13(0.38)	0.12(0.29)	−0.21(0.36)	−0.07(0.55)	0.14(0.25)
M	0.01(1.96)	0.69(1.23)	−0.48(0.90)	−0.19(0.57)	0.10(0.61)	−0.13(0.55)	−0.01(0.42)	0.09(0.41)	0.05(0.42)
N	−1.59(1.72)	−0.76(0.89)	1.09(1.00)	−0.20(0.59)	−0.19(0.43)	−0.53(0.45)	0.11(0.46)	0.18(0.29)	−0.18(0.32)
P	−2.05(1.01)	−0.94(0.94)	−0.74(1.28)	0.18(0.53)	0.21(0.24)	0.04(0.50)	0.14(0.56)	0.07(0.34)	0.04(0.39)
Q	−0.17(1.22)	−0.09(1.20)	0.04(0.37)	0.02(0.55)	0.07(0.28)	−0.04(0.29)	0.15(0.35)	0.13(0.38)	−0.09(0.40)
R	−0.68(1.02)	−0.45(0.70)	−0.52(0.86)	0.21(0.44)	−0.07(0.31)	0.01(0.33)	0.19(0.33)	0.00(0.30)	−0.17(0.54)
S	−0.52(0.66)	1.16(0.95)	2.18(0.61)	−0.08(0.47)	−0.10(0.24)	−0.34(0.58)	0.18(0.38)	0.11(0.32)	−0.34(0.36)
T	1.74(0.93)	1.23(1.13)	2.07(0.80)	−0.14(0.53)	−0.07(0.44)	−0.20(0.28)	0.06(0.30)	0.05(0.32)	0.10(0.43)
V	2.80(0.87)	2.77(1.29)	0.17(0.83)	0.04(0.40)	−0.00(0.50)	0.13(0.39)	−0.13(0.25)	−0.16(0.38)	0.27(0.36)
W	−1.06(1.11)	0.97(1.98)	−0.31(1.56)	−0.02(0.61)	−0.46(0.68)	0.15(0.85)	−0.14(0.59)	0.00(0.60)	0.19(0.75)
Y	−0.60(0.95)	0.46(1.53)	−0.67(0.97)	−0.08(0.44)	−0.03(0.60)	0.05(0.68)	−0.21(0.33)	−0.19(0.53)	0.13(0.47)

	$^{13}C^{\beta}$			$^{1}H^{N}$			$^{1}H^{\alpha}$		
X	beta	coil	helix	beta	coil	helix	beta	coil	helix
A	0.06(0.36)	−0.04(0.23)	0.01(0.22)	−0.02(0.16)	−0.07(0.12)	−0.08(0.06)	0.01(0.13)	−0.01(0.04)	−0.00(0.07)
C	0.28(0.61)	−0.18(0.39)	0.19(0.58)	0.22(0.30)	0.03(0.35)	0.12(0.17)	0.06(0.11)	−0.00(0.17)	−0.04(0.18)
D	−0.20(0.53)	−0.07(0.47)	0.20(0.27)	−0.22(0.20)	−0.01(0.14)	−0.05(0.11)	−0.05(0.18)	−0.03(0.04)	0.03(0.09)
E	−0.00(0.39)	0.04(0.29)	0.02(0.28)	0.01(0.15)	−0.01(0.08)	−0.04(0.08)	−0.05(0.08)	−0.01(0.05)	0.03(0.06)
F	0.26(0.33)	−0.06(0.36)	−0.04(0.64)	0.12(0.17)	0.01(0.29)	0.16(0.15)	0.06(0.11)	−0.01(0.08)	−0.09(0.08)
G	−0.16(0.40)	0.17(0.34)	0.14(0.37)	−0.41(0.16)	−0.10(0.08)	0.01(0.19)	−0.04(0.10)	0.04(0.08)	0.06(0.09)
H	0.05(0.54)	0.07(0.77)	0.22(0.46)	−0.02(0.22)	−0.01(0.22)	0.02(0.27)	−0.05(0.16)	−0.05(0.13)	0.00(0.11)
I	−0.04(0.31)	0.26(0.61)	−0.07(0.34)	0.15(0.12)	0.12(0.10)	0.02(0.11)	0.05(0.07)	0.04(0.07)	−0.03(0.06)
K	−0.08(0.54)	0.08(0.25)	−0.02(0.22)	−0.03(0.15)	−0.02(0.08)	−0.10(0.09)	−0.00(0.11)	−0.00(0.07)	0.03(0.09)
L	0.12(0.38)	−0.03(0.34)	−0.01(0.18)	0.12(0.13)	−0.06(0.13)	0.00(0.14)	0.04(0.09)	0.00(0.09)	−0.03(0.05)
M	0.05(0.66)	0.10(0.52)	−0.08(0.34)	0.11(0.24)	0.05(0.19)	−0.03(0.16)	0.06(0.16)	0.05(0.06)	−0.00(0.09)
N	0.15(0.43)	−0.20(0.31)	0.08(0.42)	−0.24(0.26)	0.01(0.12)	0.03(0.17)	−0.02(0.13)	−0.03(0.08)	0.05(0.13)
P	−0.22(0.71)	−0.16(0.32)	−0.11(0.33)	−0.17(0.23)	0.14(0.11)	0.08(0.31)	−0.16(0.13)	−0.04(0.08)	0.00(0.08)
Q	−0.21(0.62)	0.06(0.30)	0.15(0.28)	−0.05(0.17)	0.01(0.12)	−0.04(0.14)	−0.01(0.15)	−0.00(0.05)	0.03(0.06)
R	−0.12(0.70)	0.01(0.36)	0.03(0.40)	−0.11(0.21)	−0.03(0.10)	−0.04(0.15)	−0.00(0.12)	0.01(0.06)	0.06(0.08)
S	0.14(0.36)	−0.06(0.31)	−0.01(0.42)	−0.05(0.13)	0.02(0.12)	0.05(0.10)	0.01(0.14)	0.01(0.04)	0.07(0.07)
T	−0.14(0.44)	−0.11(0.45)	−0.16(0.28)	0.08(0.19)	0.02(0.16)	0.12(0.11)	−0.01(0.09)	−0.00(0.12)	0.03(0.07)
V	−0.17(0.36)	0.03(0.44)	−0.14(0.39)	0.14(0.09)	0.17(0.17)	0.01(0.14)	0.02(0.07)	0.04(0.11)	−0.03(0.06)
W	0.61(0.81)	0.28(0.61)	−0.25(0.80)	0.26(0.23)	−0.08(0.43)	0.20(0.25)	0.02(0.23)	−0.05(0.14)	−0.13(0.20)
Y	0.22(0.39)	−0.10(0.49)	−0.12(0.48)	0.13(0.18)	0.02(0.11)	0.11(0.18)	0.00(0.08)	0.01(0.12)	−0.11(0.10)

$\langle\Delta(Y^{Z})n,s\rangle$

	^{15}N			$^{13}C'$			$^{13}C^{\alpha}$		
Z	beta	coil	helix	beta	coil	helix	beta	coil	helix
A	−0.41(0.73)	−0.11(0.44)	0.06(0.49)	−0.12(0.49)	0.05(0.27)	−0.07(0.20)	0.03(0.24)	0.07(0.33)	−0.04(0.19)
C	−0.85(1.12)	−1.17(1.68)	−0.16(0.88)	−0.18(0.53)	0.10(0.72)	−0.07(1.00)	−0.10(0.47)	0.17(0.91)	0.09(0.57)
D	−0.10(1.04)	0.23(0.71)	0.12(1.03)	0.06(0.34)	−0.11(0.25)	−0.26(0.25)	0.03(0.43)	0.28(0.33)	0.09(0.33)
E	−0.13(1.01)	0.26(1.03)	0.25(0.51)	0.13(0.37)	0.14(0.24)	0.07(0.31)	0.18(0.32)	0.25(0.53)	0.21(0.29)
F	−0.36(1.05)	−0.35(0.97)	0.34(0.93)	−0.23(0.45)	−0.22(0.60)	0.04(0.52)	−0.12(0.38)	−0.04(0.37)	0.05(0.36)
G	−0.13(0.83)	0.13(0.70)	−1.20(0.92)	0.61(0.37)	0.47(0.24)	−0.41(0.41)	0.09(0.37)	0.12(0.28)	−1.09(0.70)
H	−0.45(1.98)	−0.09(1.44)	−0.28(1.07)	−0.20(0.64)	−0.05(0.49)	−0.21(0.55)	0.08(0.69)	0.22(0.55)	−0.04(0.68)
I	−0.01(0.70)	−0.20(0.95)	0.14(0.77)	−0.04(0.35)	−0.09(0.32)	0.17(0.28)	0.10(0.42)	0.03(0.36)	0.20(0.49)
K	0.12(0.69)	−0.13(0.51)	−0.05(0.59)	0.09(0.40)	−0.13(0.35)	−0.16(0.33)	0.22(0.26)	0.08(0.30)	−0.11(0.25)
L	−0.01(0.63)	−0.49(0.96)	0.30(0.57)	−0.27(0.28)	0.06(0.38)	0.11(0.35)	0.09(0.27)	0.10(0.44)	0.01(0.31)
M	0.34(2.07)	−0.02(1.55)	0.10(0.80)	−0.05(0.64)	0.19(0.71)	−0.08(0.48)	0.23(0.52)	0.22(0.27)	0.14(0.42)
N	−0.07(1.40)	−0.03(0.86)	−0.12(0.80)	−0.01(0.56)	−0.23(0.37)	−0.34(0.68)	0.21(0.45)	0.24(0.28)	−0.14(0.54)
P	0.89(1.31)	0.92(0.88)	0.33(1.63)	−0.92(1.17)	−1.19(0.53)	−2.49(1.45)	−1.63(0.56)	−2.04(0.45)	0.51(2.54)
Q	0.03(0.96)	−0.31(0.82)	−0.07(0.66)	0.04(0.55)	0.10(0.36)	0.10(0.56)	0.20(0.36)	0.37(0.45)	0.00(0.38)
R	−0.23(1.11)	−0.09(0.81)	−0.09(0.83)	0.00(0.47)	0.16(0.36)	0.01(0.35)	0.03(0.45)	0.19(0.33)	0.03(0.33)
S	0.32(0.83)	0.30(0.72)	−0.17(0.88)	0.26(0.35)	0.10(0.27)	0.03(0.50)	0.11(0.29)	0.10(0.30)	−0.24(0.64)
T	0.35(1.01)	0.22(0.88)	−0.54(0.74)	0.38(0.38)	0.10(0.35)	0.02(0.34)	0.07(0.27)	0.02(0.47)	−0.03(0.66)
V	0.16(0.67)	−0.05(0.60)	0.25(0.64)	0.05(0.20)	−0.08(0.29)	0.26(0.28)	0.12(0.28)	0.05(0.41)	0.07(0.38)
W	0.06(1.99)	−0.59(1.34)	0.06(1.25)	−0.46(0.62)	−0.33(0.50)	−0.00(0.54)	0.02(0.75)	−0.06(0.60)	0.14(0.71)
Y	−0.42(1.15)	−0.48(1.63)	0.03(0.93)	−0.29(0.45)	−0.51(0.68)	−0.19(0.86)	−0.12(0.38)	0.10(0.58)	0.05(0.31)

Table 2 (Continued)

	$\langle\Delta(Y^Z)n,s\rangle$								
	$^{13}C^{\beta}$			$^{1}H^{N}$			$^{1}H^{\alpha}$		
Z	beta	coil	helix	beta	coil	helix	beta	coil	helix
A	0.05(0.35)	−0.09(0.29)	−0.11(0.21)	−0.02(0.10)	−0.01(0.13)	0.05(0.11)	−0.03(0.09)	−0.03(0.05)	−0.01(0.05)
C	0.25(0.95)	0.21(0.67)	0.37(0.49)	−0.05(0.17)	0.01(0.29)	0.04(0.14)	0.06(0.13)	0.03(0.11)	0.04(0.07)
D	0.08(0.66)	0.11(0.30)	−0.07(0.39)	0.04(0.17)	0.04(0.13)	0.12(0.22)	−0.06(0.13)	−0.03(0.07)	−0.05(0.08)
E	0.15(0.38)	0.06(0.27)	−0.13(0.19)	0.02(0.08)	0.07(0.07)	0.13(0.06)	−0.04(0.11)	−0.04(0.04)	−0.05(0.05)
F	0.10(0.52)	−0.09(0.65)	0.07(0.32)	−0.11(0.14)	−0.04(0.20)	−0.10(0.15)	0.08(0.12)	−0.02(0.10)	−0.03(0.09)
G	−0.20(0.80)	−0.07(0.26)	0.14(0.45)	−0.02(0.15)	0.06(0.10)	−0.05(0.19)	−0.08(0.16)	−0.03(0.07)	0.12(0.10)
H	−0.03(0.88)	−0.24(0.74)	0.09(0.40)	−0.02(0.17)	−0.01(0.19)	−0.02(0.18)	−0.04(0.16)	−0.09(0.11)	−0.05(0.18)
I	0.05(0.27)	0.28(0.32)	0.11(0.43)	0.01(0.17)	−0.01(0.15)	−0.21(0.07)	0.06(0.16)	0.03(0.10)	0.01(0.05)
K	−0.06(0.53)	0.01(0.29)	−0.05(0.28)	0.07(0.17)	−0.04(0.10)	0.05(0.11)	−0.04(0.09)	−0.03(0.04)	−0.00(0.06)
L	−0.03(0.43)	−0.10(0.44)	0.10(0.28)	0.04(0.13)	−0.07(0.15)	−0.14(0.11)	0.01(0.08)	−0.02(0.04)	0.03(0.05)
M	0.18(0.68)	0.06(0.45)	0.05(0.36)	0.04(0.32)	0.01(0.27)	−0.06(0.15)	0.00(0.20)	−0.01(0.07)	−0.01(0.09)
N	−0.21(0.53)	−0.06(0.28)	0.04(0.29)	−0.02(0.15)	0.04(0.11)	0.11(0.17)	−0.13(0.16)	−0.04(0.04)	−0.03(0.09)
P	−0.94(0.98)	−0.20(0.52)	−1.70(1.45)	−0.07(0.13)	−0.17(0.12)	0.04(0.36)	−0.07(0.14)	0.21(0.10)	0.18(0.16)
Q	0.01(0.36)	−0.13(0.29)	−0.04(0.37)	0.04(0.16)	0.01(0.14)	0.07(0.13)	−0.05(0.11)	−0.05(0.04)	−0.02(0.05)
R	0.10(0.55)	−0.12(0.30)	−0.02(0.15)	0.09(0.15)	−0.01(0.09)	0.01(0.10)	0.01(0.15)	−0.02(0.06)	−0.02(0.10)
S	0.08(0.34)	0.14(0.30)	−0.10(0.48)	−0.03(0.18)	0.07(0.09)	0.13(0.10)	0.01(0.12)	0.02(0.05)	0.04(0.08)
T	0.14(0.38)	0.14(0.43)	0.01(0.29)	0.02(0.14)	0.04(0.13)	0.09(0.16)	0.05(0.09)	0.08(0.07)	0.03(0.10)
V	−0.11(0.28)	0.18(0.39)	0.05(0.33)	0.01(0.08)	0.01(0.14)	−0.17(0.09)	0.05(0.09)	0.05(0.09)	0.03(0.05)
W	0.28(0.78)	−0.03(0.67)	0.07(0.80)	−0.16(0.29)	−0.10(0.47)	−0.08(0.19)	−0.05(0.26)	−0.03(0.14)	0.01(0.20)
Y	0.30(0.45)	0.00(0.83)	0.07(0.48)	−0.09(0.18)	−0.07(0.15)	−0.03(0.20)	0.09(0.15)	−0.06(0.13)	−0.03(0.10)

[a] Standard deviation is in parentheses.

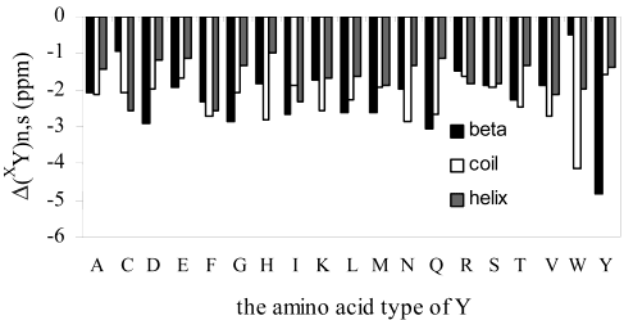

Figure 5. Variation of $\Delta(^{X}Y)n,s$ (X = Ala, n = ^{15}N) with the amino acid type of Y.

preceding residue is nearly perpendicular to the six-atom plane leaving the $C^{\beta}(i-1)$ atom at a *cis* position of the N(i) atom. Meanwhile, this also provides an explanation to the "unusual" data for Gly, which can easily adopt a positive ψ angle, and for Pro, which can form a *cis* peptide bond. The lesser influence on the ^{15}N shift by the following residue is due to the larger number of bonds between the ^{15}N atom and the side chain (C^{β} atom) of the following residue.

As shown in Figure 5, the neighboring effect varies not only with the amino acid type of the neighboring residue (X, Z) but also with the residue under study (Y). This dependence on the amino acid type of the residue under study was clearly demonstrated by Wishart et al. using model peptides.[22] The averaged standard deviations over the 20 amino acids of Y for the preceding residue effect on ^{15}N shifts were 1.03, 0.88, and 0.65 ppm for β-strand, random coil, and α-helix. These standard deviations are small compared to the size of $\Delta(^{X}Y)n,s$ (n = ^{15}N), which is in a range of −3 to 3 ppm. Thus it appears that the preceding residue effect on the ^{15}N chemical shift is a combined result from the interaction between the side chain of the preceding residue and the side chain of the residue under investigation.

The above observations indicate that the side chain of a residue can extend its influence to the following residue via

Table 3. The Second-Order Neighboring Residue Effect Correction Factors, $\langle\Delta\Delta(^{X}Y)n,s\rangle$ (n = ^{15}N) for the 20 Amino Acids of the Preceding Residue X

groups	amino acid of X	$\langle\Delta\Delta(^{X}Y)n,beta\rangle$	$\langle\Delta\Delta(^{X}Y)n,helix\rangle$
I	I	0.24	−2.69
	V	0.03	−2.60
	av (stdv)	**0.14(0.15)**	**−2.65(0.06)**
II	D	−1.63	0.86
	N	−0.83	1.85
	av (stdv)	**−1.23(0.57)**	**1.36(0.70)**
III	H	*−0.29*[a]	*−0.22*[a]
	Y	−1.06	−1.13
	W	−2.03	−1.28
	F	−1.21	−0.80
	av (stdv)	**−1.43(0.52)**	**−1.07(0.25)**
IV	L	*1.03*[a]	0.02
	M	−0.68	*−1.17*[a]
	R	−0.23	−0.07
	K	0.14	−0.01
	Q	−0.08	0.13
	E	0.21	0.23
	av (stdv)	**−0.13(0.35)**	**0.06(0.12)**
V	T	*0.51*[a]	0.84
	C	−0.56	*−0.69*[a]
	S	−1.68	1.02
	av (stdv)	**−1.12(0.79)**	**0.93(0.13)**
others	A	−0.08	0.61
	G	−2.02	2.29
	P	−1.11	0.20

[a] Not included in the calculation of average and stdv.

the peptide bond. We propose that the influence of the neighboring residue is electronic induction. This mechanism of the neighboring residue effect has potential practical applications in understanding other phenomena. For example, reduction of the S−S bridge of redox proteins, i.e., thioredoxin and glutaredoxin, usually causes substantial changes of chemical shifts for neighboring residues of cystine but not in their geometry.[32] This is not explained by the current theory in which chemical shifts are largely dominated by conformation. From the oxidized to

(32) Nordstrand, K.; Aslund, F.; Meunier, S.; Holmgren, A.; Otting, G.; Berndt, K. D. *FEBS Lett.* **1999**, *449*, 196−200.

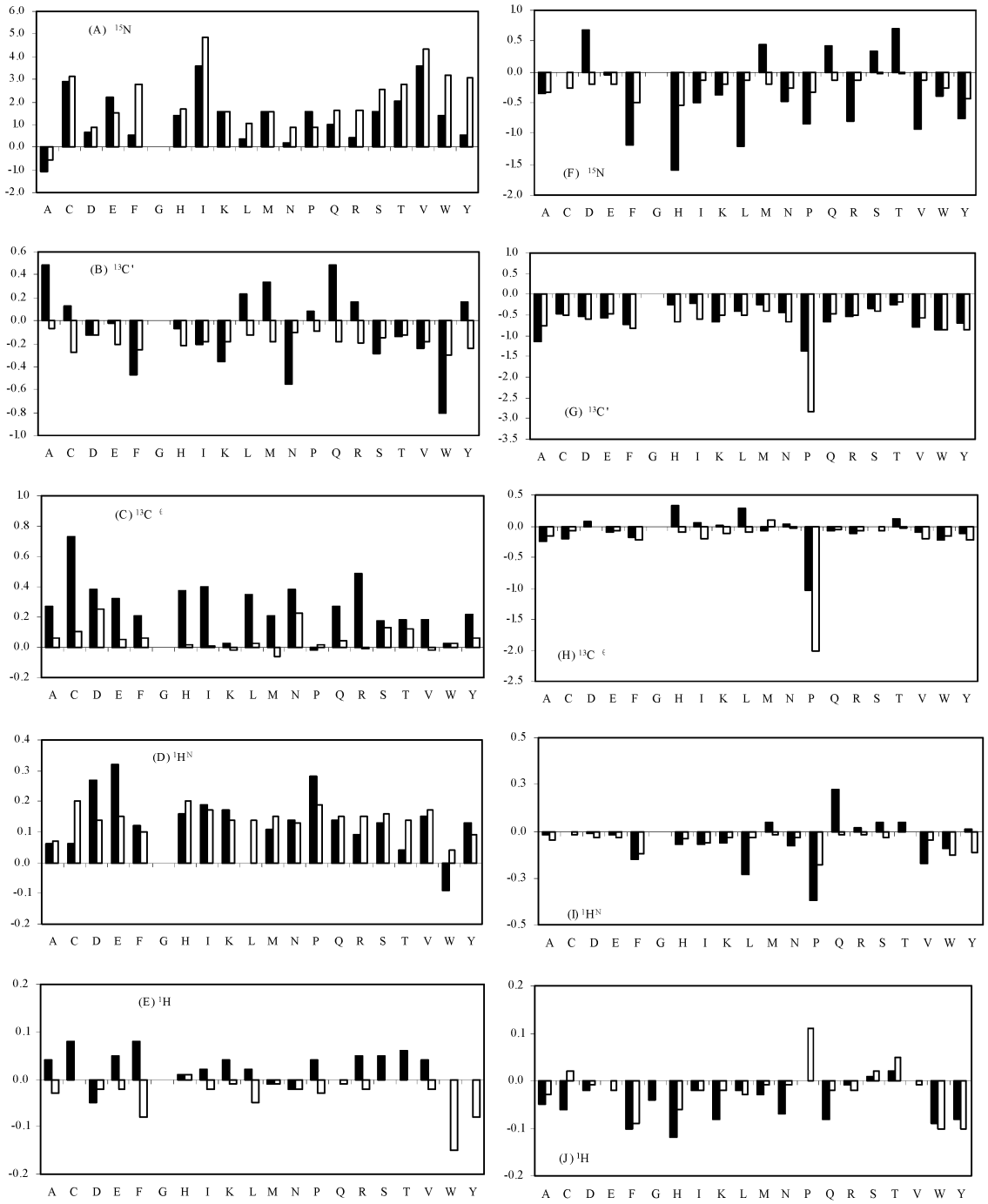

Figure 6. Comparison between the adjusted (see text) $\Delta(^{X}Y)n,s$ (solid bar) and sequence dependent correction factors (open bar) by Schwarzinger et al.[16] On the left (A, B, C, D, E) are for the preceding residue effects and the right (F, G, H, I, J) for the following residue effect.

the reduced states, cystine normally experiences changes of ∼3 and ∼13 ppm in its $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, respectively. On the basis of the mechanism we proposed, changes in the chemical shifts of its neighboring residue are understandable.

$\langle\Delta\Delta(^{X}Y)n,s\rangle$ and $\langle\Delta\Delta(Y^{Z})n,s\rangle$. In the present study, we have also defined and calculated second-order correction factors, $\langle\Delta\Delta(^{X}Y)n,s\rangle$ and $\langle\Delta\Delta(Y^{Z})n,s\rangle$, which represent the variation of the neighbor residue effect with the secondary structure. Listed in Table 3 are the second-order correction factors for the preceding residue effect on the ^{15}N shifts, $\langle\Delta\Delta(^{X}Y)n,s\rangle$ ($n = {}^{15}N$). In this table, the amino acids of the preceding residue X are also categorized into the same groups as in Table 1. As shown in Table 3, in each group, $\langle\Delta\Delta(^{X}Y)n,beta\rangle$ and $\langle\Delta\Delta(^{X}Y)n,helix\rangle$ values are close to each other with few outliers. In group V, Cys and Ser have similar values in $\langle\Delta\Delta(^{X}Y)n,beta\rangle$ but not in $\langle\Delta\Delta(^{X}Y)n,helix\rangle$. Also in group V, $\langle\Delta\Delta(^{X}Y)n,beta\rangle$ and $\langle\Delta\Delta(^{X}Y)n,helix\rangle$ values of Thr, which also has a branched side chain like Val and Ile in group I, show large discrepancies with Ser and Cys. This correlation between $\langle\Delta\Delta(^{X}Y)n,s\rangle$ ($n = {}^{15}N$) and X's side chain indicate that both Y and X share the same mechanism in their effect on the ^{15}N shift of Y.

Comparison to Other Studies. At present, there are three published correction factors for the 20 amino acids of the neighboring residues using experimental approaches.[21−23] The methods used in these earlier studies are basically the same,

Table 4. The Averaged and RMS Deviations (in Parentheses) of the Predicted Chemical Shifts to the Observed Values

data set[a]	parameters used in prediction[b]	^{15}N		$^{1}H^{N}$		$^{13}C^{\beta}$	
		before	after	before	after	before	after
A	present protein-based	0.42(3.63)	0.31(3.22)	0.05(0.54)	0.03(0.52)	0.24(1.43)	0.20(1.35)
	earlier peptide-based			n/a			
B	present protein-based	−0.08(3.39)	0.04(2.94)	0.06(0.57)	0.02(0.57)	0.13(1.03)	0.17(1.04)
	earlier peptide-based	0.06(3.51)	−1.09(3.30)	−0.18(0.57)	−0.26(0.58)	0.09(1.48)	n/a
C	present protein-based	1.09(1.61)	1.27(1.65)	0.13(0.16)	0.13(0.25)	−0.02(0.72)	0.01(0.87)
	earlier peptide-based	1.03(1.72)	−0.33(1.33)	−0.11(0.14)	−0.20(0.12)	−0.16(0.23)	N/a

data set	parameters used in prediction	$^{13}C^{\alpha}$		$^{13}C'$		$^{1}H^{\alpha}$	
		before	after	before	after	before	after
A	present protein-based	0.16(1.43)	0.11(1.30)	0.15(1.49)	0.18(1.31)	0.00(0.36)	0.01(0.34)
	earlier peptide-based			n/a			
B	present protein-based	0.28(1.49)	0.24(1.32)	0.22(1.04)	0.14(1.03)	−0.03(0.29)	0.00(0.28)
	earlier peptide-based	0.01(1.60)	0.13(1.52)	−0.87(1.21)	−0.18(1.16)	−0.10(0.30)	0.07(0.29)
C	present protein-based	0.21(0.68)	0.09(0.72)	0.37(0.74)	0.32(0.70)	−0.03(0.08)	0.01(0.11)
	earlier peptide-based	−0.09(0.53)	0.03(0.37)	−0.72(0.69)	−0.07(0.35)	−0.10(0.07)	−0.05(0.05)

[a] Data sets: A is composed of β-strand, random coil, and α-helix chemical shifts of 4 proteins (510 residues in total) under normal conditions. B is composed of random coil chemical shifts derived from 10 proteins (259 residues in total) under normal conditions. C is unfolded apomyoglobin (153 residues) under denaturing conditions (pH 2.3, and 8M urea). [b] The parameters used in the predictions are δn,coil (random coil shift), $\Delta\delta$n,s (secondary structural effect), $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s (neighboring residue effects). These parameters are from the present (protein-based) and earlier (peptide-based) studies by Schwarzinger et al.[23,34] Since the earlier peptide-based studies do not have the values of $\Delta\delta$n,s, they are not applicable to set A.

e.g., using short Gly enriched peptides under denaturing conditions. As a consequence, their results agree well to each other. The data recently reported by Schwarzinger et al.,[23] which contain the correction factors for both the preceding and the following residues, were chosen to compare with present results. To match the model peptides of sequence of Ac-GGXGG-NH_2 used by Schwarzinger et al., the present correction factors with Y = Gly and s = random coil were selected for comparison. For the present data, the statistically averaged chemical shifts were used as reference. As a consequence, they are either positive or negative, and the sum is always zero. For the "sequence dependent correction factors" by Schwarzinger et al.,[23] the "random coil" chemical shifts measured from short peptide under denaturing conditions were used as reference, leaving the "correction factors" of Gly zero by definition and that of the remaining 19 amino acids either all positive or all negative. To compare the two sets of data, the magnitude of the above selected correction factors from the present study was adjusted, so that $\Delta(^{X}Y)$n,coil and $\Delta(Y^{Z})$n,coil equal zero when X (or Z) is Gly.

The adjusted $\Delta(^{X}Y)$n,coil and $\Delta(Y^{Z})$n,coil in the present study and the correction factors by Schwarzinger et al. are graphically displayed in Figure 6. Among the 10 sets of calibration and correction factors, six (Figure 6, parts A, G, H, D, I, and J) show reasonably good agreement. Notable differences in the remaining four sets of data could result from three factors. The first factor is difference in pH and solvent, which may have notable effects on chemical shifts. The second factor is preferred conformations for certain amino acid pair adopted model peptides. As shown in Figure 6, parts G, H, and J, for the effect from the following proline, Schwarzinger's correction factors on $^{13}C'$, $^{13}C^{\alpha}$, and $^{1}H^{\alpha}$ chemical shifts are significantly larger than those obtained with the present data (−2.84, −2.0, and 0.11 vs −1.36, −1.01, and 0.0). It was found that residues directly preceding proline have a propensity to adopt a β-sheet-like or poly-proline II conformation,[33] and thus many of the correction factors for the effect from the following proline by Schwarzinger et al., as observed by Wishart et al.,[22] could be accounted for by a conformational bias. To check the influence of conformational bias on present results, we compared the backbone ϕ and ψ distributions of Y (Y = Gly, s = random coil) directly followed by proline with those followed by other amino acids. Indeed, when followed by proline, most of ϕ and ψ of Y are within the β-strand regions. The correction factors were recalculated by restricting the two sets of $\langle\delta$n,s(X)$\rangle$ − $\langle\delta$n,s(w/o X)$\rangle$ data to the same areas of (ϕ and ψ) space, resulting in modest changes in the magnitude of $\Delta(^{X}Y)$n,coil. The recalculated $\Delta(^{X}Y)$n,coil (X = Pro, Y = Gly) are −1.01, −0.63, and 0.03 ppm for 13C′, $^{13}C^{\alpha}$, and $^{1}H^{\alpha}$, respectively. The conformational bias is largely canceled out during the averaging calculation in the present study. The third factor is the different referencing chemical shifts used in this study and in the earlier experimental approaches. In our previous study,[25] we noticed that significant discrepancies exist between the averaged chemical shifts and the "random coil" values measured from model peptides. The notable discrepancies between the two sets of data, which for example can be as large as 4.1 ppm for Cys $^{13}C^{\beta}$, are, we believe, caused by low pH (∼2.0) used in studies of the peptide model.

We further prepared three sets of chemical shift data, marked as A, B, and C, to assess the quality of the present parameters, δn,coil (random coil shift), $\Delta\delta$n,s (secondary structural effect), and $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s (neighboring residue effects). The evaluations were made through comparing the observed chemical shifts with that predicted from eq 1 using the above parameters in the present study and those peptide-based parameters by Schwarzinger et al.[23,34] The averaged and root-mean-square (RMS) deviations of the predicted chemical shifts to the observed values are listed in Table 4.

Set A is composed of chemical shift assignments of four folded proteins under normal conditions. The parameters by Schwarzinger et al.[23,34] and other earlier peptide-based studies, which are limited to the unfolded protein, are not applicable to this set of data. As shown in Table 4, there are no significant averaged deviations for all the six nuclei using parameters in

(33) Williamson, M. P. *Biochem. J.* **1994**, *297*, 249−260.

(34) Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Wright, P. E.; Dyson, H. J. *J. Biomol. NMR* **2000**, *18*, 43−48.

this study. Upon introduction of the neighbor residue effect correction, the RMS deviations drop by 0.40, 0.18, 0.13, 0.08, 0.02, and 0.02 ppm for ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{N}$, and $^{1}H^{\alpha}$, respectively.

Set B contains random coil chemical shift assignments of 259 residues derived from 10 proteins under normal conditions. Using parameters in this study, there are no significant averaged deviations for all the six nuclei before and after the correction for the neighboring residue effects, and upon the correction, the RMSD notably drops by 0.44 and 0.17 ppm for ^{15}N and $^{13}C^{\alpha}$, respectively. Using peptide-based parameters, notable averaged deviations (-0.87 ppm for $^{13}C'$ before correction and -1.09 ppm for ^{15}N after correction) were found.

Set C is unfolded apomyoglobin under the denaturing conditions (pH 2.3, 8 M urea). Using present parameters, relatively larger averaged deviations (1.09 ppm) before and (1.27 ppm) after corrections for ^{15}N shifts are obtained. Corrections using present data show no improvement on RMSD for all the nuclei. Using peptide-based parameters, notable averaged deviations of 1.03 ppm for ^{15}N and -0.72 ppm for $^{13}C'$ shifts were found before correction for the neighboring residue effects. After correction, the RMSD drop by 0.39, 0.16, and 0.34 ppm for ^{15}N, $^{13}C^{\alpha}$, and $^{13}C'$, respectively. It is of interest that the RMSD of $^{13}C^{\beta}$ using present parameters is 0.45 ppm lower than that from peptide-based parameters in set B but 0.49 higher in set C. Further study shows that the differences in the RMSD of $^{13}C^{\beta}$ are caused by the Asp, Glu, and Gln, which may be partially or totally protonated at low pH.

In conclusion, for proteins under nondenaturing conditions, the present data, the averaged random coil chemical shifts (δn,coil) and the correction factor for secondary structural effect ($\Delta\delta$n,beta and $\Delta\delta$n,helix), are recommended for the purposes of secondary structural identification from chemical shifts or the chemical shift prediction from sequence and secondary structural information. Corrections for neighboring residue effect using present factors, $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s, $\Delta\Delta(^{X}Y)$n,s and $\Delta\Delta(Y^{Z})$n,s, provide moderate improvement of fit for amide ^{15}N and ^{13}C chemical shifts. Upon the neighboring residue effect correction, there is no notable improvement on ^{1}H shift. By application of the peptide-based parameters, random coil chemical shifts and neighboring residue effect correction factors for the folded proteins may cause systematic bias and large RMS deviations for ^{15}N, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts. For proteins under denaturing conditions, peptide-based parameters are recommended, and the corrections for the neighboring residue effects are needed to avoid systematic bias on ^{15}N and $^{13}C'$ chemical shifts.

Acknowledgment. NIH Grant 1R01 GM33385 is acknowledged.

Supporting Information Available: A dictionary of the neighboring residue effects, $\Delta(^{X}Y)$n,s and $\Delta(Y^{Z})$n,s, on protein chemical shifts. This material is available free of charge via the Internet at http://pubs.acs.org.

JA026811F