Journal of Biomolecular NMR **28:** 327–340, 2004.
KLUWER/ESCOM

Predicting ^{15}N chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from ϕ, ψ^{i-1}, and χ^1 torsion angles

Yunjun Wang[1,*] & Oleg Jardetzky*
Department of Molecular Pharmacology, Stanford University, Stanford, CA 94305-5174, U.S.A.
[1]*Current address: Nanomaterials and Nanofabrication Laboratories (NN-Labs), Fayetteville, AR 72704, U.S.A.*

Received 25 June 2003; Accepted 30 October 2003

Key words: chemical shift prediction, neighboring residue effect, NMR, protein, side-chain geometry effect

Abstract

Empirical shielding surfaces are most commonly used to predict chemical shifts in proteins from known backbone torsion angles, ϕ and ψ. However, the prediction of ^{15}N chemical shifts using this technique is significantly poorer, compared to that for the other nuclei such as $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$. In this study, we investigated the effects from the preceding residue and the side-chain geometry, χ^1, on ^{15}N chemical shifts by statistical methods. For an amino acid sequence XY, the ^{15}N chemical shift of Y is expressed as a function of the amino acid types of X and Y, as well as the backbone torsion angles, ϕ and ψ^{i-1}. Accordingly, 380 empirical 'Preceding Residue Specific Individual (PRSI)' ^{15}N chemical shift shielding surfaces, representing all the combinations of X and Y (except for Y=Pro), were built and used to predict ^{15}N chemical shift from ϕ and ψ^{i-1}. We further investigated the χ^1 effects, which were found to account for differences in ^{15}N chemical shifts by $\sim$ 5 ppm for amino acids Val, Ile, Thr, Phe, His, Tyr, and Trp. Taking the χ^1 effects into account, the χ^1-calibrated PRSI shielding surfaces (XPRSI) were built and used to predict ^{15}N chemical shifts for these amino acids. We demonstrated that ^{15}N chemical shift predictions are significantly improved by incorporating the preceding residue and χ^1 effects. The present PRSI and XPRSI shielding surfaces were extensively compared with three recently published programs, SHIFTX (Neal et al., 2003), SHIFTS (Xu and Case, 2001 and 2002), and PROSHIFT (Meiler, 2003) on a set of ten randomly selected proteins. A set of Java programs using XPRSI shielding surfaces to predict ^{15}N chemical shifts in proteins were developed and are freely available for academic users at http://www.pronmr.com or by sending email to one of the authors Yunjun Wang (yunjunwang@yahoo.com).

Introduction

For several decades, NMR spectroscopists have been attempting to understand chemical shifts in proteins. An important step toward this goal is to predict chemical shifts in proteins from known structures. To date, several fundamentally different techniques have been developed for this purpose, including semiempirical methods (Ösapay and Case, 1991; Williamson and Asakura, 1993), empirical ϕ/ψ shielding surfaces (Spera and Bax, 1991; Le and Oldfield 1994; Wishart and Nip, 1998), *ab initio* quantum mechanical calculations (de Dios et al., 1993; Pearson et al., 1997; Havlin et al., 1997 and 2001), and sequence homology based approaches (Wishart et al., 1997; Xu and Case, 2001). Among them, the empirical ϕ/ψ shielding surface is probably the most practical approach, which allows $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ shifts to be predicted with reasonable accuracy. However, application of this method to ^{15}N chemical shifts usually yields considerable discrepancies between predicted and observed values, indicating a need to identify other parameters influencing ^{15}N chemical shifts in proteins.

*To whom correspondence should be addressed. E-mails: yunjunwang@yahoo.com, jardetzky@stanford.edu

In addition to the well-studied backbone conformational effects, the preceding residue and side-chain geometry have also been known for a long time to have a significant influence on ^{15}N chemical shifts (Kricheldorf, 1981; Glushka et al., 1989; de Dios et al., 1993). However, the separate identification and quantitative determination of each individual contribution to the overall observed chemical shifts has proved to be very difficult. Consequently, quantitative measurements of the preceding residue effects are typically limited to the random coil state using short model peptides – e.g., AcGGXGG-HN_2 studied under denaturing experimental conditions (Braun and Wüthrich, 1994; Wishart et al., 1995; Schwarzinger et al., 2001). The parameters obtained from the short model peptides cannot be applied to folded proteins for chemical shift prediction, secondary structure identification, or structural refinement. For the same reason, some earlier *ab initio* quantum chemical shift investigations for the χ^1 effects on ^{15}N chemical shifts have not been successful (de Dios et al., 1993; Havlin et al., 2001).

With the rapid growth of the protein chemical shift and three dimensional structure databases, statistical analysis is playing a more and more important role in deciphering the relationship between the observed chemical shift and protein structure. In the past several years, secondary structure effects (Wishart et al., 1991; Wang and Jardetzky, 2002a) correlation of the chemical shift with backbone torsion angles (Spera and Bax, 1991; Le and Oldfield, 1994), the hydrogen-bond and side-chain geometry effects (Iwadate et al., 1999), contributions of specific structural features, such as the helix capping box (Gronenborn and Clore, 1994) and β-hairpin (Santiveri et al., 2001) have been successfully identified from the empirical chemical shift database. Recently, we investigated the nearest neighboring residue effects on $^1H^{\alpha}$, ^{15}N, $^1H^N$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts, and reported that the preceding residue effects on the ^{15}N shift depend on both the amino acid type of the preceding residue and on the secondary structural status (β-strand, random coil, or α-helix) for the folded proteins (Wang and Jardetzky, 2002b). We have also shown that the preceding residue effect on ^{15}N chemical shifts in protein is traced on electronic induction. Our initial motivation of this study is to improve the ^{15}N chemical shifts prediction by incorporating the preceding residue effect in the empirical shielding surfaces. Based on our previous work, for an amino acid sequence XY, we propose that ^{15}N shift of Y is dominated by four independent variables–the amino acid type of X, the amino acid type of Y, phi torsion angle of Y(ϕ), and psi torsion angle of X (ψ^{i-1}). Accordingly, we construct 380 Preceding Residue Specific Individual (PRSI) ϕ/ψ^{i-1} shielding surfaces for the prediction of ^{15}N chemical shifts. Compared to the currently used ϕ/ψ^{i-1} shielding surfaces, which do not specify the preceding residue (hereinafter called non-PRSI), PRSI gives significant improvements in ^{15}N chemical shift prediction.

In this study, we also statistically determined another factor that has a significant influence on ^{15}N chemical shifts in proteins– the side-chain geometry (χ^1). We demonstrated that ^{15}N chemical shift prediction can be further improved by incorporating the χ^1 effect into the shielding surface for amino acids Val, Ile, Thr, Phe, His, Tyr, and Trp.

^{15}N chemical shifts in proteins span a range of up to nearly 40 ppm and contain very rich structural information. Unlike other nuclei such as $^1H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$, the theoretical interpretation of ^{15}N chemical shifts has never been sufficiently accurate to allow the information it contained to be used for purposes of structure determination. We believe this work will help to understand the origins of ^{15}N chemical shifts in proteins, and will eventually lead to the application of ^{15}N chemical shifts in secondary structure identification, and the tertiary structure determination and refinement.

Methods

Preparation of chemical shift database. Assigned ^{15}N chemical shifts were downloaded from BioMagResBank (BMRB; http://www.bmrb.wisc.edu) and those meeting the following criteria were selected. (1) The length of protein sequence > 50. (2) The most commonly used materials, DSS, TMS, TSP, and liquid NH_3 were used as either direct or indirect ^{15}N chemical shift reference (Wishart et al., 1995). 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 a ^{15}N chemical shift, and excluded. The assignments of the very first two N-terminal and last two C-terminal residues of each protein were also excluded to avoid terminal effects on ^{15}N chemical shifts. The 3d-coordinates of the selected protein were downloaded from Protein Data Bank (http://www.rcsb.org/pdb),

from which torsion angles ϕ, ψ^{i-1} and χ^1 were calculated, and then combined with assigned ^{15}N chemical shifts to form a database, $\delta(\phi, \psi^{i-1})$. As noticed by several earlier studies (Iwadate et al., 1999; Wishart and Case, 2001, Zhang et al., 2003), improper ^{15}N chemical shift referencing is one of the problems associated with the protein ^{15}N chemical shift assignments deposited in BMRB. A recent study by Zhang et al. (2003) shows that a significant amount (27%) of BMRB entries with protein ^{15}N chemical shift assignments required significant (>1.0 ppm) reference readjustments. In this study, each protein was carefully checked for improper ^{15}N chemical shift reference, and reference readjustments were made accordingly using the protocols by Zhang et al. (2003). As an external check, we compared the reference corrections calculated in this study with those by SHIFTX (Zhang et al., 2003) for a set of fifteen randomly selected proteins. The averaged difference between the two sets of calculated reference correction is less than 0.5 ppm–that is statistically insignificant for ^{15}N chemical shift in proteins. The resulting database, $\delta(\phi, \psi^{i-1})$, is composed of 54239 reference-adjusted ^{15}N chemical shift entries derived from 511 distinct non-paramagnetic proteins. Each entry contains the assigned ^{15}N chemical shift, torsion angles ϕ, ψ^{i-1}, and χ^1, the amino acid types of the preceding residue (X) and the target residue (Y). A specific identification number was also generated and assigned to each entry. Subsets of the database to be used for comprising different types of the ^{15}N-shielding surface as described below, were generated from this database.

Non-PRSI database. The above prepared ^{15}N chemical shift database, $\delta(\phi, \psi^{i-1})$, was split into 19 subsets, $\delta Y(\phi, \psi^{i-1})$, based on the amino acid type of the target residue Y (except for Y = Pro). During the calculation of ^{15}N chemical shifts using non-PRSI database, the large number of chemical shift entries in each of $\delta Y(\phi, \psi^{i-1})$ were trimmed to 500.

PRSI database. Similarly, 380 subset databases, $\delta^X Y(\phi, \psi^{i-1})$, representing all the combinations of 20 amino acid types of X and Y (except for Y = Pro), were generated from $\delta(\phi, \psi^{i-1})$ based on the amino acid types of X and Y.

Preceding residue-combined PRSI database. Despite our efforts to collect all assigned protein chemical shifts as well as the corresponding 3d structures, the database we have built in this study is still not large enough to build three dimensional PRSI (ϕ, ψ^{i-1}, χ^1) shielding surfaces. Overcoming this problem, we built the preceding residue (X) combined databases, $\delta^{X-com}Y(\phi, \psi^{i-1})$, to investigate the side-chain geometry (χ^1) effects on ^{15}N chemical shift. The 20 amino acids of the preceding residues were clustered into eight groups, X = {Val, and Ile}; {Asp, and Asn}; {Phe, Tyr, His, and Trp}; {Leu, Met, Lys, Arg, Glu, and Asn}; {Thr, Cys, and Ser}; {Ala}; {Gly}; and {Pro}. Such clustering is based on our observations in this study as well as in our earlier work (Wang and Jardetzky, 2002b): the amino acids in the same group are very similar in their side-chain structures and show very similar effects on the ^{15}N shifts of the following residue. As a consequence, the 380 PRSI databases $\delta^X Y(\phi, \psi^{i-1})$ were grouped into $19 \times 8 = 152$ preceding residue (X) combined databases, $\delta^{X-com}Y(\phi, \psi^{i-1})$. Based on the χ^1 value, each $\delta^{X-com}Y(\phi, \psi^{i-1})$ database was further split into three χ^1 specific ones, $\delta^{X-com}Y\chi^1(\phi, \psi^{i-1})$, where $\chi^1 = 180 - 30°$ or $-180 + 30°$ (hereinafter referred to as $180 \pm 30°$), $60 \pm 30°$, or $-60 \pm 30°$.

To quantify the χ^1 effect, we define one of three staggered conformation of χ^1($180 \pm 30°$ for Val; $-60 \pm 30°$ for Ile, Thr, Phe, His, Tyr, and Trp) as the reference. For each amino acids, the staggered conformation we selected as the reference is the one with the largest population and energetically most stable. We propose that an observed ^{15}N chemical shift, $\delta^{obs}(\chi^1)$, can be written as:

$$\delta^{obs}(\chi) = \delta(\chi = \text{ref}) + \Delta(^X Y)\chi^1,$$

where $\Delta(^X Y)\chi^1$ is χ^1 effect correction factor. By definition, $\Delta(^X Y)\chi^1$ equals zero for the reference conformation, $\chi^1 = -60 \pm 30°$ ($180 \pm 30°$ for Val). In this study, the statistically averaged correction factor, $< \Delta(^X Y)\chi^1 >$, was calculated and used to represents the χ^1 effect. $< \Delta(^X Y)\chi^1 >$ was obtained by calculating the averaged difference between the observed ^{15}N chemical shift in the database $\delta^{X-com}Y\chi^1(\phi, \psi^{i-1})$ and that predicted using database $\delta^{X-com}Y\chi^1 = \text{ref}(\phi, \psi^{i-1})$.

XPRSI database. $380\chi^1$-calibrated PRSI (XPRSI) databases, $\delta^X Y\chi^1(\phi, \psi^{i-1})$, were generated from the corresponding $\delta^X Y(\phi, \psi^{i-1})$. More specifically, each of the ^{15}N chemical shift in the PRSI database, $\delta^X Y(\phi, \psi^{i-1})$, was calibrated with, $\Delta(^X Y)\chi^1$, if its χ^1 angle lie within any of the three above defined ranges, $180 \pm 30°$, $60 \pm 30°$, and $-60 \pm 30°$.

Chemical shift prediction. ^{15}N chemical shift prediction is performed by convoluting each of the observed values in the database of interest with a Gaussian function as suggested by Le and Oldfield (Le and Oldfield, 1994). For example, the ^{15}N shift of Y (preceded by X), $\delta^{X}Y(\phi,\ \psi^{i-1})$, was predicted from its ϕ, ψ^{i-1} angles by convoluting each of the chemical shifts in the corresponding PRSI database, $\delta_{n}^{X}Y(\phi_{n},\ \psi_{n}^{i-1})$.

$$\delta^{X}Y(\phi,\ \psi^{i-1}) = \frac{\sum_{n}\delta_{n}^{X}Y(\phi_{n},\psi_{n}^{i-1})\times\exp\{-[\sin^{2}(\frac{\phi_{n}-\phi}{2})+\sin^{2}(\frac{\psi_{n}^{i-1}-\psi^{i-1}}{2})]/A\}}{\sum_{n}\exp\{-[\sin^{2}(\frac{\phi_{n}-\phi}{2})+\sin^{2}(\frac{\psi_{n}^{i-1}-\psi^{i-1}}{2})]/A\}},$$

where A is a constant and is set to 0.03 unless otherwise indicated; the summations extend over all residues *n* in the corresponding PRSI database. When the XPRSI database, $\delta^{X}Y\chi^{1}(\phi,\ \psi^{i-1})$, is used for the prediction, the calculated shift from above Gaussian function, when applicable, is re-calibrated with the side-chain effect correction factor, $\Delta(^{X}Y)\chi^{1}$. During all the calculations, when a given chemical shift was computed, that shift itself was identified via its specifically assigned ID and was excluded during the convoluting calculation.

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

The preceding residue effect

As representative examples, four calculated PRSI shielding surfaces are shown in Figure 1. They demonstrate the variation of ^{15}N chemical shifts with the preceding residue as well as with the backbone torsion angles, ϕ and ψ^{i-1}. Shown on the left of this figure are the PRSI shielding surfaces for amino acid Glu preceded by Val (^{V}E) and Ala (^{A}E) respectively. The magnitudes of the ^{V}E surface are consistently higher than the one of ^{A}E with difference as large as nearly 8 ppm in certain areas (Figure 1C). These two surfaces, ^{V}E and ^{A}E, differ by $\sim$ 4–5 ppm in the β-strand region ($\psi^{i-1} = 90$–$180°$), and $\sim$ 2 ppm in the α-helical region ($\psi^{i-1} = -90$–$0°$). In our previous paper (Wang and Jardetzky, 2002b), we have statistically studied the preceding residue effect on ^{15}N chemical shift in proteins and shown that it is an electronic inductive effect, which, therefore, is determined by side chain structure of the preceding residue and the backbone dihedral angle ψ^{i-1}. Ala and Val differ greatly in their side-chain and on their electron withdrawing or donating properties. Consequently they have a different effect on the following residues. In the other hand, those amino acids with similar side chains, as for example, Val and Ile both with branched side chains at their C^{β} atoms, have a similar effect on the ^{15}N chemical shift of the following residue. To demonstrate this, two PRSI shielding surfaces for amino acid Asp preceded by Val (^{V}D) and Ile (^{I}D) are shown on the right of Figure 1. As shown in this figure, these two shielding surfaces, ^{V}D and ^{I}D, show a high degree of similarity. The difference in magnitude between the two surfaces is below 0.5 ppm in α-helical and β-strand regions – the most populated ϕ and ψ^{i-1} areas.

The PRSI shielding surfaces, which incorporate the preceding residue effects, significantly improve the accuracy of ^{15}N chemical shift prediction. To quantitatively determine the preceding residue effects, the ^{15}N chemical shift of each entry in the database, $\delta(\phi,\ \psi^{i-1})$, was calculated from its torsion angles, ϕ and ψ^{i-1}, using PRSI and non-PRSI surfaces respectively. The predicted ^{15}N shifts were compared with the observed values, and their differences were statistically analyzed based on the amino acid types of the preceding and the target residues. The average deviations between predicted and observed ^{15}N chemical shifts are listed in Table 1. As shown in this table, the ^{15}N chemical shifts predicted using PRSI surfaces show no significant differences (<0.15 ppm for most XY pairs) to the observed values for all the amino acid types of X and Y. On the other hand, when non-PRSI surfaces are used, systematic and large deviations were found for certain preceding residues. As highlighted in Table 1, regardless of the amino acid type of the target residue Y, the deviations are consistently higher (>1.0 ppm) when the preceding residue is Ile, Val, Ala, Thr, or Pro. The average (over all the 20 amino acids of Y) deviations between observed and predicted ^{15}N chemical shifts are plotted versus the amino acid type of the preceding residue in Figure 2.

Side-chain geometry (χ^{1}) effect

A major problem associated with the statistical investigation of the χ^{1} effect is insufficient chemical shift data. In this study, we construct the preceding residue combined database, $\delta^{X-com}Y(\phi,\ \psi^{i-1})$, to overcome this problem. We found that nearly 90%

Table 1. Averaged deviation between the observed ^{15}N chemical shifts and that predicted using PRSI and non-PRSI shielding surfaces respectively. The data are categorized based on the amino acid types of the preceding residue (X) and the target residue (Y)

X	I		V		D		N		F		Y		H		W		L		M	
Y	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PRSI	Non-PRSI	PSRI	Non-PRSI
A	0.00	−1.09	−0.27	−1.19	−0.05	−0.02	−0.04	0.21	−0.13	0.18	0.04	0.24	0.28	0.48	−0.29	−0.02	−0.08	0.62	0.13	−0.45
R	0.06	−0.78	−0.19	−1.37	−0.03	0.53	0.10	0.94	−0.16	0.60	−0.10	0.44	0.35	−0.04	0.06	0.35	0.07	0.33	−0.20	−0.54
N	0.16	−2.17	0.00	−1.95	0.17	0.92	−0.03	−0.22	0.02	−0.70	−0.15	−0.64	0.05	0.45	0.37	0.62	0.01	0.68	−0.13	0.11
D	0.14	−2.46	0.03	−1.98	0.04	0.07	−0.20	0.21	0.00	0.07	−0.09	−0.17	0.16	0.53	−0.22	0.91	0.06	0.48	−0.16	0.01
C	0.33	−1.20	0.29	−1.38	0.65	0.20	0.17	0.58	0.02	0.77	−0.57	−0.30	0.42	0.14	−1.75	5.63	0.75	1.87	0.86	0.32
Q	0.17	−1.25	0.10	−1.73	0.00	−0.04	0.15	1.20	−0.18	0.37	0.06	0.46	0.03	0.02	0.33	0.23	0.13	0.42	0.23	0.29
E	0.02	−1.82	0.13	−1.28	0.00	0.20	−0.05	0.21	−0.11	0.69	−0.08	0.18	−0.30	0.14	0.27	0.05	−0.04	0.69	0.34	−0.35
G	−0.13	−1.22	0.40	−0.84	0.14	0.41	0.28	1.00	0.05	0.69	0.29	0.79	0.71	0.51	−0.40	0.01	0.42	1.47	0.61	−0.62
H	0.48	−1.86	0.38	−1.76	0.23	0.82	0.23	1.18	0.66	0.52	0.13	−0.60	0.79	0.86	0.14	2.84	0.22	−0.27	−0.67	0.20
I	0.34	−2.12	0.10	−1.52	0.01	0.08	0.17	0.57	0.17	1.20	0.54	0.23	0.15	1.09	0.46	−0.54	−0.12	0.58	0.07	1.06
L	0.05	−1.02	0.04	−1.39	−0.02	−0.29	−0.06	−0.31	0.17	0.93	0.01	0.64	0.17	−0.07	0.11	−0.77	0.09	0.48	−0.15	0.29
K	−0.12	−1.82	0.14	−1.81	−0.02	0.70	0.08	0.38	0.06	−0.09	0.00	0.47	−0.12	−0.04	0.42	0.06	0.10	0.45	0.26	0.40
M	0.03	−1.04	0.58	−0.88	0.16	−0.47	−0.13	−1.18	0.05	0.18	0.03	0.54	−0.23	−0.54	0.19	0.96	0.14	0.48	0.45	0.59
F	0.24	−1.88	0.11	−1.21	−0.03	−0.17	0.08	0.07	−0.28	1.92	−0.16	1.00	0.61	0.18	0.31	−0.17	0.25	0.32	−0.65	0.40
S	0.01	−1.49	0.12	−1.82	0.13	0.51	−0.07	1.14	0.14	0.77	−0.24	0.49	−0.39	1.04	−0.10	−0.39	0.15	0.55	0.02	0.60
T	0.37	−2.31	0.20	−2.07	0.30	1.48	0.07	1.00	0.02	0.48	0.37	0.21	0.02	0.35	−0.21	−0.35	0.02	0.65	−0.07	0.41
W	−0.06	−2.08	−0.07	−1.49	0.05	0.80	0.37	0.51	−0.08	2.38	0.45	0.51	−1.21	1.71	1.35	2.50	0.18	0.07	0.00	0.75
Y	0.18	−1.72	0.14	−2.28	−0.08	0.46	0.25	0.14	−0.20	1.68	0.51	1.47	−0.30	2.08	0.01	0.81	0.21	0.89	1.47	0.87
V	0.07	−1.93	−0.22	−1.61	−0.16	0.84	−0.03	0.31	−0.03	0.88	0.25	0.95	0.26	−0.28	0.47	0.98	0.23	0.13	0.00	0.00
ave	0.12	−1.65	0.11	−1.56	0.08	0.37	0.07	0.42	0.01	0.71	0.07	0.36	0.08	0.45	0.08	0.72	0.15	0.57	0.13	0.23
X	R		K		Q		E		T		C		S		A		G		P	
A	0.04	0.37	0.07	0.35	−0.01	−0.08	0.00	0.04	−0.24	−1.44	−0.28	−0.89	−0.03	−1.43	−0.01	1.55	−0.10	−0.40	0.00	1.49
R	−0.06	0.02	−0.13	0.30	0.02	0.53	0.17	−0.10	0.01	−1.27	−0.22	−0.54	0.01	−1.47	−0.04	1.96	0.10	−0.04	0.05	1.30
N	0.08	0.23	0.07	−0.21	−0.02	0.54	−0.05	0.57	0.14	−1.37	0.39	0.31	0.02	−1.58	0.04	2.18	0.01	0.02	0.26	2.23
D	0.06	0.57	0.01	1.02	0.20	−0.38	0.07	0.65	−0.02	−1.29	0.75	−1.31	−0.02	−1.07	−0.02	1.90	0.08	−0.96	0.14	2.53
C	−0.03	0.95	0.43	−0.21	1.19	−0.31	0.29	0.97	0.15	−1.63	0.00	−0.67	0.70	−1.13	−0.12	2.63	0.23	0.03	−0.74	0.19
Q	0.21	0.60	0.16	−0.14	0.05	0.01	0.05	0.14	−0.04	−1.09	−0.07	−0.69	−0.06	−1.13	0.07	1.90	0.19	−0.30	−0.07	2.40
E	−0.03	0.59	0.07	0.59	0.16	−0.16	0.04	0.07	0.03	−1.32	0.13	−1.28	0.01	−1.21	0.00	1.22	0.00	−0.22	0.03	1.82
G	0.34	0.00	0.22	0.16	−0.01	0.34	−0.06	−0.47	0.10	−0.72	−0.21	−1.73	0.18	−0.66	0.45	2.41	0.09	1.07	−0.06	0.68
H	−0.20	1.10	0.13	0.56	0.22	1.11	−0.48	0.30	0.41	−1.78	−0.67	−1.08	0.18	−1.41	−0.08	2.14	0.02	−0.95	−0.41	1.00
I	0.10	0.47	−0.01	−0.63	−0.06	−0.42	0.08	−0.11	−0.04	−2.22	0.32	−1.66	−0.03	−0.52	−0.21	1.99	0.08	−0.20	−0.26	1.14
L	0.20	0.14	−0.15	−0.05	−0.04	−0.49	0.09	−0.53	0.08	−2.07	−0.06	−0.26	−0.03	−1.37	0.03	1.44	0.01	−0.54	0.03	1.47
K	−0.18	0.28	−0.08	−0.26	0.36	0.28	0.09	0.96	−0.01	−1.30	−0.12	−1.23	0.01	−0.64	0.05	1.86	0.04	−0.12	−0.04	1.77
M	0.05	−0.98	0.04	−0.60	−0.21	−0.33	0.02	−0.56	0.07	−0.50	0.72	−1.39	0.23	−0.90	−0.15	1.66	0.32	−0.35	0.19	1.88
F	−0.08	0.09	0.08	0.02	−0.02	−0.55	0.30	−0.26	0.17	−1.43	0.55	−0.70	−0.14	−0.43	0.16	2.12	0.03	0.36	−0.36	0.82
S	0.15	0.56	0.10	0.96	−0.24	−0.89	−0.11	0.83	0.14	−1.33	−0.09	−1.50	0.03	−1.12	−0.01	2.24	−0.02	0.65	−0.15	1.73
T	0.25	0.56	0.27	−0.35	0.12	0.60	0.25	0.19	0.02	−0.67	0.50	−2.15	−0.09	−0.44	0.12	1.96	0.03	−0.26	−0.29	2.24
W	0.56	0.31	−0.13	0.06	0.27	−0.77	0.29	−0.87	0.00	−2.00	0.01	−1.59	0.09	0.41	0.21	1.61	−0.57	−0.23	−0.60	2.79
Y	0.35	0.58	0.10	0.47	0.18	0.45	0.18	0.11	0.09	−1.34	0.40	0.54	−0.15	−0.67	0.02	2.11	0.01	0.18	0.10	0.78
V	0.02	0.35	0.21	−0.13	−0.13	0.52	0.24	0.03	0.14	−1.70	0.02	−0.23	0.07	−0.11	0.09	2.20	0.06	−0.23	−0.01	1.42
ave	0.10	0.36	0.08	0.10	0.11	0.00	0.08	0.10	0.06	−1.39	0.11	−0.95	0.05	−0.89	0.03	1.95	0.03	−0.13	−0.12	1.56

Table 2. Averaged side-chain effect correction factor, $< \Delta(^{X}Y)\chi^{1} >$. The data are categorized based on the amino acid types of the preceding residue (X, grouped) and the target residue (Y). In brackets are the RMS deviations

Amino acids	χ^1	X = I, V	X = D, N	X = F,Y,H,W	X = L,M,R,K,E,Q	X = T,C,S	X = A	X = G	X = P	Average (std)
F	180 ± 30	3.56 (2.47)	0.91 (3.20)	4.43 (3.35)	2.78 (2.41)	2.83 (2.80)	4.1 (1.62)	1.94 (2.45)	1.61 (3.90)	2.77 (1.23)
	60 ± 30	−2.46 (4.08)	−1.74 (4.30)	0.11 (4.23)	−1.98 (4.04)	−2.21 (3.93)	−0.61 (4.53)	−3.13 (3.40)	−2.8 (4.76)	−1.85 (1.10)
Y	180 ± 30	1.94 (2.73)	2.08 (4.01)	3.53 (3.46)	2.64 (2.39)	1.65 (2.75)	2.7 (2.61)	3.15 (3.42)	3.63 (3.34)	2.67 (0.74)
	60 ± 30	−2.78 (1.63)	1.14 (3.01)	−3.61 (3.99)	−1.19 (3.18)	−1.87 (3.29)	−2.15 (1.96)	−1.63 (3.49)	−2.35 (5.08)	−1.81 (1.40)
H	180 ± 30	0.83 (2.49)	2.78 (2.98)	2.05 (3.09)	1.93 (2.86)	1.25 (2.58)	1.32 (2.31)	3.37 (4.16)	2.11 (3.12)	1.96 (0.84)
	60 ± 30	−3.26 (3.02)	−1.64 (4.06)	−1.04 (5.43)	−1.03 (3.90)	−2.5 (5.10)	−1.02 (2.38)	−1.74 (3.86)	0.32 (2.51)	−1.49 (1.08)
W	180 ± 30	3.61 (2.80)	2.37 (1.56)	2.79 (4.38)	1.38 (2.01)	1.26 (2.49)	0.33 (3.12)	3.46 (1.74)	2.7 (6.84)	2.24 (1.15)
	60 ± 30	−1.25 (1.48)	−1.57 (2.98)	1.05 (4.15)	−1.79 (3.98)	−2.87 (3.84)	−3.28 (3.59)	3.97 (5.55)[a]	−3.51 (6.17)	−1.89 (1.56)
I	180 ± 30	−1.01 (2.54)	0.32 (3.38)	0.44 (3.38)	−1.51 (3.54)	−1.17 (3.77)	−1.1 (2.43)	−0.02 (3.43)	0.4 (3.17)	−0.46 (0.82)
	60 ± 30	−5.81 (3.43)	−4.21 (4.52)	−5.66 (5.21)	−5.17 (4.69)	−6.44 (5.14)	−5.83 (5.21)	−6.02 (4.06)	−6.45 (5.58)	−5.7 (0.73)
V	60 ± 30	−1.23 (3.24)	−1.48 (3.27)	−0.9 (4.38)	−1.57 (3.17)	−1.77 (3.99)	−2.06 (2.97)	−2.49 (3.20)	−2.14 (3.58)	−1.71 (0.52)
	-60 ± 30	−5.7 (4.33)	−4.01 (4.28)	−4.87 (5.54)	−5.57 (4.20)	−5.7 (4.36)	−4.71 (5.13)	−5.35 (5.12)	−4.5 (5.04)	−5.05 (0.63)
T	180 ± 30	−2.94 (3.59)	−3.8 (4.29)	−2.19 (4.38)	−2.89 (4.15)	−1.96 (3.86)	0.23 (2.60)	−3.84 (4.09)	0.77 (6.93)	−2.08 (1.73)
	60 ± 30	−5.29 (3.65)	−4.55 (4.15)	−5.22 (3.72)	−6.01 (3.62)	−4.58 (3.82)	−3.64 (3.85)	−5.57 (3.40)	−4.08 (4.34)	−4.87 (0.79)

[a]Not included in the calculation of the averaged value.

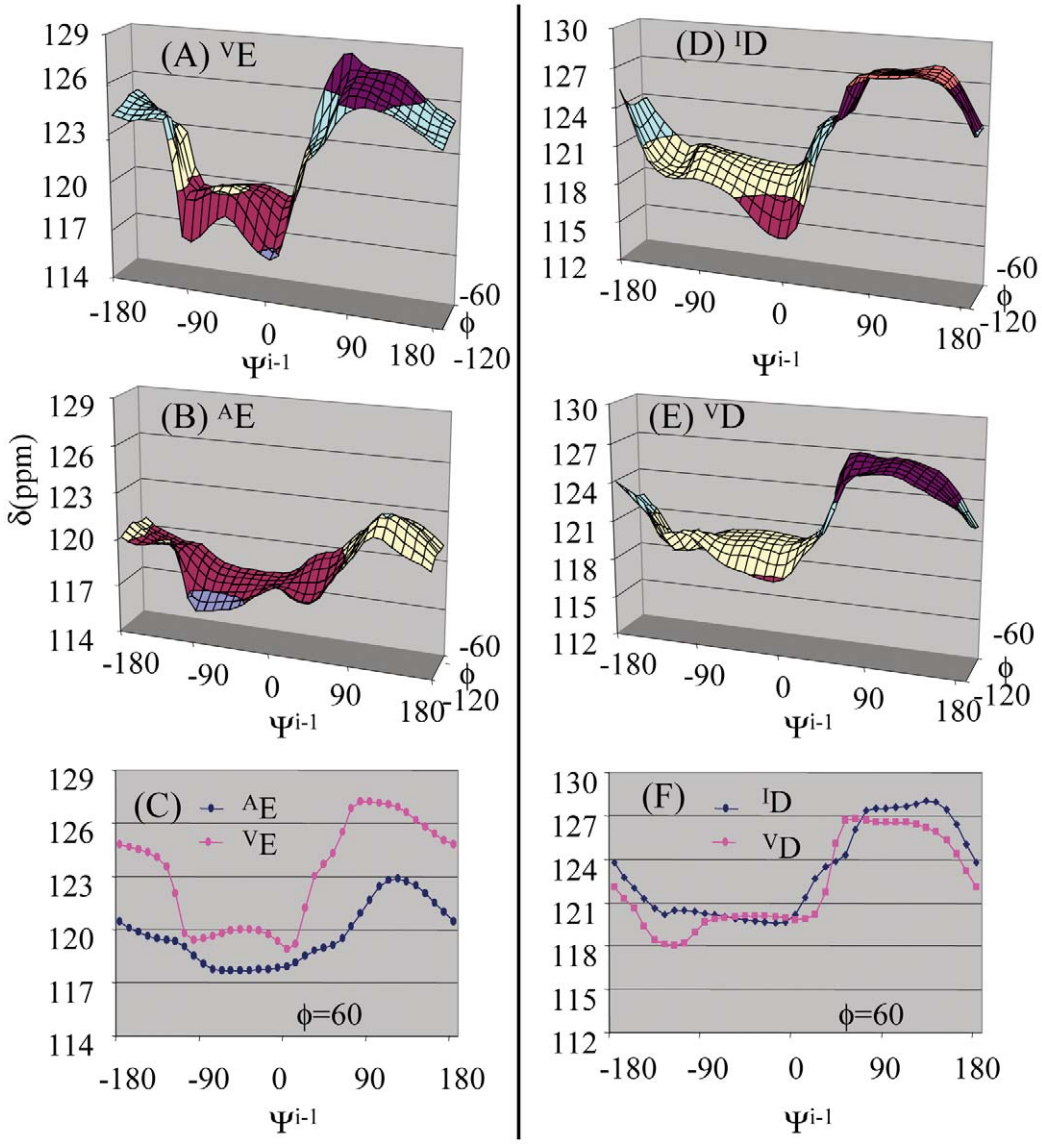

Figure 1. Representative PRSI ^{15}N-shielding surfaces show how ^{15}N shifts vary with the preceding residue and backbone torsion angles, ϕ and ψ^{i-1}. On the left are ^{15}N shifts of Glu preceded by Val (A) and Ala (B) respectively; for comparison a slice from A and B is shown in (C). On the right are ^{15}N shifts of Asp preceded by Ile (D) and Val (E) respectively; similarly their slices are shown in (C). These surfaces are calculated in 10 degree intervals with A=0.045 in the Gaussian convolution function.

of χ^1 torsion angles of residues Val, Ile, Thr, Phe, Tyr, His, and Trp can be clustered into one of the three staggered conformations–$180 \pm 30°$, $60 \pm 30°$, or $-60 \pm 30°$. This is obviously due to the steric hindrance caused either by the branched side-chain at C^β atom of Ile, Val, and Thr or by the aromatic ring of Phe, His, Tyr, and Trp. Three subset databases, $\delta^{X-com}Y\chi^1(\phi, \psi^{i-1})$ (where $\chi^1 = 180 \pm 30°$, $60 \pm 30°$, or $-60 \pm 30°$), were generated from each of the preceding residue combined database, $\delta^{X-com}Y(\phi, \psi^{i-1})$. The $\delta^{X-com}Y\chi^1(\phi, \psi^{i-1})$ databases were used for the investigation of the χ^1effects.

As a representative example showing the χ^1 effects, three ^{15}N-shielding surfaces of Val calculated using databases $\delta^{X-com}Y\chi^1(\phi, \psi^{i-1})$ are shown in Figure 3. Among the three surfaces, the one with $\chi^1 = 60 \pm 30°$ differ greatly (up to 8 ppm in certain areas) from the other two. Inspection of the steric

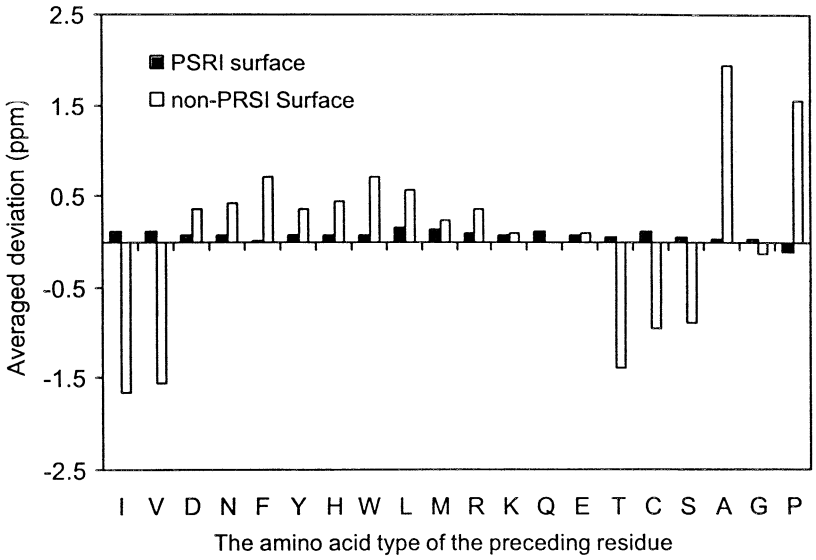

Figure 2. Average deviation between the observed ^{15}N chemical shifts and those predicted using PRSI and non-PRSI shielding surfaces respectively.

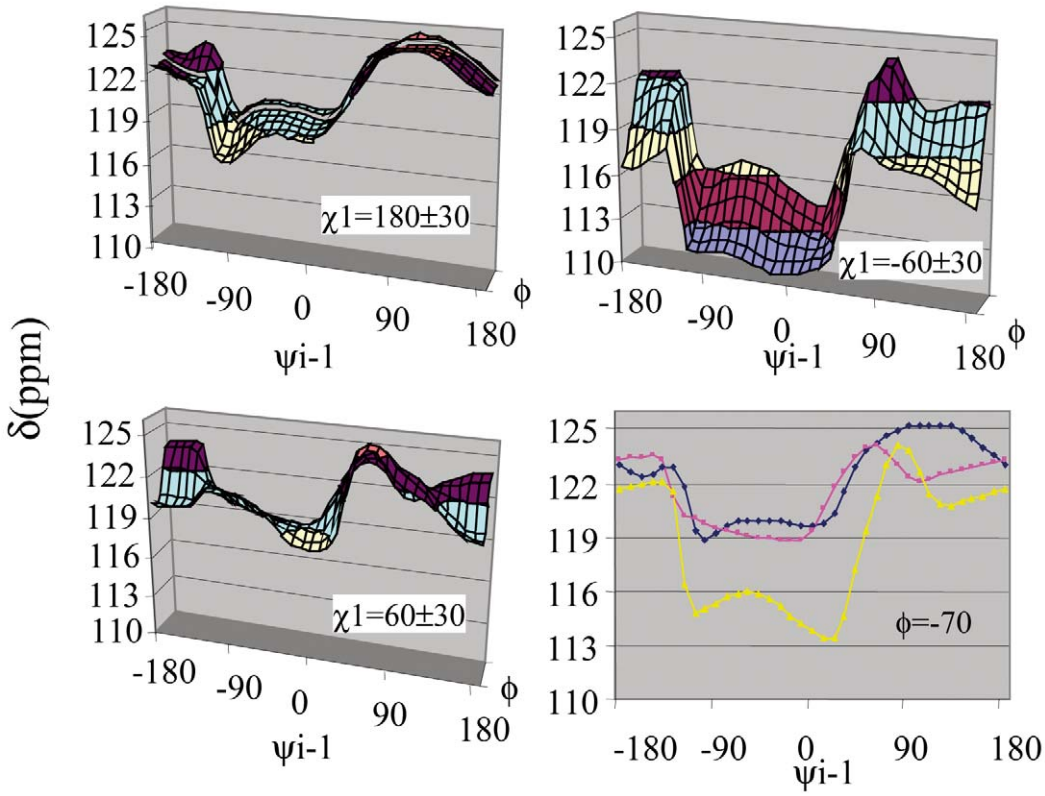

Figure 3. ^{15}N-shielding surfaces of Val calculated using the database $\delta^{X-com}Y\chi^1(\phi, \psi^{i-1})$, where, Y=Val; X={L, M, K, R, E, Q}; $\chi^1 = 180 \pm 30°, 60 \pm 30°$, and $-60 \pm 30°$ respectively.

structures of three staggered conformations revealed that, as early proposed by Le and Oldfield (Le and Oldfield, 1994) the influence on ^{15}N chemical shift from the side-chain geometry of Val is caused by the strong shielding from the methyl constituent on the β-carbon. The shielding from the methyl group also accounts for the strong χ^1 effect observed for Ile and Thr. Meanwhile, the strong χ^1 effect that we have observed for amino acids Phe, Tyr, His, and Trp are obviously caused by the aromatic ring. In this study, no significant side-chain (χ^1) effects were detected for the remaining amino acids, Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, Ser, Leu, Lys, and Met.

We have defined and calculated the χ^1 effect correction factor, $\Delta(^{X}Y)\chi^1$, which represents the change in ^{15}N chemical shift for a amino acid when its side-chain orientation shift from the referenced (as defined in this study) conformation to another. The averaged correction factors–over all the data points in each of the database of interest, $< \Delta(^{X}Y)\chi^1 >$, are calculated and listed in Table 2. As shown in this table, the χ^1 induced change in ^{15}N chemical shift can be as high as 5 ppm for the amino acids Val, Ile, Thr, Phe, Tyr, His, and Trp. The result in Table 2 is quantitatively in very good agreement with that calculated by deDios et al. using *ab initio* quantum mechanical methods (deDios et al., 1993). In general, the variation of the χ^1 effect with preceding residue types can be ignored. As shown in Table 2, there are a few data points standing out for being significantly different than others in the same row, particularly for the $60 \pm 30°$ staggered conformation. These unusual $< \Delta(^{X}Y)\chi^1 >$ values are very likely caused by the extremely low number (less than 10) points used in the analysis. On average, the population is only 13%, 19%, and 68% on the staggered conformations of $60 \pm 30°$, -60 ± 30, and $180 \pm 30°$ respectively for the above-mentioned amino acids (except for Val).

The present XPRSI shielding surfaces, which simultaneously take into account the backbone conformational effects, the preceding residue effects, and the χ^1effects, were constructed to predict ^{15}N chemical shifts in proteins from the torsion angles ϕ, ψ^{i-1}, and χ^1. To evaluate the χ^1 effects, the chemical shift entries of Val, Ile, Phe, Tyr, Trp, Thr, and His were clustered into three groups with their χ^1 angle in the range of $180 \pm 30°$, $60 \pm 30°$, and $-60 \pm 30°$ respectively. The ^{15}N chemical shifts in each cluster were predicted using PRSI shielding surfaces (from ϕ and ψ^{i-1}) and XPRSI shielding surfaces from (ϕ, ψ^{i-1}, and χ^1) respectively. The averaged deviations between

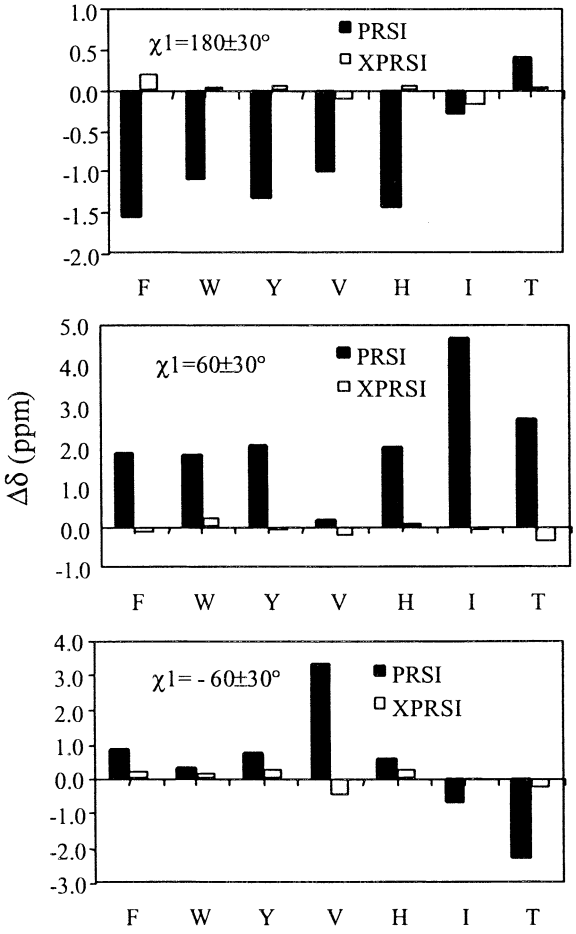

Figure 4. Average deviation of the predicted from the observed ^{15}N chemical shifts for amino acids Ile, Val, Thr, Phe, Tyr, His, and Trp.

predicted and observed ^{15}N chemical shifts are plotted in Figure 4. As shown in this figure, there are no significant deviations between observed ^{15}N chemical shifts and those predicted using the XPRSI shielding surface for all the three χ^1 conformations. In contrast, significant discrepancies (for example as large as up 5 ppm for Ile at $60 \pm 30°$) exist between the observed ^{15}N chemical shifts and those predicted using PRSI shielding surfaces.

Comparison between non-PRSI, PRSI, and XPRSI

250 chemical shift entries for each of the 20 amino acids (target residue Y, except for Pro) were randomly selected from the database $\delta(\phi, \psi^{i-1})$ and used to test the efficiency of PRSI and XPRSI. ^{15}N chemical shift of each selected entry was predicted from its torsion angles using non-PRSI, PRSI and XPRSI shielding

Table 4. Correlation coefficients, RMS deviations, and slopes for linear regression fits of the observed ^{15}N chemical shifts versus those predicted using non-PRSI, PRSI, and XPRSI shielding surfaces respectively

Amino acid	Correlation coefficient			RMS deviation			Slope		
	Non-PRSI	PRSI	XPRSI	Non-PRSI	PRSI	XPRSI	Non-PRSI	PRSI	XPRSI
A	0.530	0.657		2.81	2.56		0.356	0.513	
R	0.582	0.668		3.02	2.71		0.385	0.540	
N	0.497	0.662		3.31	2.92		0.311	0.482	
D	0.5759	0.710		3.17	2.80		0.344	0.526	
C	0.564	0.635		3.70	3.32		0.341	0.488	
Q	0.584	0.692		2.97	2.59		0.405	0.595	
E	0.564	0.666		2.96	2.68		0.323	0.497	
G	0.588	0.630		3.28	2.98		0.354	0.517	
S	0.456	0.734		3.25	2.87		0.241	0.536	
L	0.620	0.719		2.86	2.61		0.397	0.600	
K	0.624	0.731		3.07	2.76		0.345	0.580	
M	0.676	0.695		2.91	2.83		0.399	0.540	
Average	0.567	0.678		3.11	2.80		0.347	0.529	
F	0.627	0.723	0.803	3.47	3.21	2.98	0.334	0.52	0.658
I	0.475	0.662	0.764	3.54	3.27	2.99	0.231	0.437	0.615
H	0.676	0.732	0.762	3.37	3.30	3.16	0.403	0.601	0.646
T	0.489	0.565	0.694	4.50	4.18	3.69	0.229	0.382	0.53
W	0.688	0.696	0.767	3.21	3.07	2.90	0.423	0.576	0.678
Y	0.531	0.631	0.722	3.56	3.18	2.95	0.282	0.479	0.596
V	0.421	0.549	0.715	3.91	3.61	3.20	0.215	0.423	0.576
Average	0.558	0.651	0.747	3.65	3.40	3.12	0.302	0.488	0.614

surfaces respectively. Three parameters are used for the evaluation. They are (1) chemical shift dispersion (here defined as six times the standard deviation), (2) RMS deviation, and (3) linear regression fits (slope and correlation coefficient) between the observed and the predicted ^{15}N chemical shifts.

For an amino acid, the dispersion of its chemical shifts represents a number of factors contributing to its chemical shift. For each of 20 amino acids (except for Pro), the dispersions of the observed ^{15}N chemical shift and those predicted using non-PRSI, PRSI, and XPRSI shielding surfaces are listed in Table 3. The results shown in this table (as well as in Table 4) were categorized into two groups. Group I contains those amino acids, of which the ^{15}N chemical shifts are less sensitive to χ^1 conformation. Group II is composed by Val, Ile, Thr, Phe, Tyr, His, and Trp, for which ^{15}N chemical shifts are strongly influenced by the χ^1 conformation. On average, the dispersions of the observed ^{15}N chemical shifts for Group I amino acids are notably narrower than those in Group II – the averaged dispersions are 22.9 and 26.1 ppm for Group I and II respectively. The χ^1 effect, which is stronger for Group II amino acids, could be a good explanation for the differences between Group I and II. Among the Group I amino acids, Cys stands out for having the largest observed ^{15}N chemical shift dispersion of 27.4 ppm, much higher than the averaged value of 22.9 ppm. This is because, an additional factor–the redox state of the side chain of Cys, greatly affect its chemical shifts. The second largest dispersion in Group I is that for Gly, 26.64 ppm. Close inspection of the ^{15}N chemical shift distribution reveals that around 4% of observed chemical shifts of Gly fall in a range of 120–130 ppm, that is far from its normally observed values (105–115 ppm). As pointed out by one the reviewers of this manuscript, these higher field Gly shifts could be simply a mis-assignment– the folded peaks caused by the narrow spectral window used for the ^{15}N dimension during the experiments.

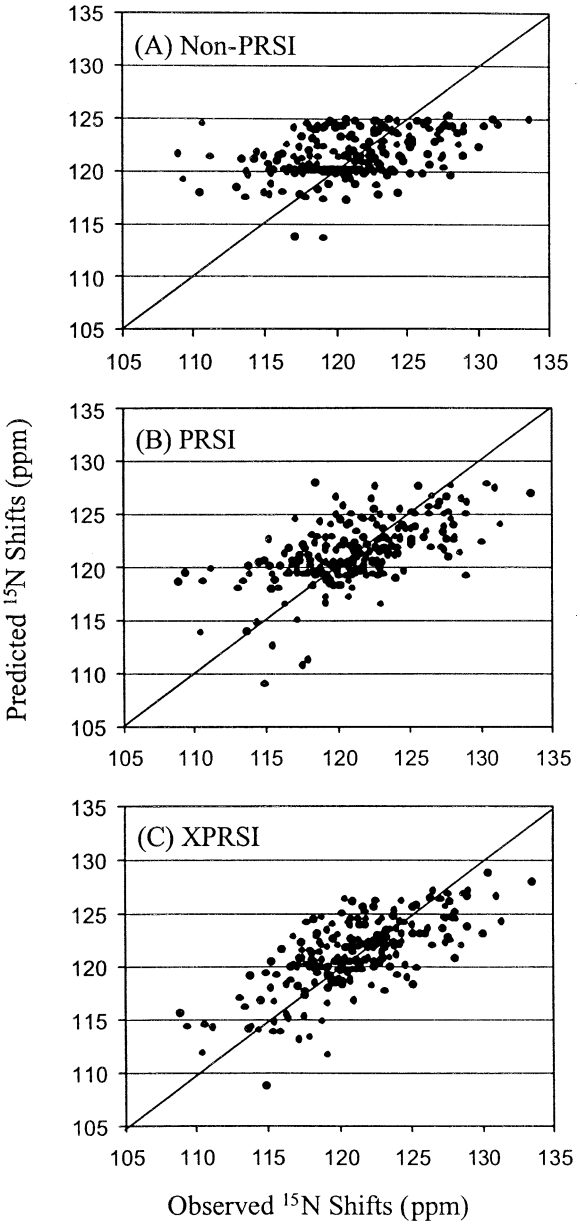

Figure 5. Correlations between the observed ^{15}N chemical shifts and those predicted from ϕ and ψ^{i-1} backbone dihedral angles using Non-PRSI (A), PRSI, and XPRSI (C) shielding surfaces for the amino acid Val.

On the other hand, the absence of a side-chain on Gly, which allows this amino acid to have more flexibility in its backbone conformation, could also make its ^{15}N shifts more sensitive to the environment than other amino acids. Therefore, the observed higher field Gly ^{15}N shifts might indicate the existence of unknown factors that greatly affect its ^{15}N chemical shift. A further study of these abnormally high Gly shifts is clearly needed. Excluding the 4% extremely high values in the calculation yield a narrowed dispersion of 22.4 ppm, which is much closer to that for the other amino acids in this group. As shown in Table 3, dispersions of the predicted ^{15}N chemical shifts significantly increase when PRSI and XPRSI shielding surfaces are used. For Group I amino acids, the averaged dispersion for the predicted shifts using PRSI surfaces is 16.8 ppm (6.1 ppm less than that for the observed ^{15}N chemical shifts). For Group II, the average dispersion for the predicted shifts using XPRSI surfaces is 20.2 ppm (5.8 ppm less than that for the observed ^{15}N chemical shifts). This indicates that incorporating the χ^1 effect improves the prediction of Group II amino acids to the same level of accuracy as for Group I amino acids. More important, the narrower dispersion of the predicted ^{15}N shifts (6 ppm less than that for the observed value) clearly demonstrates the existence of other unknown factors that have a strong influence on ^{15}N chemical shift in proteins.

Table 3. Chemical shift dispersions (six times of the RMS deviation, in ppm)

Amino acids	Non-PRSI	PRSI	XPRSI	Observed
A	11.58	14.52		19.68
R	13.14	16.08		21.30
N	12.00	18.42		21.78
D	13.74	18.12		24.30
C	14.40	21.30		27.42
Q	13.26	15.96		21.30
E	11.88	15.72		20.94
G	14.04	14.46		26.64
S	12.72	14.64		22.92
L	13.98	18.06		22.20
K	11.16	16.62		21.96
M	15.48	17.52		24.30
Average	13.12	16.79		22.90
F	12.66	16.14	18.90	24.60
I	11.46	15.06	17.94	24.06
H	15.06	18.72	20.46	25.50
T	11.28	17.28	22.02	28.32
W	15.72	21.66	22.86	27.36
Y	13.08	18.12	20.04	26.04
V	11.76	16.20	19.08	26.58
Average	13.00	17.60	20.19	26.07

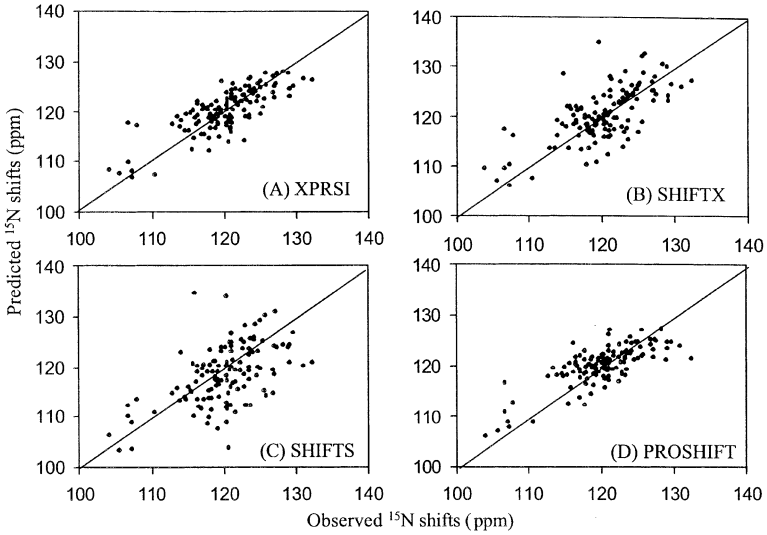

Figure 6. Correlations between the observed ^{15}N chemical shifts and those predicted using (A) XPRSI, (B) SHIFTX, (C) SHIFTS, and (D) PROSHIFT for protein MTP1598.

The predicted ^{15}N shifts versus these observed for Val are plotted in Figure 5. As shown in this figure, the observed ^{15}N chemical shifts of Val spread over a range of ~25 ppm (from 107 to 132), however, those predicted using non-PRSI surfaces are limited to a relatively narrow range of 16 ppm (from 119–125). When PRSI and XPRSI surfaces are used for the prediction, the predicted ^{15}N chemical shifts spread over a wider range of ~20 ppm.

Linear regression fittings of the observed ^{15}N chemical shifts versus these predicted using non-PRSI, PRSI, and XPRSI shielding surfaces were performed for each of the 20 amino acids (except for Pro). Table 4 lists RMD deviations, correlation coefficients (r), and slopes of such a fit for each amino acid. From non-PRSI to PRSI and to XPRSI shielding surfaces, better predictions are clearly demonstrated by the improved fitting (increased correlation coefficients and slopes) and lower RMS deviations for each amino acids as listed in this table. The prediction for Gly is at first glances rather poor in comparison with all other amino acids (r = 0.492, RMS deviation = 3.76, slope = 0.289 compared to the average values of 0.68, 2.80, and 0.53 respectively). Close examination of the predicted and the observed ^{15}N chemical shifts reveals that this poorer prediction for Gly is, again, caused by the 4% unusual chemical shifts in the range of 120–130 ppm, that can't be accurately predicted by any of the shielding surfaces. Again, excluding these 4% extremely high shifts yields a better prediction for Gly ^{15}N shifts: Correlation coefficient = 0.588, RMS deviation = 2.98, slope = 0.517, which is much closer to that for the other amino acids in the same group.

Comparison with other approaches

In the past decades, several techniques have been developed to predict the chemical shifts in proteins for ^{15}N as well as for other nuclei. Among those available, *ab initio* quantum mechanical calculation needs very complicated computations; sequence homology based approach is largely dependent on the availability of homologous proteins with assigned chemical shifts. We have noticed that three programs that can automatically predict protein chemical shifts from its 3d coordinates have been published recently. They are SHIFTS which uses density functional theory (Xu and Case, 2001, 2002), SHIFTX which is based on a hybrid predictive approach (Neal et al., 2003), and PROSHIFTS which based on a neural network (Meiler, 2003). Similar to the present approach, all these three algorithms are utilizing chemical shifts and 3D structural correlations that are empirically derived from BMRB-PDB database. However, unlike the present study, each individual parameter affecting chemical shifts has not been quantitatively determined and applied in these earlier studies. The XPSRI shielding surfaces we proposed in this study use three quantitatively determined effects – backbone conformation (ϕ and ψ^{i-1}) effect, the preceding residue effect, and the side-chain

Table 5. Correlation coefficients, RMS deviations, and slopes for linear regression fits of the observed ^{15}N chemical shifts versus those predicted using XPRSI, SHIFTX[a], SHIFTS[b], and PROSHIFT[c] for the ten test proteins

Proteins (BMRB & PDB #)	Correlation coefficient				RMS deviation				Slope			
	XPRSI	SHIFTX	SHIFTS	PRO-SHIFT	XPRSI	SHIFTX	SHIFTS	PRO-SHIFT	XPRSI	SHIFTX	SHIFTS	PRO-SHIFT
I–FABP (5285, 3IFB)	0.755	0.627	0.573	0.692	3.54	4.57	5.76	3.93	0.617	0.607	0.726	0.565
NNOS PDZ (4304, 1B8Q)	0.785	0.703	0.623	0.765	3.68	4.62	6.40	3.72	0.741	0.757	0.878	0.633
FimC (4070, 1QUN)	0.856	0.868	0.751	0.827	2.86	2.82	4.67	3.12	0.759	0.857	0.961	0.641
Bet v 1-L (4417, 1B6F)	0.849	0.833	0.712	0.824	2.91	3.23	5.08	3.26	0.713	0.79	0.892	0.632
Scaffoldin (4589, 1EHX)	0.804	0.72	0.649	0.749	2.37	4.78	6.29	4.21	0.647	0.65	0.78	0.555
Apaf-1 CARD (4661, 2YGS)	0.847	0.845	0.735	0.791	2.10	2.13	3.06	2.42	0.682	0.79	0.819	0.594
Ub1D8 (4663, 1C3T)	0.897	0.841	0.759	0.858	2.56	3.31	4.34	3.01	0.785	0.791	0.816	0.674
HEDA (5027, 1J8K)	0.838	0.841	0.717	0.814	3.23	5.06	5.59	3.43	0.697	0.712	0.971	0.655
MTP1598 (5165, 1JW3)	0.782	0.692	0.534	0.763	3.18	4.17	5.35	3.28	0.672	0.711	0.624	0.586
RRF (5190, 1EK8)	0.814	0.8	0.605	0.807	2.46	2.65	4.49	2.48	0.72	0.794	0.799	0.653
Average	0.823	0.777	0.664	0.789	2.89	3.72	5.12	3.29	0.703	0.746	0.824	0.619

[a]SHIFTX– version 1.0, Neal et al. (2003).
[b]SHIFTS– version 4.1, Xu and Case (2001, 2003). Side-chain refinements were applied during the prediction.
[c]PROSHIFT– current version, Meiler (2003).

orientation (χ^1) effect, in ^{15}N chemical shift prediction. Ten proteins were randomly selected from our database and their ^{15}N chemical shifts were predicted using the present XPRSI, SHIFTX (version 1.0; http://redpoll.pharmacy.ualberta.ca), SHIFTS (version 4.1; http://www.scripps.edu/case), and PROSHIFT (current version, http://www.jens-meiler.de) respectively. Performance evaluations for the four programs were made through the comparison the observed shifts with that predicted using each program respectively. As a representative example, the observed ^{15}N shifts versus these predicted using XPRSI, SHIFTX, SHIFTS, and PROSHIFT for one of the test proteins, Methanobacterium Thermoautotrophicum Protein 1598 (MTP1598), are plotted in Figure 6. Meanwhile, Table 5 lists the correlation coefficients, RMS deviations, and slopes for linear regression fits of the observed ^{15}N chemical shifts versus those predicted for all the ten testing proteins. The results summarized in Table 5 clearly demonstrate performance of each of the four programs in predicting the ^{15}N chemical shifts. Briefly, the present XPRSI shielding surface and PROSHIFT yield significantly lower RMS deviations than that SHIFTS and SHIFTX for all the ten testing proteins. The averaged RMS deviations of the prediction for the ten testing proteins are 2.89, 3.29, 3.72, and 5.12 ppm for XPRSI, PROSHIFT, SHIFTX, and SHIFTS respectively. The present XPRSI and PROSHIFT also give higher correlation coefficients—the averaged values are 0.823, 0.789, 0.777, and 0.664 for XPRSI, PROSHIFT, SHIFTX, and SHIFTS respectively. On the other hand, XPSRI and PROSHIFT give lower fitting slopes (0.703 and 0.619 respectively) than SHIFTS and SHIFTX (0.746 and 0.824 respectively). We believe, better ^{15}N shift prediction using present PRSI (and XPRSI) shielding surfaces could

be reached by taking into account additional parameters, such as aromatic ring effects, hydrogen-bond, and solvent effect, etc.

References

Braun, D., Wider, G. and Wüthrich, K. (1994) *J. Am. Chem. Soc.*, **116**, 8466–8469.

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993) *Science*, **260**, 1491–1496.

Glushka, J., Lee, M., Coffin, S. and Cowburn, D. (1989) *J. Am. Chem. Soc.*, **111**, 7716–7722.

Gronenborn, A.M. and Clore, G.M. (1994) *J. Biomol. NMR*, **4**, 455–458.

Havlin, R.H., Laws, D.D., Bitter, H.L.M., Sanders, L.K., Sun, H., Grimley, J.S., Wemmer, D.E., Pines, A. and Oldfield, E. (2001) *J. Am. Chem. Soc.*, **123**, 10362–10369.

Havlin, R.H., Le, H., Laws, D.D., deDios A.C. and Oldfield, E. (1997) *J. Am. Chem. Soc.*, **119**, 11951–11958.

Iwadate, M., Asakura, T. and Williamson, M.P. (1999) *J. Biomol. NMR*, **13**, 199–211

Kricheldorf, H.R. (1981) *Org. Magn. Reson.*, **15**, 162–177.

Le, H. and Oldfield, E. (1994) *J. Biomol. NMR*, **4**, 341–348.

Meiler, J. (2003) *J. Biomol. NMR*, **26**, 25–37.

Neal, S, Nip, A.M., Zhang, H. and Wishart, D.S. (2003) *J. Biomol. NMR*, **26**, 215–240.

Ösapay, K. and Case, D.A. (1991) *J. Am. Chem. Soc.*, **113**, 9436–9444.

Pearson, J.G., Le, H., Sanders, L.K., Godbout, N., Havlin, R.H. and Oldfield, E. (1997) *J. Am. Chem. Soc.*, **119**, 11941–11950.

Santiveri, C.M., Rico, M. and Jimenez, M.A. (2001) *J. Biomol. NMR*, **19**, 331–345.

Schwarzinger, S., Kroon, G. J.A., Foss, T.R., Chung, J., Wright, P.E. and Dyson, H.J. (2001) *J. Am. Chem. Soc.*, **123**, 2970–2978.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Wang, Y. and Jardetzky, O. (2002a) *Protein Sci.*, **11**, 852–861.

Wang,Y. and Jardetzky, O. (2002b) *J. Am. Chem. Soc.*, **124**, 14075–14084.

Williamson, M.P. and Asakura, T. (1993) *J. Magn. Reson. B*, **101**, 63–71.

Wishart, D.S. and Case, D.A. (2001) *Meth. Enzymol.*, **338**, 3–34.

Wishart, D.S. and Nip, A.M. (1998) *Biochem. Cell Biol.*, **76**, 1–10.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Wishart, D.S.,Watson, M.S., Boyko, R.F. and Sykes, B.D. (1997) *J. Biomol. NMR*, **10**, 329–336.

Xu, X-P. and Case, D.A. (2001) *J. Biomol. NMR*, **21**, 321–333.

Xu, X-P. and Case, D.A. (2002) *Biopolmers*, **65**, 215–240.

Zhang, H.Y., Neal, S. and Wishart, D.S. (2003) *J. Biomol. NMR*, **25**, 173–195.