Sequence-Dependent Correction of Random Coil NMR Chemical Shifts

Stephan Schwarzinger, Gerard J. A. Kroon, Ted R. Foss, John Chung, Peter E. Wright,* and H. Jane Dyson*

Contribution from the Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037

Received October 23, 2000. Revised Manuscript Received January 17, 2001

Abstract: Random coil chemical shifts are commonly used to detect secondary structure elements in proteins in chemical shift index calculations. While this technique is very reliable for folded proteins, application to unfolded proteins reveals significant deviations from measured random coil shifts for certain nuclei. While some of these deviations can be ascribed to residual structure in the unfolded protein, others are clearly caused by local sequence effects. In particular, the amide nitrogen, amide proton, and carbonyl carbon chemical shifts are highly sensitive to the local amino acid sequence. We present a detailed, quantitative analysis of the effect of the 20 naturally occurring amino acids on the random coil shifts of $^{15}N^{H}$, $^{1}H^{N}$, and ^{13}CO resonances of neighboring residues, utilizing complete resonance assignments for a set of five-residue peptides Ac-G-G-X-G-G-NH_2. The work includes a validation of the concepts used to derive sequence-dependent correction factors for random coil chemical shifts, and a comprehensive tabulation of sequence-dependent correction factors that can be applied for amino acids up to two residues from a given position. This new set of correction factors will have important applications to folded proteins as well as to short, unstructured peptides and unfolded proteins.

Introduction

Chemical shift index (CSI) calculations, in which characteristic deviations of the chemical shifts of certain nuclei relative to random coil chemical shifts are exploited to detect secondary structure elements, have become a standard procedure in the characterization of proteins by NMR.[1−5] Since the chemical shifts of several nuclei are usually available from the assignment of the protein backbone resonances, this is a quick and reliable procedure. However, the results are influenced by the quality of the random coil shifts used as reference data in the calculation. Over time a number of different sets of random coil shifts have been determined.[6−15] Because data derived from protein databases[6,7] display some bias, random coil shifts obtained with series of short peptides under a variety of experimental conditions are commonly used as reference data in CSI calculations.

The major current usage for CSI calculations is for the detection of secondary structure elements in folded proteins and of residual structure in unfolded and partly folded proteins. A local sequence dependence of the chemical shifts of some nuclei, notably ^{13}CO and ^{15}N, has been noted and exploited to perform sequential assignments on unfolded proteins,[16−18] while other nuclei, notably $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$, show a much lower influence of sequence and can be reliably used as indicators of residual secondary structure. For folded proteins, a consensus chemical shift index calculated over the backbone nuclei ($^{1}H^{N}$, $^{1}H^{\alpha}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO) gives a good prediction of secondary structure elements. However, because of the sequence dependence of the chemical shifts of several of these nuclei, rather poor results are obtained for unstructured or partly structured proteins. Correction factors have been published for the sequence-dependent effect of the preceding residue X on the ^{15}N chemical shift of the Ala4 residue in a set of H-GGXA-OH peptides.[11] The chemical shift predictions made for unfolded proteins using these correction factors fitted experimental ^{15}N shifts with an RMSD of 0.8−1 ppm. In another study Wishart and colleagues[12] measured random coil chemical shifts for a series of peptides

* Corresponding authors: Dr. Peter E. Wright and Dr. H. Jane Dyson, Telephone: (858) 784-9721. Fax: (858) 784-9822.

(1) Gross, K.-H.; Kalbitzer, H. R. *J. Magn. Reson.* **1988**, *76*, 87−99.
(2) Szilágyi, L.; Jardetzky, O. *J. Magn. Reson.* **1989**, *83*, 441−449.
(3) Pastore, A.; Saudek, V. *J. Magn. Reson.* **1990**, *90*, 165−176.
(4) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490−5492.
(5) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647−1651.
(6) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311−333.
(7) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363−392.
(8) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285−297.
(9) Richarz, R.; Wüthrich, K. *Biopolymers* **1978**, *17*, 2133−2141.
(10) Jimenez, M. A.; Nieto, J. L.; Rico, M.; Santoro, J.; Herranz, J.; Bermejo, F. J. *J. Mol. Struct.* **1986**, *143*, 435−438.
(11) Braun, D.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1994**, *116*, 8466−8469.
(12) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67−81.
(13) Merutka, G.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **1995**, *5*, 14−24.
(14) Thanabal, V.; Omecinsky, D. O.; Reily, M. D.; Cody, W. L. *J. Biomol. NMR* **1994**, *4*, 47−59.
(15) Plaxco, K. W.; Morton, C. J.; Grimshaw, S. B.; Jones, J. A.; Pitkeathly, M.; Campbell, I. D.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 221−230.
(16) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559−1563.
(17) Logan, T. M.; Thériault, Y.; Fesik, S. W. *J. Mol. Biol.* **1994**, *236*, 637−648.
(18) Yao, J.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1997**, *419*, 285−289.

10.1021/ja003760i CCC: $20.00
Published on Web 03/06/2001

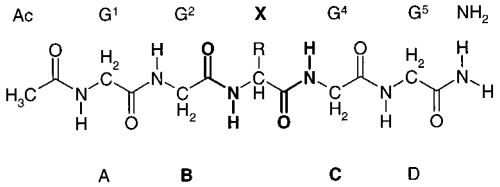

Figure 1. Schematic representation showing nomenclature used for the Ac-GGXGG-NH_2 model peptides. The two peptide bonds involving the central residue X are depicted in bold print. Throughout this paper we will use the terms G1, G2, G4, and G4 for observed chemical shifts (δ), while *A*, *B*, *C*, and *D* will be used for the correction factors ($\Delta\delta$ values).

Ac-GGXAGG-NH_2 and Ac-GGXPGG-NH_2 and investigated the effect on the chemical shifts of residue X (^{13}C and ^{1}H) of the following residue (A or P, respectively). Some significant differences were found between these two data sets. However, to obtain the sequence dependence of the chemical shift of all dipeptide sequences using this method would require data to be measured on 400 peptides. To investigate the sequence-dependent effect on the preceding and the following residues in a tripeptide sequence, it would be necessary to measure data on 8000 peptides of this type.

We have used a simplified approach that combines those of Wishart et al.[12] and Braun et al.[11] to analyze data for a series of model peptides of sequence Ac-GGXGG-NH_2 to determine the variations in the chemical shifts of the four glycine residues caused by substitutions at position X. Complete assignments have been made for all resonances in these peptides in 8 M urea at pH 2.3, and the changes caused by changing the central residue X have been analyzed. A complete tabulation is made of correction factors for the random coil chemical shifts of the most sequence-dependent nuclei, ^{15}N, $^{1}H^{N}$, and ^{13}CO. We demonstrate the usefulness of the obtained correction factors for CSI calculations in the case of urea-denatured apomyoglobin, employing a modified CSI-module for the program NMRView.[19]

Experimental Section

Sample Preparation. Synthesis and purification of the model peptides Ac-GGXGG-NH_2, where X is one of the 20 naturally occurring amino acids, have been described elsewhere.[20] The nomenclature for the model peptides is depicted in Figure 1, where G1, G2, G4, and G5 represent the four glycine residues, X represents one of the 20 naturally occurring amino acids, and *A*, *B*, *C*, and *D* represent the respective correction factors to be applied to the chemical shifts of glycines G1, G2, G4, and G5. By keeping residues 1, 2, 4, and 5 constant in the peptide series, it is possible to quantitatively describe the effect of the variant central residue X on the chemical shifts of these residues.

Urea-denatured apomyoglobin was prepared by buffer exchange of recombinant apomyoglobin (sperm whale sequence)[21] into 10 mM acetic acid pH 2.3 containing 8 M urea.

NMR Spectroscopy. The NMR spectra were recorded on a Bruker DMX 750 spectrometer as previously described.[20] Resonance assignments for each peptide were made with reference to one-dimensional ^{1}H and natural-abundance ^{13}C spectra, while ^{15}N assignments were made using natural abundance $^{15}N-^{1}H$ HSQC spectra.[20] The probe temperature was calibrated to 293 K with neat methanol.[22] Proton and carbon shifts were referenced to an internal standard DSS, and ^{15}N was referenced indirectly to DSS using a ratio of 0.10132912.[23]

Data Analysis. Correction factors were calculated relative to the Ac-GGGGG-NH_2 peptide as a reference. The four correction factors were derived for each nucleus ($^{1}H^{N}$, $^{1}H^{\alpha}$, ^{15}N, $^{13}C^{\alpha}$, ^{13}CO) as follows:

$$A\ (\text{ppm}) = \delta(\text{G1}) - \delta(\text{G1}_{\text{ref}}) \quad (1)$$

$$B\ (\text{ppm}) = \delta(\text{G2}) - \delta(\text{G2}_{\text{ref}}) \quad (2)$$

$$C\ (\text{ppm}) = \delta(\text{G3}) - \delta(\text{G3}_{\text{ref}}) \quad (3)$$

$$D\ (\text{ppm}) = \delta(\text{G4}) - \delta(\text{G4}_{\text{ref}}) \quad (4)$$

where δ(G1), δ(G2), and so forth, refer to the chemical shift observed for G1, G2, and so forth, in the peptide Ac-GGXGG-NH_2 and $\delta(\text{G1}_{\text{ref}})$, $\delta(\text{G2}_{\text{ref}})$, and so forth, refer to the chemical shift of the corresponding nucleus in the peptide Ac-GGGGG-NH_2. Clearly, the most significant effects will be observed on the glycine residues immediately adjacent to the central residue X; that is, we expect that the correction factors *B* and *C* to be the most significant.

Application of the correction factors to random coil shifts to compensate for sequence dependence is done according to eq 5:

$$\begin{aligned}\delta_{\text{R}}(\text{corrected}) &= \delta_{\text{random}}(\text{R}) + \Delta\delta(\text{R}_{-1}) + \Delta\delta(\text{R}_{+1}) + \Delta\delta(\text{R}_{-2}) + \Delta\delta(\text{R}_{+2}) \\ &= \delta_{\text{random}}(\text{R}) + B + C + A + D \end{aligned} \quad (5)$$

where R is the residue of interest, R_{-1} and R_{-2} are the preceding residues, and R_{+1} and R_{+2} are the residues following R in the peptide $R_{-2}-R_{-1}-R-R_{+1}-R_{+2}$. In practice, eq 5 can be reduced to eq 6 in most cases, since the correction factors *A* and *D* are small.

$$\delta_{\text{R}}(\text{corrected}) = \delta_{\text{random}}(\text{R}) + B + C \quad (6)$$

If we consider in the first approximation only the tripeptide sequence $R_{-1}-R-R_{+1}$, then the chemical shift of the residue of interest (R) has to be corrected by the correction factor *B* corresponding to the residue R_{+1}. This correction factor is obtained by determining the effect on the chemical shift of G2 in the peptide Ac-GGR_{+1}GG-NH_2. Similarly, R has to be corrected by the correction factor *C* corresponding to the residue R_{-1}, which is obtained by determining the effect on the chemical shift of G4 in the peptide Ac-GGR_{-1}GG-NH_2. For example, the random coil chemical shift of serine in the tripeptide fragment -Arg-Ser-Thr- has to be corrected by the correction factor B obtained from the peptide Ac-GGTGG-NH_2 (effect of Thr on the preceding residue G2) and by the correction factor C obtained from the peptide Ac-GGRGG-NH_2 (effect of Arg on the following residue G4).

Implementation into NMRView. The calculations were carried out using a modified version of the CSI-module for NMRView.[19] The original treatment for chemical shift index sequence correction has been extended to allow for corrections involving all four factors (*A*, *B*, *C*, and *D*) and implemented into a recently introduced improved CSI interface[20] for NMRView. CSI calculations on the data sets *Wishart-peptides*, *Wüthrich*, and *urea* are now automatically corrected for local sequence-dependent effects. The varying amount of information available for each of these data sets required differences in the implementation of the corrections. In case of the data set *urea* (Schwarzinger et al.[20] and present work) all four correction factors were utilized. Generally, we correct for the preceding and the following residue of all nuclei (except H^{α} where we only correct for effects caused by the aromatic residues on the direct neighbors). In certain cases (e.g., proline) we correct for long-range effects. The correction factors determined in the present work are used as sole source of the sequence corrections. Sequence corrections for the data set *Wishart-peptides* are possible for residues followed by proline[12] and for the effect of the preceding residue on the nuclei H^{N} and N (present work). The later correction employs the same algorithm that is used in the *urea* data set. Residues followed by proline take the random coil shifts published for the Ac-GGXPGG-

(19) Johnson, B. A.; Blevins, R. A. *J. Chem. Phys.* **1994**, *29*, 1012−1014.

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

(21) Jennings, P. A.; Stone, M. J.; Wright, P. E. *J. Biomol. NMR* **1995**, *6*, 271−276.

(22) Van Geet, A. L. *Anal.Chem.* **1970**, *42*, 679−680.

(23) 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.

Table 1. Assignments and Sequence-Dependent Chemical Shift Correction Factors for $^1H^\alpha$

	Observed Chemical Shifts[a]				Calculated Correction Factors[b]			
Residue	G1-H^α	G2-H^α	G4-H^α	G5-H^α	*A*-H^α	*B*-H^α	*C*-H^α	*D*-H^α
Ala	3.97	4.00	3.99	3.95	-0.02	-0.03	-0.03	0.00
Arg	3.97	4.01	4.00	3.95	-0.02	-0.02	-0.02	0.00
Asn	3.98	4.02	4.00	3.94	-0.01	-0.01	-0.02	-0.01
Asp	3.97	4.02	4.00	3.94	-0.02	-0.01	-0.02	-0.01
Cys	3.98	4.05	4.02	3.95	-0.01	0.02	0.00	0.00
Gln	3.98	4.01	4.01	3.95	-0.01	-0.02	-0.01	0.00
Glu	3.97	4.01	4.00	3.95	-0.02	-0.02	-0.02	0.00
Gly	3.99	4.03	4.02	3.95	(0)[c]	(0)	(0)	(0)
His	3.96	3.97	4.03	3.96	-0.03	-0.06	0.01	0.01
Ile	3.96	4.01	4.00	3.94	-0.03	-0.02	-0.02	-0.01
Leu	3.95	4.00	3.97	3.94	-0.04	-0.03	-0.05	-0.01
Lys	3.97	4.01	4.01	3.95	-0.02	-0.02	-0.01	0.00
Met	3.97	4.02	4.01	3.95	-0.02	-0.01	-0.01	0.00
Phe	3.93	3.94	3.94	3.91	-0.06	***-0.09***[d]	***-0.08***[d]	-0.04
Pro	3.98	4.14	3.99	3.94	-0.01	**0.11**[d]	-0.03	-0.01
Ser	3.98	4.05	4.02	3.94	-0.01	0.02	0.00	-0.01
Thr	3.98	4.08	4.02	3.94	-0.01	0.05	0.00	-0.01
Trp	3.91	3.93	3.87	3.79	-0.08	***-0.10***[d]	***-0.15***[d]	***-0.16***[d]
Tyr	3.94	3.93	3.94	3.91	-0.05	***-0.10***[d]	***-0.08***[d]	-0.04
Val	3.97	4.02	4.00	3.94	-0.02	-0.01	-0.02	-0.01
Average					-0.03	-0.02	-0.03	-0.02

[a] $^1H^\alpha$ chemical shifts for the glycine residues in the peptide series Ac-GGXGG-NH_2, where X represents one of the 20 naturally occurring amino acids and G1, G2, G4, and G5 are the glycines at positions 1, 2, 4, and 5 of each peptide. The $^1H^\alpha$ and $^{13}C^\alpha$ resonances for G2 and G3 of Ac-GGGGG-NH_2 could not be resolved either in 1D spectra or in 2D $^1H-^{13}C$ HSQC spectra. [b] Calculated by subtraction of the corresponding $^1H^\alpha$ chemical shift value for the peptide Ac-GGGGG-NH_2 according to eqs 1−4. [c] Correction factors for Gly are by definition zero due to the method of calculation. [d] Unusually high sequence-dependent correction factors (≥0.08 ppm for $^1H^\alpha$) are identified as follows: for aromatic residues in ***bold-italic***, for hydrophobic residues in *italic*, for hydrophilic residues underlined, and for proline **bold**.

NH_2 peptide as a reference.[12] The data set *Wüthrich* allows corrections for the effect of the preceding residue on the amide nitrogen only.[11]

Results and Discussion

Resonance Assignments. Resonance assignments for the residues X in the 20 model peptides Ac-GGXGG-NH_2, comprising a set of reference set of random coil chemical shifts in 8 M urea at pH 2.3, have been published.[20] The present work utilizes complete assignments for the glycine residues G1, G2, G4, and G5, as a means of estimating corrections that need to be made to random coil chemical shifts to account for differences in neighboring amino acids in the primary sequence. Such corrections are likely to be particularly important in studies of unfolded and partly folded proteins, where correct estimates of residual secondary structure depend critically on the absolute values of the random coil shifts employed.

The resonance assignments for the glycine residues are presented in Tables 1−5 for the nuclei $^1H^\alpha$, $^{13}C^\alpha$, $^1H^N$, ^{15}N, and ^{13}CO, respectively. For each nucleus, the sequence-dependent chemical shift correction has been calculated using the reference chemical shifts obtained from the Ac-GGGGG-NH_2 peptide chemical shifts according to eq 1−4, and these values are also shown in Tables 1−5 (values for Cβ are not available from this type of model peptide). Average values for the correction factors *A*−*D* for each of the nuclei are also included in each table.

Effect of X on H^α and C^α. Tables 1 and 2 show that the effect of local sequence variations on $^1H^\alpha$ and $^{13}C^\alpha$ is quite small. Even the immediately adjacent residues are relatively undisturbed by changes in the identity of residue X. Thus, the random coil chemical shifts of these two nuclei depend mainly on the attached side chain, with little necessity for corrections based on the local amino acid sequence. This makes both H^α and C^α ideally suited for the estimation of secondary structure in proteins based on chemical shift analysis; because of its larger chemical shift dispersion, the C^α usually gives clearer results.

Table 2. Assignments and Sequence-Dependent Chemical Shift Correction Factors for C^{α}

	Observed Chemical Shifts[a]				Calculated Correction Factors[b]			
Residue	G1-C^{α}	G2-C^{α}	G4-C^{α}	G5-C^{α}	*A*- C^{α}	*B*- C^{α}	*C*- C^{α}	*D*- C^{α}
Ala	45.53	45.22	45.48	44.99	-0.02	*-0.17*[d]	0.06	0.01
Arg	45.55	45.32	45.41	45.00	0.00	-0.07	-0.01	0.02
Asn	45.49	45.36	45.65	44.99	-0.06	-0.03	0.23[d]	0.01
Asp	45.52	45.39	45.67	44.97	-0.03	0.00	0.25[d]	-0.01
Cys	45.52	45.32	45.52	44.97	-0.03	-0.07	0.10	-0.01
Gln	45.53	45.33	45.46	44.99	-0.02	-0.06	0.04	0.01
Glu	45.54	45.31	45.47	44.99	-0.01	-0.08	0.05	0.01
Gly	45.55	45.39	45.42	44.98	(0)[c]	(0)	(0)	(0)
His	45.50	45.30	45.44	44.99	-0.05	-0.09	0.02	0.01
Ile	45.48	45.19	45.41	45.00	-0.07	*-0.20*[d]	-0.01	0.02
Leu	45.54	45.29	45.45	45.00	-0.01	-0.10	0.03	0.02
Lys	45.54	45.28	45.40	45.00	-0.01	-0.11	-0.02	0.02
Met	45.55	45.49	45.36	44.99	0.00	0.10	-0.06	0.01
Phe	45.48	45.16	45.48	44.99	-0.07	***-0.23***[d]	0.06	0.01
Pro	45.33	**44.62**	45.44	45.02	**-0.22**[d]	**-2.0**[e]	0.02	0.04
Ser	45.55	45.31	45.55	44.98	0.00	-0.08	0.13	0.00
Thr	45.54	45.35	45.54	44.98	-0.01	-0.04	0.12	0.00
Trp	45.53	45.22	45.45	44.90	-0.02	***-0.17***[d]	0.03	-0.08
Tyr	45.48	45.17	45.48	44.97	-0.07	***-0.22***[d]	0.06	-0.01
Val	45.48	45.18	45.40	44.99	-0.07	*-0.21*[d]	-0.02	0.01
Average					-0.06	-0.12	0.06	0.01

[a] $^{13}C^{\alpha}$ chemical shifts for the glycine residues in the peptide series Ac-GGXGG-NH_2, where X represents one of the 20 naturally occurring amino acids and G1, G2, G4 and G5 are the glycines at positions 1, 2, 4 and 5 of each peptide. The $^{1}H^{\alpha}$- and $^{13}C^{\alpha}$ resonances for G2 and G3 of Ac-GGGGG-NH_2 could not be resolved either in 1D spectra or in 2D $^{1}H-^{13}C$ HSQC spectra. [b] Calculated by subtraction of the corresponding $^{13}C^{\alpha}$ chemical shift value for the peptide Ac-GGGGG-NH_2 according to eqs 1−4. [c] Correction factors for Gly are by definition zero due to the method of calculation. [d] Unusually high sequence-dependent correction factors (≥0.15 ppm for $^{13}C^{\alpha}$) are identified as follows: for aromatic residues in ***bold-italic***, for hydrophobic residues in *italic*, for hydrophilic residues underlined, and for proline **bold**. [e] In the single case of the B correction factor for the presence of Pro, the consistency of the correction factor calculation breaks down. The value of −0.77 ppm derived from the observed chemical shifts in the GGPGG peptide is too small. The value of −2.0 ppm was calculated using random coil data of Wishart et al.,[12] as described in the text.

Some amino acids show systematically higher than average values for the $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ correction factors (Tables 1, 2). In particular, the aromatic residues and proline appear to have extensive effects on all of the nuclei studied; these effects will be discussed in later sections. Small polar residues affect the C^{α} of the following residue (G4). The effect observed is strongest for the amino acids Asp and Asn (deviations of 0.25 and 0.23 ppm, compared to the average value of 0.06 ppm). Smaller effects can be observed for Ser (0.13 ppm) and Thr (0.12 ppm), but there is no effect for Glu, Gln, Lys, and Arg, presumably due to the longer distance between the affected backbone nuclei and the polar group of the side chain. Because the average correction factors for C^{α} and H^{α} are small relative to the cutoff used in CSI-calculations (generally 0.7 ppm for C^{α} [24] and 0.1 ppm for H^{α} [5]), corrections for sequence-specific effects on the chemical shift of C^{α} and H^{α} will generally not be necessary unless aromatic residues and prolines are involved.

Effect of X on $^{1}H^{N}$, ^{15}N, and ^{13}CO. The nuclei $^{1}H^{N}$, ^{15}N, and ^{13}CO form the peptide bond and are involved in hydrogen bonding either within the protein molecule or to the solvent. Hydrogen bonding, which contributes to the systematic deviations from random coil values in secondary structure elements, does not significantly contribute to random coil chemical shifts since unfolded polypeptides rarely engage in hydrogen bonding other than to the solvent. This is supported by findings of Plaxco et al.,[15] who detected only comparatively small differences in the random coil chemical shifts upon addition of high concentrations of a strong denaturant. Local sequence dependence arises for these nuclei because of their involvement in the peptide bond. The random coil chemical shift depends on the identity of both of the amino acids in the bond. We would therefore expect that

(24) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171−180.

Table 3. Assignments and Sequence-Dependent Chemical Shift Correction Factors for $^1H^N$

	Observed Chemical Shifts[a]				Calculated Correction Factors[b]			
Residue	G1-H^N	G2-H^N	G4-H^N	G5-H^N	*A*-H^N	*B*-H^N	*C*-H^N	*D*-H^N
Ala	8.28	8.40	8.48	8.26	-0.01	-0.05	0.07[d]	-0.10
Arg	8.29	8.43	8.56	8.30	0.00	-0.02	0.15	-0.06
Asn	8.28	8.42	8.54	8.29	-0.01	-0.03	0.13	-0.07
Asp	8.27	8.42	8.55	8.25	-0.02	-0.03	0.14	-0.11
Cys	8.29	8.43	8.61	8.29	0.00	-0.02	0.20	-0.07
Gln	8.28	8.43	8.56	8.30	-0.01	-0.02	0.15	-0.06
Glu	8.28	8.42	8.56	8.29	-0.01	-0.03	0.15	-0.07
Gly	8.29	8.45	8.41	8.36	(0)[c]	(0)	(0)	(0)
His	8.28	8.41	8.61	8.36	-0.01	-0.04	0.20	0.00
Ile	8.28	8.39	8.58	8.27	-0.01	-0.06	0.17	-0.09
Leu	8.29	8.42	8.55	8.28	0.00	-0.03	0.14	-0.08
Lys	8.29	8.42	8.55	8.30	0.00	-0.03	0.14	-0.06
Met	8.29	8.43	8.56	8.30	0.00	-0.02	0.15	-0.06
Phe	8.26	8.33	8.51	7.99	-0.03	***-0.12***[d]	0.10	***-0.37***[d]
Pro	8.25	8.27	8.60	8.24	-0.04	**-0.18**[d]	0.19	**-0.12**[d]
Ser	8.29	8.42	8.57	8.28	0.00	-0.03	0.16	-0.08
Thr	8.30	8.45	8.55	8.30	0.01	0.00	0.14	-0.06
Trp	8.21	8.32	8.45	7.74	-0.08	***-0.13***[d]	0.04	***-0.62***[d]
Tyr	8.25	8.34	8.50	7.94	-0.04	***-0.11***[d]	0.09	***-0.42***[d]
Val	8.28	8.40	8.58	8.28	-0.01	-0.05	0.17	-0.08
Average					-0.01	-0.05	0.15	0.14

[a] $^1H^N$ chemical shifts for the glycine residues in the peptide series Ac-GGXGG-NH_2, where X represents one of the 20 naturally occurring amino acids and G1, G2, G4, and G5 are the glycines at positions 1, 2, 4, and 5 of each peptide. [b] Calculated by subtraction of the corresponding $^1H^N$ chemical shift value for the peptide Ac-GGGGG-NH_2 according to eqs 1−4. [c] Correction factors for Gly are by definition zero due to the method of calculation. [d] Unusually high sequence-dependent correction factors (≥0.11 ppm for $^1H^N$) are identified as follows: for aromatic residues in ***bold-italic***, for hydrophobic residues in *italic*, for hydrophilic residues underlined, and for proline **bold**.

the amide proton and the amide nitrogen would be influenced most strongly by the preceding residue, while the carbonyl carbon should be influenced by the following residue. For $^1H^N$ and ^{15}N, large deviations in the chemical shifts are observed for glycine G4 as residue X is varied (correction factor *C* is 0.15 ppm on average for $^1H^N$ and 2.1 ppm on average for ^{15}N). These average values are equal to the cutoff employed in CSI calculations for the amide proton (0.15 ppm[19]) and exceed the cutoff value for the amide nitrogen (0.5 ppm[19]) by a factor of 4. We therefore caution that application of uncorrected random coil chemical shift values for these nuclei in CSI calculations may result in severely biased estimations of residual secondary structure content in unfolded or partially structured proteins.

The effect of residue X on the $^1H^N$ and ^{15}N chemical shifts of the preceding residue (correction factor *B*) is generally small compared to the effect on the following residue. For $^1H^N$ this contribution is on average −0.05 ppm, and only −0.23 ppm for ^{15}N. An exception is His, which has a strong effect on the preceding (G2) amide nitrogen, comparable to the effects arising from the aromatic residues. Interestingly, such an effect of His could not be observed for the amide proton of G2, which may be an indication that the histidine ring is acting as a hydrogen bond donor to the G2 amide nitrogen, without involvement of the amide proton.

Large deviations are found for the ^{13}CO chemical shifts of glycine G2 as X is varied, with an average value of −0.7 ppm (correction factor *B*). This value also exceeds the cutoff value used for ^{13}CO in CSI calculations (0.5 ppm[24]). Biased estimations of the secondary structure content of proteins will result if these values are not corrected for sequence dependence. It should be noted that in the case of ^{13}CO, the correction factor *C* also has a small but significant contribution of −0.18 ppm on average, in contrast to the behavior of the amide nuclei, where the following residue has an effect with the opposite sign.

Influence of Proline. Large deviations in the chemical shifts of residue G2 are observed for all nuclei except ^{15}N when residue X is Pro (Tables 1−5). In addition, the ^{15}N and $^1H^N$ nuclei of residue G4 and the $^{13}C^\alpha$ and ^{13}CO of G1 are affected.

Table 4. Assignments and Sequence-Dependent Chemical Shift Correction Factors for ^{15}N

	Observed Chemical Shifts[a]				Calculated Correction Factors[b]			
Residue	G1-^{15}N	G2-^{15}N	G4-^{15}N	G5-^{15}N	*A*-^{15}N	*B*-^{15}N	*C*-^{15}N	*D*-^{15}N
Ala	113.46	107.14	106.81	108.16	-0.12	-0.33	-0.57[d]	-0.15
Arg	113.52	107.33	109.00	108.25	-0.06	-0.14	1.62	-0.06
Asn	113.40	107.21	108.25	108.14	-0.18	-0.26	0.87	-0.17
Asp	113.46	107.27	108.24	108.02	-0.12	-0.20	0.86	-0.29
Cys	113.52	107.21	110.45	108.31	-0.06	-0.26	3.07	0.00
Gln	113.52	107.33	109.00	108.25	-0.06	-0.14	1.62	-0.06
Glu	113.52	107.27	108.89	108.19	-0.06	-0.20	1.51	-0.12
Gly	113.58	107.47	107.38	108.31	(0)[c]	(0)	(0)	(0)
His	113.46	106.92	109.06	108.48	-0.12	**-0.55**[d]	1.68	0.17
Ile	113.40	107.33	112.25	108.31	-0.18	-0.14	4.87	0.00
Leu	113.52	107.33	108.43	108.25	-0.06	-0.14	1.05	-0.06
Lys	113.52	107.27	108.95	108.25	-0.06	-0.20	1.57	-0.06
Met	113.52	107.27	108.95	108.25	-0.06	-0.20	1.57	-0.06
Phe	113.40	106.98	110.16	107.85	-0.18	***-0.49***[d]	2.78	***-0.46***[d]
Pro	113.40	107.15	108.25	108.14	-0.18	-0.32	0.87	-0.17
Ser	113.52	107.44	109.93	108.14	-0.06	-0.03	2.55	-0.17
Thr	113.52	107.44	110.16	108.19	-0.06	-0.03	2.78	-0.12
Trp	113.58	107.21	110.57	107.67	0.00	-0.26	3.19	***-0.64***[d]
Tyr	113.34	107.04	110.39	107.79	-0.24	***-0.43***[d]	3.01	***-0.52***[d]
Val	113.34	107.33	111.72	108.25	-0.24	-0.14	4.34	-0.06
Average					-0.11	-0.23	2.07	-0.16

[a] ^{15}N chemical shifts for the glycine residues in the peptide series Ac-GGXGG-NH_2, where X represents one of the 20 naturally occurring amino acids and G1, G2, G4, and G5 are the glycines at positions 1, 2, 4, and 5 of each peptide. [b] Calculated by subtraction of the corresponding ^{15}N chemical shift value for the peptide Ac-GGGGG-NH_2 according to eqs 1−4. [c] Correction factors for Gly are by definition zero due to the method of calculation. [d] Unusually high sequence-dependent correction factors (≥0.4 ppm for ^{15}N) are identified as follows: for aromatic residues in ***bold-italic***, for hydrophobic residues in *italic*, for hydrophilic residues underlined, and for proline and histidine **bold**.

These effects can be attributed to the absence of an amide proton and the reduced conformational degrees of freedom of the proline ring. The effects caused by proline are larger than the CSI-cutoff value for all five nuclei, making correction highly desirable. The effect can still be detected on the carbonyl group of the N-terminal acetyl protecting group, which resonates 0.19 ppm upfield compared to the reference peptide Ac-GGGGG-NH_2. A correction is clearly indicated for the presence of a proline on the $^{13}C^{\alpha}$ and ^{13}CO chemical shifts of the preceding two residues (corrections *A* and *B*), and is probably advisable for $^{1}H^{\alpha}$ and $^{1}H^{N}$ for the residues immediately preceding and following proline (correction factors *B* and *C*).

The Gly-Pro Correction. In the particular case of the C^{α} correction *B* for the presence of Pro, the excellent consistency of the systematic corrections described here breaks down. In the course of analysis of the chemical shift deviations from random coil of several sets of unfolded proteins, it was noticed that the application of the sequence corrections to the C^{α} chemical shifts gave deviations for the residues preceding prolines that was greater than the average, by about 1.2 ppm. This deviation is apparently caused by the unique properties of the sequence Gly-Pro, which is present in the peptide GGPGG for which the measurements leading to the sequence corrections were made. It appears that, in this sequence, there is a conformational preference for a particular local structure with a backbone conformation that induces an anomalous value of the C^{α} chemical shift of Gly 2. This is confirmed by consideration of the data of Wishart et al.,[12] which shows tabulations of random coil chemical shifts for peptides of the sequence GGXAGG and GGXPGG. The C^{α} random coil chemical shifts from the Wishart paper for these two series of peptides are shown in Table 6. It is noticeable that the difference between the two values (δ_{RC}(GGXAGG) − δ_{RC}(GGXPGG)) is close to 2.0 ppm for all residues except glycine, where the value is 0.6 ppm. The value of the *B* correction factor has been modified in Table 2 to reflect this apparently real anomaly in this single case. The value present in Table 2 was derived from the average

Table 5. Assignments and Sequence-Dependent Chemical Shift Correction Factors for ^{13}CO

	Observed Chemical Shifts[a]				Calculated Correction Factors[b]			
Residue	G1-^{13}CO	G2-^{13}CO	G4-^{13}CO	G5-^{13}CO	*A*-^{13}CO	*B*-^{13}CO	*C*-^{13}CO	*D*-^{13}CO
Ala	175.14	174.20	174.67	176.88	-0.11	-0.77[d]	-0.07	-0.02
Arg	175.19	174.48	174.55	176.87	-0.06	-0.49	-0.19	-0.03
Asn	175.16	174.31	174.64	176.87	-0.09	-0.66	-0.10	-0.03
Asp	175.17	174.39	174.61	176.86	-0.08	-0.58	-0.13	-0.04
Cys	175.17	174.46	174.46	176.83	-0.08	-0.51	-0.28	-0.07
Gln	175.20	174.49	174.56	176.87	-0.05	-0.48	-0.18	-0.03
Glu	175.16	174.49	174.54	176.87	-0.09	-0.48	-0.20	-0.03
Gly	175.25	174.97	174.74	176.90	(0)[c]	(0)	(0)	(0)
His	175.15	174.32	174.52	176.83	-0.10	-0.65	-0.22	-0.07
Ile	175.05	174.39	174.56	176.88	-0.20	-0.58	-0.18	-0.02
Leu	175.12	174.47	174.61	176.89	-0.13	-0.50	-0.13	-0.01
Lys	175.17	174.47	174.56	176.87	-0.08	-0.50	-0.18	-0.03
Met	175.17	174.56	174.56	176.88	-0.08	-0.41	-0.18	-0.02
Phe	174.98	174.14	174.49	176.80	-0.27	-0.83	-0.25	-0.10
Pro	174.78	172.13	174.65	176.88	**-0.47**[d]	-2.84	-0.09	-0.02
Ser	175.17	174.57	174.59	176.84	-0.08	-0.40	-0.15	-0.06
Thr	175.17	174.78	174.61	176.85	-0.08	-0.19	-0.13	-0.05
Trp	174.99	174.12	174.44	176.73	-0.26	-0.85	-0.30	-0.17
Tyr	174.97	174.12	174.50	176.77	-0.28	-0.85	-0.24	-0.13
Val	175.05	174.40	174.56	176.87	-0.20	-0.57	-0.18	-0.03
Average					-0.15	-0.71	-0.18	-0.05

[a] ^{13}CO chemical shifts for the glycine residues in the peptide series Ac-GGXGG-NH_2, where X represents one of the 20 naturally occurring amino acids and G1, G2, G4, and G5 are the glycines at positions 1, 2, 4, and 5 of each peptide. [b] Calculated by subtraction of the corresponding ^{13}CO chemical shift value for the peptide Ac-GGGGG-NH_2 according to eqs 1−4. [c] Correction factors for Gly are by definition zero due to the method of calculation. [d] Unusually high sequence-dependent correction factors (≥0.4 ppm for ^{13}CO) are identified as follows: for aromatic residues in ***bold-italic***, for hydrophobic residues in *italic*, for hydrophilic residues underlined, and for proline **bold**.

of the differences δ_{RC}(GGXAGG) − δ_{RC}(GGXPGG), a value of 2.0, with a standard deviation of 0.2 (Table 6).

Influence of Aromatic Residues. Relatively long-range effects caused by aromatic residues have been reported, particularly for protons.[12,13] Ring-current effects may be pronounced for protons that spend significant proportions of time close to aromatic rings, and this is observed for the series Ac-GGXGG-NH_2. The effect is strongest for Trp, and is of about similar magnitude for Phe and Tyr. For the H^{α} nucleus an influence of the aromatic residue on both neighboring glycine residues can be observed (Table 1), which is close to or even exceeding the CSI cutoff value of 0.1 ppm. For Trp a correction factor *D* of −0.15 ppm is found for the C-terminal residue (G5), which is 4 times larger than the deviations for Phe and Tyr found for this residue. This is presumably due the larger size of the indole ring of tryptophan. An effect on both neighboring residues can also be found for $^1H^N$ (Table 3). The effect on the preceding residue is of comparable size to the effect found in H^{α}, but the effect on the following residue is systematically lower for the aromatic residues than the average, an indication that the effect of the aromatic ring at position X is opposite to that for all of the other residues. In general, the correction factor *C* for $^1H^N$ is quite large and positive (see previous discussion), due to the presence of the peptide bond. For the aromatic residues the lower values of *C* could be due to electron withdrawal from the peptide bond by the aromatic ring.

For the amide nitrogen, correction factors *B* for the preceding residue of −0.49, −0.26 ppm, and −0.43 ppm are found for Phe, Trp, and Tyr. While the strongest effect is usually observed on Trp and the weakest effect on Phe, the opposite is found here. In contrast to H^N, the effects show the same sign as that of the remaining correction factors. In addition, they are all larger than the average value obtained for *B*. Both amide nuclei, $^1H^N$ and ^{15}N, display a strong influence of the aromatic residue on the C-terminal residue (G5) (correction factor *D*). The effect is strongest for Trp (−0.62 ppm for H^N, −0.64 ppm for N), and weaker for Phe (−0.37 ppm HN, −0.46 ppm N). These large deviations suggest that a correction is necessary. However, when the sequence correction *D* is applied to urea unfolded apomyoglobin, it does not improve the result (data not shown).

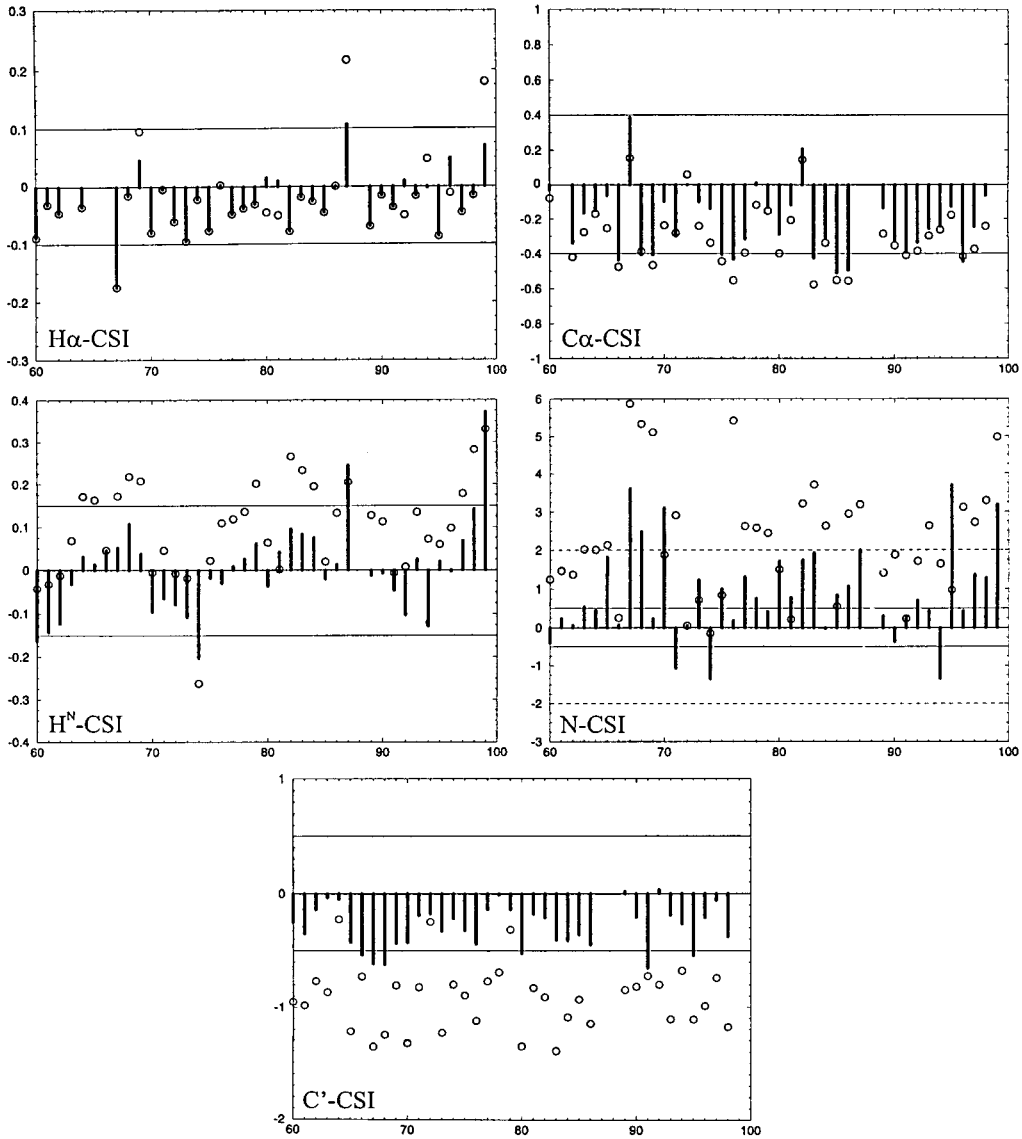

Figure 2. Application of the sequence correction factors for the preceding and the following residue to the chemical shifts of urea unfolded sperm whale apomyoglobin. A representative sequence ranging from amino acid 60 to 100 is shown here. Circles mark CSI calculations obtained without sequence correction, while bars represent CSI calculations with sequence corrections.

This may be due to the fact that short peptides and protein chains behave differently. In particular, the peptides used for determining chemical shifts are optimized to be highly flexible and to adopt the least amount of structure possible. It seems that the C-terminal residue can flip under the aromatic ring, giving a close contact between the aromatic ring and the amide of the residue two positions after it. This has been observed in peptides[13,25,26] and has been attributed to a hydrogen-bonding interaction between the amide proton and the aromatic ring.[26]

The aromatic residues also have an effect on the ^{13}C resonances of all four glycine residues. We showed that ^{13}CO nuclei should in general be corrected for the preceding residue (correction factor *B*). The most significant corrections for the following residue (correction factor *C*) are for the aromatic residues, and the N-terminal residue is comparably affected (correction factor *A*). The effect found on the C-terminal residue is comparatively small (correction factor *D*: −0.17 ppm for Trp to −0.1 ppm for Phe compared to the average value of −0.05 ppm). These effects are comparable in magnitude to effects caused by other hydrophobic amino acids. Correction for these rather small (compared to the CSI cutoff values) long-range effects is not recommended.

Comparison to Other Data Sets. Two previous studies have presented data which may be compared with the present data. Braun et al.[11] investigated the effect of the sequence on the ^{15}N chemical shift of the alanine in unprotected H-GGXA-OH model peptides, while Wishart et al.[12] published a list of random coil shifts for alanine when preceded by each of the 20 naturally occurring amino acids, from which sequence-dependent correction factors similar to those of Tables 1−5 can be calculated for $^{1}H^{\alpha}$, $^{1}H^{N}$, $^{13}C^{\alpha}$, ^{13}CO, and ^{15}N. Table 7 shows a comparison of the ^{15}N correction factors from these two data sets with those derived from the present data. In general there is good agreement in the trends between measurements. Slight differences between the data sets could be the result of a variety of factors, such as the differences in temperature or pH, the presence of different concentrations of urea, the presence or absence of protecting groups on the peptides or the known tendency of the H-GGXA-OH peptides to adopt preferred conformations.[11]

On the basis of the three sets of correction factors obtained on three different peptides, an average correction factor for the

(25) Dyson, H. J.; Sayre, J. R.; Merutka, G.; Shin, H.-C.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992**, *226*, 819−835.

(26) Kemmink, J.; Creighton, T. E. *J. Mol. Biol.* **1993**, *234*, 861−878.

Table 6. Random Coil Values for C^{α} from Peptides GGXAGG and GGXPGG[12]

residue	$\delta_{RC}(C^{\alpha})$ (GGXAGG)	$\delta_{RC}(C^{\alpha})$ GGXPGG)	δ_{RC}(GGXAGG) − δ_{RC}(GGXPGG)
Ala	52.5	50.5	2.0
Arg	56.0	54.0	2.0
Asn	53.1	51.3	1.8
Asp	54.2	52.2	2.0
Cys	58.2	56.4	1.8
Gln	55.7	53.7	2.0
Glu	56.6	54.2	2.4
Gly	45.1	44.5	0.6
His	55.0	53.3	1.7
Ile	61.1	58.7	2.4
Leu	55.1	53.1	2.0
Lys	56.2	54.2	2.0
Met	55.4	53.3	2.1
Phe	57.7	55.6	2.1
Pro	63.3	61.5	1.8
Ser	58.3	56.4	1.9
Thr	61.8	59.8	2.0
Trp	57.5	55.7	1.8
Tyr	57.9	55.8	2.1
Val	62.2	59.8	2.4
average difference[a]			2.0 (1.9)
standard deviation[a]			0.2 (0.4)

[a] Values excluding those for glycine (values in parentheses are the corresponding values including those for glycine).

Table 7. Comparison of Correction Factors *C* for ^{15}N, Derived from Three Data Sets

residue	set 1[a] (present)	set 2[b] (Braun)	set 3[c] (Wishart)	average (sets 1−3)	standard deviation
Ala	−0.6	−0.2	−0.8	−0.5	0.3
Arg	1.6	2.3	1.4	1.8	0.5
Asn	0.9	0.7	0.7	0.8	0.1
Asp	0.9	0.8	0.8	0.8	0.0
Cys	3.1	2.0	2.7	2.6	0.5
Gln	1.6	2.1	1.3	1.7	0.4
Glu	1.5	2.0	1.2	1.6	0.4
Gly	(0)[d]	0.0	(0)	(0)	(0)
His	1.7	2.8	1.8	2.1	0.6
Ile	4.9	4.5	4.2	4.5	0.3
Leu	1.1	1.5	1.0	1.2	0.3
Lys	1.6	2.1	1.6	1.8	0.3
Met	1.6	2.1	1.1	1.6	0.5
Phe	2.8	2.3	2.4	2.5	0.3
Pro	0.9	0.6	0.4	0.6	0.2
Ser	2.6	2.0	1.9	2.2	0.4
Thr	2.8	3.2	2.4	2.8	0.4
Trp	3.2	2.3	2.8	2.8	0.4
Tyr	3.0	2.3	2.8	2.7	0.4
Val	4.3	4.2	3.9	4.1	0.2
average	2.1	2.0	1.8	2.0	0.3
σ	1.3	1.2	1.2	1.2	0.1

[a] For peptides Ac-GGXGG-NH_2, 8 M urea, pH 2.3, 20 °C (see Table 4). [b] For peptides H-GGXA-OH, pH 2.0, 35 °C.[11] [c] For peptides Ac-GGXAGG-NH_2, 1 M urea, pH 5.0, 25 °C.[12] [d] Correction factors for Gly are by definition zero due to the method of calculation.

effect of X on the chemical shift of the following residue was calculated (Table 7). The standard deviation resulting from these three data sets gives an estimate of the uncertainty associated with CSI calculations employing N. In general, it appears that these standard deviations are small compared to the size of the correction factor, an excellent validation of the methodology. The same trends in sign and magnitude could be observed on all three data sets, indicating that correction factors for sequence-dependent influences on the random coil chemical shift derived by the observation of a effect of a varying residue on a single amino acid can be applied to any combination of amino acids. This assumption is valid over a wide variety of experimental conditions.

Application of Sequence-Specific Correction Factors to a Denatured Protein. The newly derived correction factors were applied to urea-denatured apomyoglobin (8 M urea, pH 2.3, and 20 °C). Under these conditions, apomyoglobin is largely denatured, with very little residual structure. Nevertheless, the chemical shift index for the protein shows a number of systematic deviations from standard random coil values. Sequence-dependent corrections derived from Tables 1−5 were applied to these data in order to evaluate the source of the systematic deviations in chemical shift. Chemical shift data for a representative portion of the sequence of apomyoglobin (residues 60−100) is shown in Figure 2. The results for CSI calculations without sequence-dependent corrections are shown as circles, while the values corrected for the effect of the preceding and the following residue are shown as bars. This treatment improves the quality of the CSI values, judged by the removal of systematic effects, especially for ^{15}N and ^{13}CO. We find that correction for long-range effects, for example, the effects of aromatic residues on $^{1}H^{N}$, discussed previously, did not improve the accuracy of the result. Figure 2 clearly shows that even in the case of H^{α} and C^{α}, which are not normally thought to be very dependent on sequence, the accuracy of the chemical shift index can be dramatically improved by the local sequence-dependence correction, and systematic overestimation of secondary structure elements can be prevented.

Conclusions

We have presented a complete analysis of the effect of variations in the amino acid sequence on random coil chemical shifts. The derived correction factors were compared to data obtained from other peptides showing a generally good agreement. The application of these correction factors to urea unfolded apomyoglobin clearly demonstrates the necessity for sequence corrections in the case of unstructured proteins, the study of which is of growing importance for understanding of protein-folding mechanisms. In the case of folded proteins the accuracy of, for example, prediction of the boundaries of secondary structure elements by the chemical shift index method ought to be improved by correction for sequence, especially where large corrections due to proline or aromatics are involved. More importantly, sequence-dependent corrections of secondary shifts will undoubtedly increase the accuracy of protein backbone angle restraints obtained from database searching.[27] Finally, the detection of residual local structure in otherwise unfolded parts of proteins will be more reliable. Finally, the presented data may serve as a basis for computational studies, which can provide a better understanding of the factors governing the chemical shift.[28,29]

The improved CSI module for NMRView and a table with the correction factors determined in this study is available for download as apart of the TSRI contribution for NMRView (http://garbanzo.scripps.edu/nmrgrp/wisdom/pipe/tips_scripts.html).

Acknowledgment. This work was supported by Grant GM57374 from the National Institutes of Health. S.S. acknowledges an Erwin Schrödinger fellowship by the Austrian Science Fund (FWF Project J1736-CHE).

JA003760I

(27) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289−302.

(28) Szilágyi, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **1995**, *27*, 325−443.

(29) Case, D. A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 197−203.