J Biomol NMR (2007) 38:139–150
DOI 10.1007/s10858-007-9156-8

ARTICLE

A new model for chemical shifts of amide hydrogens in proteins

Seongho Moon · David A. Case

Received: 23 January 2007 / Accepted: 14 March 2007 / Published online: 14 April 2007

Abstract We propose a new computational model to predict amide proton chemical shifts in proteins. In addition to the ring-current, susceptibility and electrostatic effects of earlier models, we add a hydrogen-bonding term based on density functional calculations of model peptide–peptide and peptide–water systems. Both distance and angular terms are included, and the results are rationalized in terms of natural bond orbital analysis of the interactions. Comparison to observed shifts for 15 proteins shows a significant improvement over existing structure-shift correlations. These additions are incorporated in a new version of the SHIFTS program.

Keywords Chemical shifts · Proteins · Amide hydrogen

Introduction

Over the past decades, the structures and dynamics of thousands of proteins have been studied using NMR. Among the parameters measured by NMR, the proton chemical shift plays a key role in peak assignments, but is generally less used in structural studies, in spite of its sensitivity to the covalent structure of the molecule and to through-space interactions (Wishart and Case 2001). A general problem is that many electronic and geometric factors contribute to shift dispersion, so that using this for protein structure determination is not straightforward. The use of abundant shift information for structural studies depends on the level of our understanding of various structural contributions to the shift.

Empirical efforts to elucidate the relationship between the chemical shift and protein structure express the shift as a sum of effects such as aromatic ring current (Haigh and Mallion 1980), magnetic anisotropy (McConnell 1957) and electrostatic effects (Buckingham et al. 1960). These effects are included into shift calculations by using the empirical formulas derived from classical physics and experimental data (Ösapay and Case 1991; Asakura et al. 1995; Sitkoff and Case 1997). In this approach, the local covalent effects were approximated by the observed shifts of short peptides in random-coil conformations. This approach works fairly well for protons bonded to carbon, but has been much less successful for protons bonded to oxygen or nitrogen, presumably because of strong hydrogen-bonding effects that are not adequately accounted for in the classical decompositions. An alternative approach, which can be based on either experimental or quantum-mechanical calculations, using chemical shift 'hyper-surfaces', which are obtained from databases of observed chemical shifts, mainly focus on local structural changes. The hyper-surfaces relate chemical shifts to various local structural parameters (backbone dihedral angles, side-chain orientations, nearest neighbors, secondary structures, etc). These methods are widely used to predict the backbone conformation of an amino acid from its chemical shifts and to restrict the possible ranges for the protein backbone angles in structure determination protocols (Spera and Bax 1991; Le and Oldfield 1994; Beger and Bolton 1997; Iwadate et al. 1999; Cornilescu et al. 1999). Recently, Neal et al. (2003) have shown that an accuracy for predicting chemical shifts (including amide proton shifts) could be improved by the combination of the empirical formulas for the through-space interactions with the 'hyper-surfaces' for

S. Moon · D. A. Case (✉)
Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
e-mail: case@scripps.edu

local covalent interactions. However, the empirical approaches still afford only a partial understanding of conformational effects on chemical shifts because of some missing interactions (e.g. close-contact interactions and angular dependences of proton shifts in hydrogen bonding regions) and the loss of accuracy arising from different experimental conditions between X-ray and NMR (e.g. different structures, solvent and dynamic effects). In addition, the limited number of available experimental data, which are essential for the parameterization of empirical formulas, may reduce the accuracy of predicted shifts.

Recent advances in quantum mechanical (QM) approaches, which can include most crucial covalent effects and through-space interactions, make them valuable candidates for predicting shifts in proteins. This is true for the case of C and N shifts for selected classes of residues in proteins, where the QM calculations use protein fragments (de Dios et al. 1993; Le and Oldfield 1996; Pearson et al. 1997). These fragmental approaches can be rationalized by the fact that the shifts of heavy atoms are mainly affected by local structures. The shielding parameters of amide protons in proteins, which are more sensitive to the local structure than those of aliphatic protons, have been studied in the same way (Sharma et al. 2002; Barfield 2002). Since QM methods allow one to investigate structural and environmental shift effects in a systematic and controlled manner, they have been also used in the pre-calculated C and N shift 'hyper-surfaces' (Xu and Case 2001, 2002; Neal et al. 2003) and in the parameterization of physical models of environmental proton shift effects (Sitkoff and Case 1997).

In the present study, we provide a new empirical approach to predict shifts of amide protons in proteins, where quantum hydrogen bond (HB) effects are combined with other classical through-space interactions. Systematic quantum shielding calculations are performed to investigate the hydrogen bonding (HB) effects of amide proton shifts in proteins, by examining shifts in the complexes of N-methylacetamide (NMA) and formamide in various geometries. In addition, the complexes of NMA and water molecule in various geometries are used for the solvent effect study. Angles and distances are investigated and used for making empirical formulas by training artificial neural networks that are powerful in recognizing shift patterns to the structural variation (Meiler 2003). The formulas for HB effects are integrated into the empirical method, which includes formulas for ring-currents, magnetic anisotropies of peptide groups, and electrostatic interactions, to predict amid proton shifts in proteins. For the better description of solvent effects, HB effects of water molecules are investigated in combination with MD simulations. The new empirical method will be valuable for predicting amide proton shifts because it can include most important short-range interactions and environmental effects in a systematic way.

Methods

Quantum shielding calculations for H-bonded model systems

All quantum calculations were performed using the Gaussian98 program (Frisch et al. 1998). The N-methylacetamide (NMA) and formamide (FA) molecules were fully optimized (in vacuum) at the MP2(FC)/6-31G* level. The optimized NMA and FA structures were used to generate the H-bonded complexes that were used for shielding calculations to investigate the structural dependencies of amide proton shifts in the hydrogen bonding regions. The structures of the H-bonded complexes are graphically shown in Fig. 1. The NMA and water molecule complexes were also generated in the same way as the NMA–FA complexes to measure HB effects arising from solvent molecules. The amide proton shieldings of the H-bonded complexes were calculated using the B3PW91/6-31+G** level (B3PW91 makes use of Becke's three-parameter hybrid exchange functional (Becke 1993) and the non-local Perdew/Wang 91 correlation functional (Perdew and Wang 1992). Since the diffuse function can greatly reduce the basis set superposition errors (BSSEs) of shielding calculations of H-bonded model peptides (Cui and Karplus 2000), corrections for superposition errors were not carried out. The 6-31+G** basis set gives absolute shieldings that are significantly different from

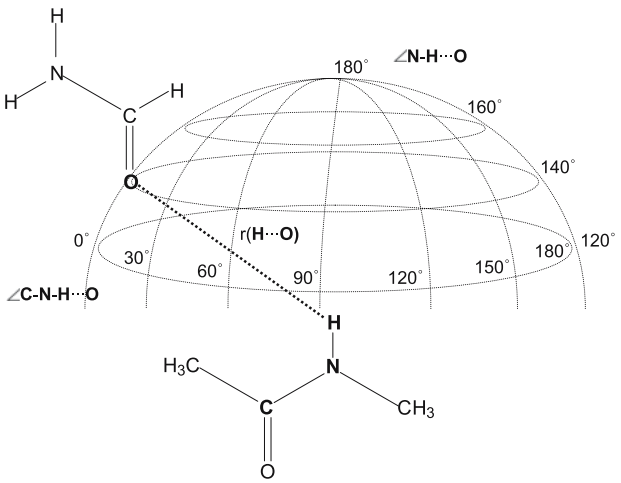

Fig. 1 Structure of the H-bonded complex (N-methylacetamide and formamide) used in quantum calculations where the molecular planes of two are in parallel

complete basis set (CBS) extrapolations, but previous work has shown that the relative shifts needed to analyze hydrogen-bond interactions are comparable to CBS for the model peptides (Moon and Case 2006). The natural chemical shielding (NCS) analyses (Bohmann et al. 1997) were done using the NBO 5.0 module linked to the Gaussian98 program. The B3PW91/6-31+G** level of theory was also used for the NCS calculations.

Neural network training

Fitting quantum data to empirical models can be carried out in a variety of ways. Here we use artificial neural networks as part of a general fitting strategy for many types of NMR data (Zupan and Gasteiger 1993; Meiler et al. 2002; Meiler 2003). Specifically, feed-forward multilayer perception (MLP) model neural networks were employed (Rumelhart and McClelland 1986). The input layer, which receives the input signals and distributes them forward to the network, uses 196 geometrical parameter sets, where each set has three nodes (the r(H...O) hydrogen bond length, the N–H...O angle, and the C–N–H...O dihedral angle). In the hidden layer where each node receives the weighted sum of the outputs of the input layer nodes as a signal, 10 nodes (or hidden neurons) were used. The output of a node in a hidden layer is obtained through a nodal activation function, where a sigmoid function, was used. In the output layer, 196 calculated shielding values were used for the target of the network. Each output value has one node with a simple linear function as an activation function. Since the sigmoid function can handle only the values between 0 and 1, the input and output data were linearly scaled to lie between 0 and 1 before training. This procedure is schematically shown in Fig. 2. The empirical formula obtained by the networks for HB effects is given by

$$\sigma_{HB} = \sum_{i=1}^{10} w_i^{(2)} f\left(\sum_{j=1}^{3} w_{ij}^{(1)} x_j + b_i^{(1)}\right) + b^{(2)} \qquad (1)$$

Here, is the weight for the connection between node j in the input layer and node i in the hidden layer, is the weight for the connection between node i in the hidden layer and the node in the output layer, and b are corresponding biases, respectively. The network weights and the bias terms were determined by training, using the NNDT software package. The network connections were trained with the training data set until root-mean-square (rms) errors were minimized. Since the shielding patterns used here were not complicated and the number of hidden neurons was not large, the input data was not divided into a training set and test sets. Rather, we tested for signs of over-fitting by generating multiple data that included geometries that were interspersed among the points used for network training. The network gave smooth interpolations for such data sets. We expect to be able to simplify Eq. 1 in future work, but it has served us well in the current work, and is incorporated into version 4.2 of the SHIFTS software, as discussed below.

Hybrid shielding calculations combined with molecular dynamics

For realistic shielding simulations that include solvent and motional averaging effects on chemical shielding, molecular dynamics (MD) can be combined with quantum mechanics (QM) or empirical approaches. This hybrid scheme extracts MD snapshots selected in a certain time interval, applies quantum or empirical shielding calculations, and averages the results. In particular, solvent effects can be included into the shielding calculations by using the explicit water models in the MD simulations.

The MD simulations were performed using the AMBER8 program suite (Case et al. 2005). The oxidized thioredoxin (PDB ID: 1XOA) (Jeng et al. 1994) was chosen as a target protein and the initial structure was taken (arbitrarily) from the 7th of 20 NMR structures. The protein was put in a box with 4,438 water molecules and three sodium ions. The *ff03* force field (Duan et al. 2003) was used with a TIP3P water solvent model (Jorgensen et al. 1983). Electrostatic interactions were treated using the Particle Mesh Ewald (PME) method with 10 Å non-bonded cutoff. After 1 ns equilibration procedure, 10 ns constant pressure simulation was carried out with SHAKE at 300 K. From the MD trajectory, 100 snapshots with 100 ps time

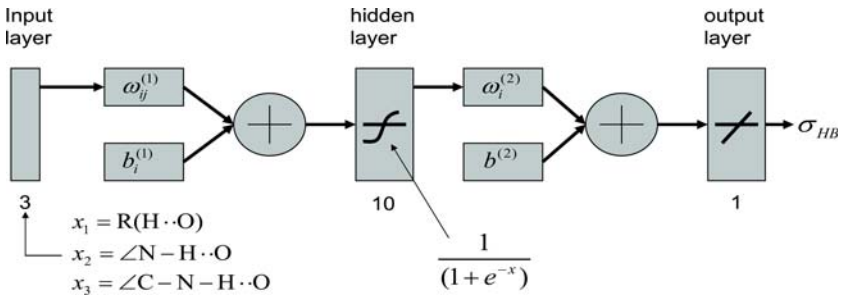

Fig. 2 Structure of the artificial neural network where three geometrical parameters (r(H...O) bond length, $\angle \mathrm{N-H\ldots O}$ angle and $\angle \mathrm{C-N-H\ldots O}$ dihedral angle) are used for input and the corresponding HB shieldings were used for the target. 10 sigmoid functions are used in the hidden layer

interval were extracted and analyzed. The protein structure in each snapshot was refined by 100 step energy-minimization (50 step steepest descent and 50 step conjugate gradient). Quantum shielding calculations (at the B3LYP/6-31+G** level) were carried out using the fragments taken from the 100 structure-refined snapshots.

For quantum shielding calculations, 30 protein fragments (including 30 target residues) were chosen randomly. Each fragment includes at least three peptide units. For example, if the peptide of interest is the *i*-th peptide in a protein, the $(i-1)$th peptide and parts of $(i+1)$th and $(i-2)$th peptides are always included with the fragments. If additional residues (including solvent molecules) are within 3.0 Å distance from the amide proton of the *i*th peptide and the carbonyl oxygen of the $(i-1)$th peptide, the fragments are extended to include the spatially close residues (or parts of the residues). Each fragment was truncated by quantum capping potentials (QCPs) without adding any artificial link-atoms, as described earlier (Moon and Case 2006). An example of this fragment procedure is graphically shown in Fig. 3.

Empirical shift calculations

The empirical shift calculations of amide proton were carried out using the SHIFTS program suite (Xu and Case 2001, 2002). In this program, the proton shift is predicted from several empirical contributions arising from random coil, ring current, peptide group magnetic anisotropy, and electrostatic effects. The additional contributions from HB effects were calculated from the empirical formulas obtained by quantum model studies (using NMA–FA and NMA–water complexes) and neural networks. For this purpose, the empirical terms for HB effects were added to the SHIFTS program, creating version 4.2, which is available for download from http://www.scripps.edu/case

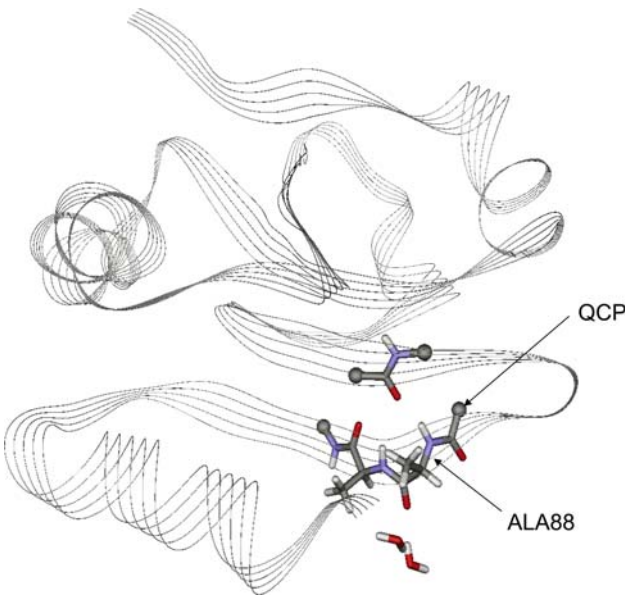

Fig. 3 Structure of a fragment (A88) in the oxidized thioredoxin where the fragment is saturated by quantum capping potentials

Results and discussion

Structural dependence of hydrogen bond effects on amide proton shifts

HB effects on chemical shifts have been often described by solely distance-dependent terms, but density functional model studies show that adding angle-dependent terms can improve calculations for amide proton shifts (Barfield 2002). Here, three main geometrical parameters were chosen (see Fig. 1): (1) the r(H..O) hydrogen bond length (1.8 ~ 4.0 Å), (2) the $\angle$N – H..O angle (120 ~ 180°), and (3) the $\angle$C – N – H..O dihedral angle (0 ~ 180°). The proton shielding dependence on the H-bond structure is illustrated in Fig. 4. Here, the amide proton of the complex is deshielded by the hydrogen bond and the deshieding effects weaken as the distance increases (especially in the 1.8 ~ 3.0 Å range). Figure 4 also shows a strong and somewhat complicated angle dependence ($\angle$N – H..O and $\angle$C – N – H..O) of the proton shielding: the amide proton becomes less deshielded as the dihedral angle approaches 90° (perpendicular to the NMA molecular plane). This probably arises from the change of electronic interactions between lone pair electrons of FA and N–H bond of NMA.

To provide a more quantitative interpretation, natural chemical shielding (NCS) analysis was performed for two families of structures: (1) the FA molecule was in the same plane as the NMA molecule and the $\angle$N – H..O angle was changed from 120 to 240°. (2) the $\angle$N – H..O angle was fixed at 120° and the $\angle$C – N – H..O dihedral angle was changed from 0 to 180°. In both cases, the r(H..O) length was 1.8 Å. Figure 5 compares the relative shieldings (to the average values) with two main contributions: (1) hyperconjugation (HC) contribution, defined as the sum of Lewis shieldings (n(O) and n(O) of FA) and non-Lewis shielding (σ*(N–H) of NMA); and (2) electrostatic (ES) contribution that can be approximately represented by the sum of Lewis shieldings (and of FA) and Lewis shielding (of NMA). For the $\angle$N – H..O angle variation (Fig. 5a), the HC contribution deshields the proton as the angle moves from 180° towards 120 and 240° because the *n*(FA)-*(NMA) interactions are getting stronger. The ES contribution shows the opposite trend with a similar magnitude. Thus, the net proton shieldings are little affected by the $\angle$N – H..O angle in the range (120 ~ 180°). Similar trends are shown for the $\angle$C – N – H . . . O dihedral angle

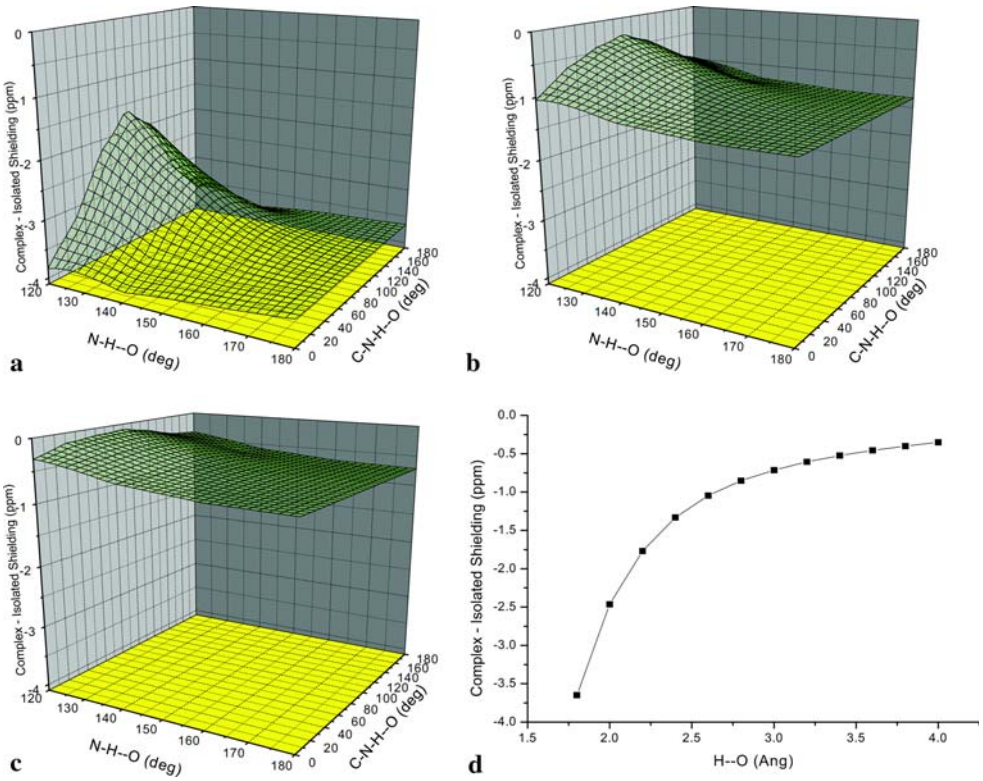

Fig. 4 Shielding variation of amide proton in the H-bonding region (NMA–FA complex): (**a**) r(H...O) = 1.8 Å; (**b**) r(H...O) = 2.4 Å; (**c**) r(H...O) = 3.0 Å; (**d**) r(H...O) = 1.8 ~ 4.0 Å

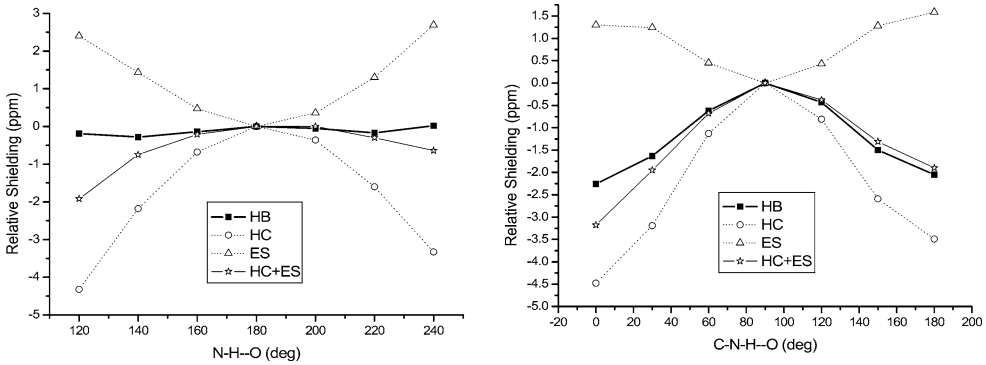

Fig. 5 Natural shielding analysis on the shielding variation of amide proton in the H-bond region: (**a**) for the $\angle N-H\ldots O$ angle; (**b**) for $\angle C-N-H\ldots O$ dihedral angle. In both cases, r(H...O) = 1.8 Å

variation (Fig. 5b) where the HC contribution deshields the proton as the angle goes from 90° towards 0 and 180° whereas the ES contribution makes the proton more shielded. However, in this case, the HC contribution dominates the shielding over the ES contribution.

The NMA–water molecule complexes were also used to estimate HB effects arising from solvent molecules on amide proton shielding. The same procedure as for the NMA–FA complexes was used. The shielding variation of amide proton by H-bonding with a water molecule is shown in Fig. 6. In this case, a strong shielding dependence of the amide proton shift is seen not only for distance changes, but also to angular movement of the water molecule. Unlike the case of NMA–FA complex, amide proton is more shielded as the $\angle N-H..O$ angle decreases (from 180° to 120°) and is more deshieded as the $\angle C-N-H..O$ dihedral angle goes to 90°. This is probably related to the fact that the lone pair electrons of the water molecule lie perpendicular to its molecule plane. Hence, their distance from the amide proton increases as the $\angle N-H..O$ angle decreases and decreases as the $\angle C-N-H..O$ dihedral angle goes to 90°.

The hyper-surfaces of the proton shifts in the H-bonding region have rather simple shape and thus they can be fitted by the combination of sine, cosine and exponential functions depending on the geometrical factors. This kind of approach is, however, limited to only simple cases. To overcome the limitation, artificial neural networks, which can handle even very complicated patterns, were employed to build empirical formulas for HB effects on the proton

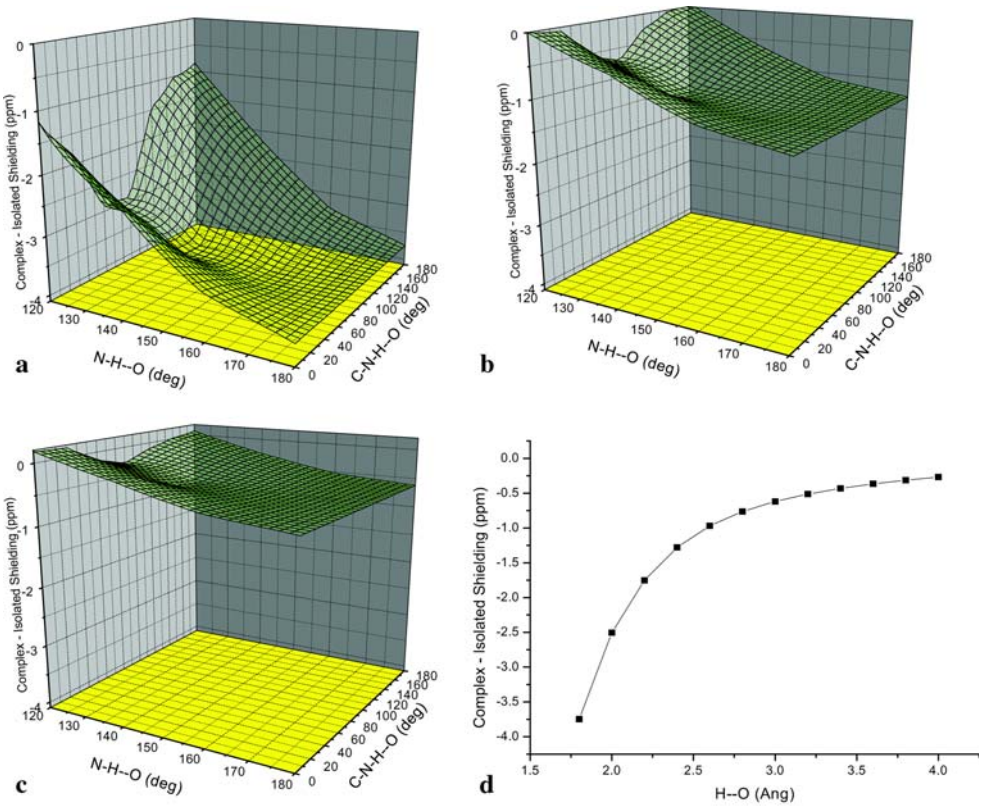

Fig. 6 Shielding variation of amide proton in the H-bonding region (NMA–water complex): (**a**) r(H...O) = 1.8 Å; (**b**) r(H...O) = 2.4 Å; (**c**) r(H...O) = 3.0 Å; (**d**) r(H...O) = 1.8 ~ 4.0 Å

shielding. The parameters generated by neural networks are collected in Table 1. Figure 7 shows the correlation between DFT shielding values and the ones generated by the empirical HB formulas for the NMA–FA complexes (Fig. 7a) and the NMA–water complexes (Fig. 7b). The neural networks give an excellent correlation with DFT model studies (NMA–FA: $R = 0.999$, standard deviation (SD) = 0.05 ppm, and slope = 0.998; NMA–water: $R = 0.999$, SD = 0.05 ppm, and slope = 0.987). These HB empirical terms were added into the SHIFTS program that has the empirical contributions from random coil, ring current, peptide group magnetic anisotropy, and electrostatic interactions.

Applications to protein structures based on X-ray crystallographic data

Using the new empirical method including quantum HB effects, amide proton shifts in 15 proteins (1,584 residues) were calculated and compared with experimental shifts. Since crystallographic techniques do not usually find accurate hydrogen atom positions, the X-ray structures were protonated by the program AMBER8 without further structure refinement. The overall information is collected in Table 2, and is compared with experimental data and other empirically predicted shifts in Fig. 8. Without HB effects, the SHIFTS program gives the weak correlation coefficient of 0.52, a root-mean-square (rms) error of 0.52 ppm, and a best-fit slope of 0.30. Much better results are provided by the SHIFTX program (Neal et al. 2003) that includes more accurate local covalent effects and distance-dependent HB effects. Here the correlation coefficient and rms error are 0.70 and 0.52 ppm, respectively. However, the dispersion is large and the best-fit slope of 0.55 suggests that some features leading to shift dispersion are not included in the model. There are several possible reasons for such discrepancies: (1) the use of X-ray structures different from the solution structures used in NMR experiments; (2) the uncertainty of the added proton positions; (3) the absence of dynamic averaging in the calculations; (4) the limited accuracy of empirical formulas used; and (5) the missing contributions of inter-residual HB effects and solvent effects. The first two problems may be addressed by more sophisticated theoretical methods that can handle molecular dynamics and electronic structure. This point will be discussed below. In the present context, we focus on the inter-residual HB effects and solvent effects by crystallographic water molecules. As seen in Fig. 8c, the inclusion of quantum HB effects separates the distribution into two parts, depending on whether or not they have H-bonds to other peptide groups. The residues exposed to solvent are little affected by HB effects but the inter- residual residues and residues interacting with crystallographic water molecules are greatly changed by the

Table 1 Parameters from Eq. 1 for the NMA–FA and NMA–H_20 complexes

	$w_{i1}^{(1)}$	$w_{i2}^{(1)}$	$w_{i3}^{(1)}$	$b_i^{(1)}$	$w_i^{(2)}$	$b^{(2)}$
NMA–FA						
i = 1	–24.2764	–59.2397	–9.7149	10.9108	–13.0623	–0.2389
i = 2	–28.9939	–72.3055	–13.1514	13.5998	8.3768	
i = 3	–52.4944	4.3521	0.4065	3.8866	9.2756	
i = 4	–284.9738	–320.0569	98.1535	76.7046	0.0817	
i = 5	–6.6013	–38.7931	52.9082	0.5972	0.2374	
i = 6	10.1706	42.7202	–43.3521	–3.4999	0.2747	
i = 7	–18.6367	–6.7689	0.1266	3.3561	1.2442	
i = 8	–48.7727	–58.5471	–13.9458	8.1723	462.1509	
i = 9	–27.6527	–61.7773	–7.4624	13.5389	0.9744	
i = 10	–29.9911	–73.5552	–42.1239	17.4542	0.1091	
NMA–H_2O						
i = 1	27.7788	47.5218	–22.3307	–7.0012	1.6469	–1.6656
i = 2	–28.8366	–56.9108	43.6805	4.9632	0.7936	
i = 3	–28.6409	–59.5435	–3203.534	13.8806	0.3249	
i = 4	–31.4878	–57.2285	–25.5537	12.4508	4.5670	
i = 5	–31.7386	–2.9465	3.1429	4.2145	1.8183	
i = 6	–16.8916	6074.2595	–2068.236	–536.5301	0.0209	
i = 7	–11.3876	1053.6816	480.2143	–159.4665	0.0231	
i = 8	–23.7841	19.3506	0.4662	3.1633	0.0761	
i = 9	–28.4133	–24.9065	–22.3992	7.4803	3.3093	
i = 10	–28.3461	–43.9640	–28.1438	10.1242	–8.6121	

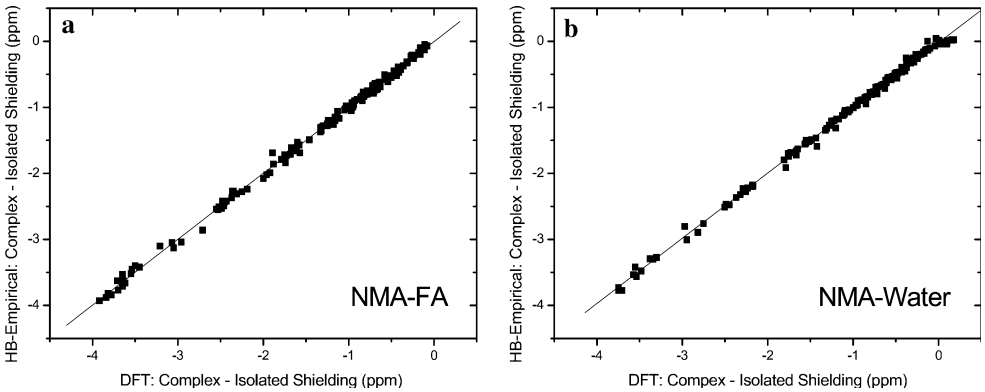

Fig. 7 Correlations between DFT shieldings and empirical shieldings where quantum HB effects were added: (**a**) for NMA–FA, R = 0.999, SD = 0.05 ppm, and slope = 0.998; (**b**) for NMA–water, R = 0.999, SD = 0.05 ppm, and slope = 0.987

HB effects (Fig. 8c). In Fig. 8d, the numerical statistical factors can be obtained by removing unchanged parts where a constant (2.19 ppm) is added to the predicted shifts for a direct comparison to the experimental data. The quantum HB effects improve the correlation to 0.661 but the shifts are more scattered (rms error = 0.78). The great improvement is seen in the slope of 1.02 that is almost unity. The HB effects can reflect correct sensitivities of amide proton shifts to the structural variation. In this approach, it is likely that some contributions are doubly counted because the magnetic anisotropies and electrostatic interactions of H-bond partner peptides are already included in the HB effects. However, their magnitudes by classical formulas are much smaller than those of quantum HB effects and thus it was not corrected in this approach.

Combination with molecular dynamics

The comparisons above clearly suffer from the fact that interactions with solvent are not well-represented by the coordinates present in an X-ray structure. Even when ''crystallographic'' waters are present, they represent only a fraction of the water actually present, and their static positions are probably not representative of the

Table 2 Proteins used for the shift predictions

#	Protein	PDB	BMRB	Resolution	# Residues
1	Hen lysozyme	2lzt	1,093	2.0	129
2	Dihydrofolate reductase	3dfr	(a)	1.7	162
3	Ribonuclease T1	2rnt	1,658	1.8	104
4	Oxidized thioredoxin	2trx	62	1.7	108
5	Trypsin inhibitor	5pti	5,359	1.0	58
6	PTI precursor	6pti	485	1.7	58
7	Ribonuclease A	3rn3	385	1.5	124
8	Chymotrypsin inhibitor	2ci2	1,869	2.0	83
9	Alpha-amylase inhibitor	1hoe	60	2.0	74
10	Cytochrome b5	1ehb	294	1.9	82
11	T4 lysozyme	3lzm	915	1.7	164
12	Ubiquitin	1ubq	5,387	1.8	76
13	Human lysozyme	1lz1	(b)	1.5	130
14	Myoglobin	1mbc	4,062	1.5	153
15	Cytochrome c551	451c	1,333	1.6	82

(a) NMR data from Polshakov et al. (1999); (b) NMR data from Redfield and Dobson (1990)

environment seen by an amide group. Molecular dynamics simulations offer one way to sample the solvent environment. To see how such data might be used, we extracted 100 snapshots (100 ps time interval) from a 10 ns solvated simulation of thioredoxin (see the Methods section.) Amide proton shifts were calculated using empirical methods and MD snapshots and averaged to be compared with experimental data. Figure 9 shows the correlations between the calculated shifts and experimental shifts. As seen in Fig. 9a, the shifts calculated by the pure empirical method (SHIFTS) show a similar correlation ($R = 0.551$) as the calculated shifts using X-ray structures. In this case, the slope (0.36) is still very low and there are no significant contributions from the motional averaging and solvent effects. By adding HB effects to the empirical shift calculations, things are greatly changed (in Fig. 9b). The correlation ($R = 0.671$) and slope (0.97) are improved and there is no clear shift separation between the water-exposed parts and inner parts of the protein as shown in the results using X-ray structures. However, the scatter is much worse than the pure empirical shifts. This discrepancy may mainly arise from the bad description of water–protein interactions. In fact, the water–NMA models used for HB effects seem to be too much simplified and the TIP3P water model using the MD simulations also seem to be not accurate enough to describe the actual local interactions between the protein and water solvent. Figure 9c shows the correlation without HB effects arising from water molecules. In this case, most far scatter points are removed and the correlation without the shifts of water-exposed parts is greatly improved ($R = 0.852$ and slope = 1.06). Even its scatter factor (SD = 0.55 ppm) is smaller than that (SD = 0.92 ppm) shown in Fig. 9b.

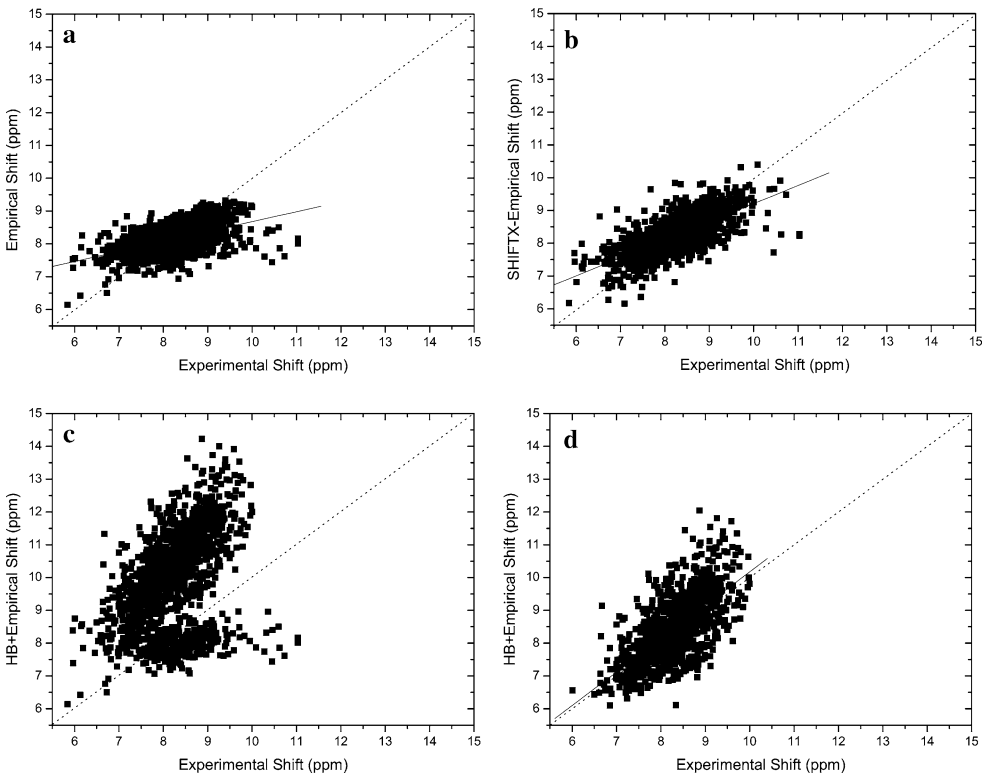

Fig. 8 Correlations between empirically calculated shifts and experimental shifts: (**a**) original SHIFTS results where $R = 0.519$, slope = 0.30, and rms error from experiment = 0.63 ppm; (**b**) SHIFTX results where $R = 0.704$, slope = 0.55, and rms error = 0.52 ppm; (**c**) SHIFTS results with quantum HB effects; (**d**) SHIFTS results with quantum HB effects for parts with inter-residual hydrogen bonds where $R = 0.661$, slope = 1.02, and rms error = 0.78 ppm

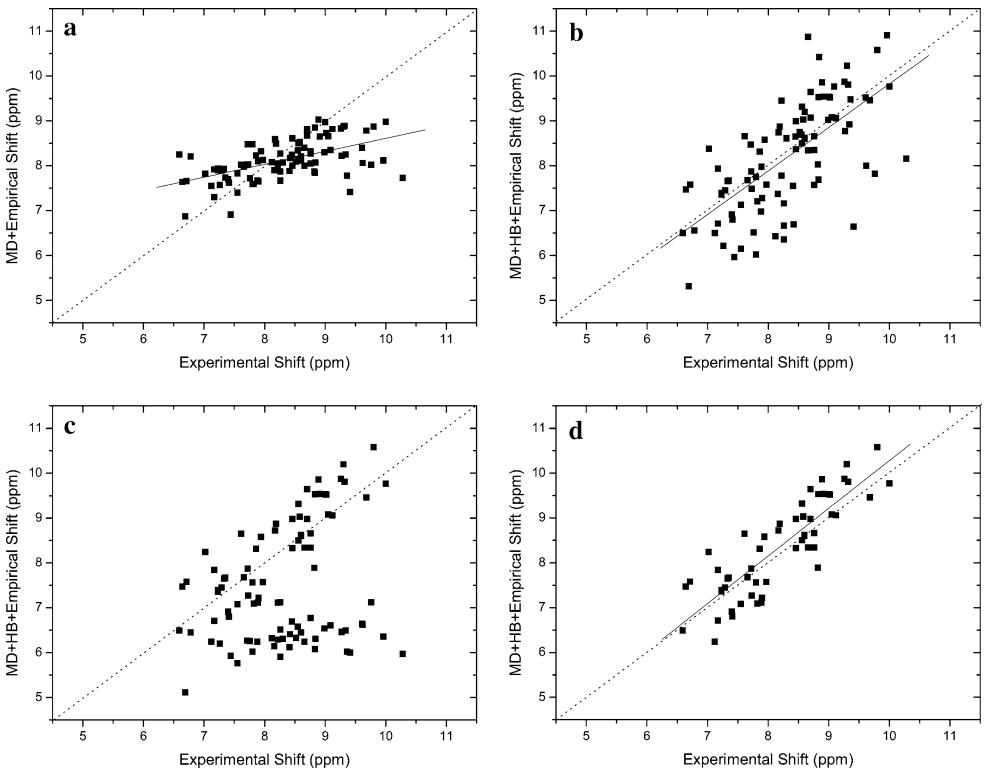

Fig. 9 Correlations between empirically calculated shifts and experimental shifts where 100 MD snapshots were used for the calculations: (**a**) original SHIFTS results where $R = 0.551$, slope = 0.38, and rms error from experiment = 0.73 ppm; (**b**) SHIFTS results with quantum HB effects where $R = 0.671$, slope = 0.97, and rms error = 0.92 ppm; (**c**) SHIFTS results with quantum HB effects where HB effects from water molecules were removed; (**d**) SHIFTS results with quantum HB effects for parts with inter-residual hydrogen bonds where $R = 0.852$, slope = 1.06, and rms error = 0.57 ppm

For a more detailed investigation of water HB effects, density-functional shielding calculations were carried out using protein fragments that include most important short-range interactions of an amide proton shift and the results are shown in Fig. 10. In this case, the calculated shifts were averaged over 100 MD snapshots to include motional and solvent effects. In Fig. 10a, it is shown that the DFT shifts are not well correlated with experimental data and some

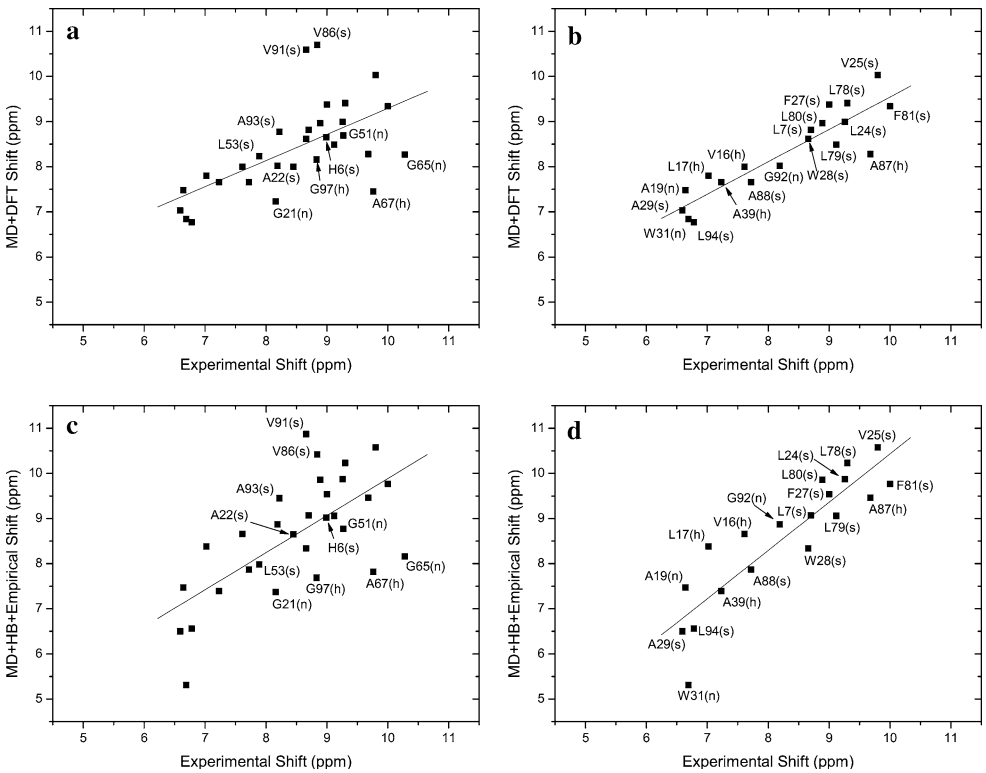

Fig. 10 Correlations between calculated shifts and experimental shifts where 100 MD snapshots were used for the calculations: (**a**) DFT results where $R = 0.613$, SD = 0.80 ppm, and slope = 0.58; (**b**) DFT results without water-exposed parts where $R = 0.901$, SD = 0.42 ppm, and slope = 0.72; (**c**) SHIFTS results with quantum HB effects where $R = 0.668$, SD = 0.82 ppm, and slope = 0.82; (**d**) SHIFTS results with quantum HB effects without water-exposed parts where $R = 0.889$, SD = 0.67 ppm, and slope = 1.07

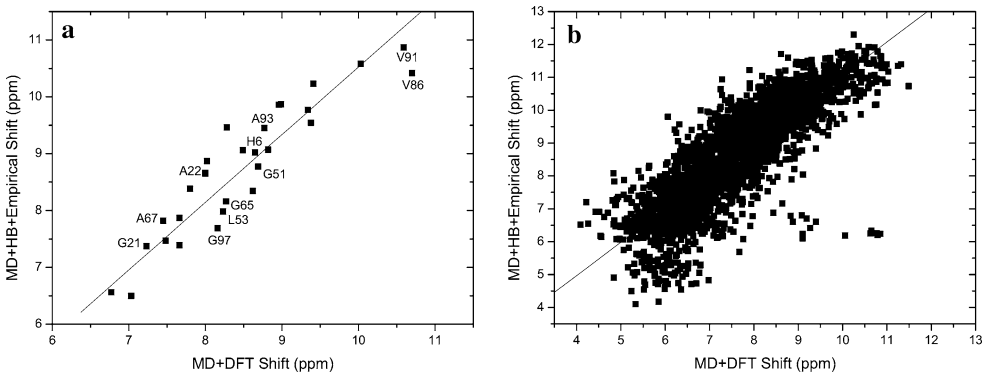

Fig. 11 Correlations between DFT and empirical shifts where 100 MD snapshots were used for the calculations: (**a**) averaged shifts over 100 snapshots were used (R = 0.914, SD = 0.54 ppm, and slope = 1.19); (**b**) all calculated shifts for 100 snapshots were used (R = 0.840, SD = 0.85 ppm, and slope = 1.02)

shifts are far scattered from the correlation line (R = 0.613, slope = 0.58, and SD = 0.80 ppm). However, the correlation (R = 0.901) excluding the shifts of water-exposed parts is better and its scatter (SD = 0.42 ppm) is smaller (Fig. 10b). The large errors arising from the water models tell us that the TIP3P model may not be good enough for the NMR study related to water–protein interactions. The empirical shifts with quantum HB effects show similar trends as the DFT ones (Fig. 10c, d) but the correlation (R = 0.889) and standard deviation (0.67 ppm) seen in Fig. 10d is slightly worse than DFT results. Finally the DFT shifts were compared with the empirical shifts that include quantum HB effects in Fig. 11. The correlation using averaged shifts over 100 MD snapshots is shown in Fig. 11a and the correlation using all shifts in 100 MD snapshots is shown in Fig. 11b. The empirical shifts in Fig. 11a are in good correlation (R = 0.914) with DFT shifts. Even the shifts of water-exposed residues are in good correlation and thus the errors shown in Fig. 10a may arise from the inaccurate TIP3P water models for water–protein interactions. The correlation of all individual shifts is slightly worse than the averaged one and the shifts are much more scattered. This means that the simplified NMA + FA and NMA + water models are not accurate enough for complicated residue and water motions but the average motions are similar to the model systems.

Conclusions

It has been recognized for some time that automated predications of amide proton shifts, starting from protein structure information, are significantly poorer than are comparable predictions for other nuclei. Values of correlation coefficients for multiple proteins are currently not much larger than 0.5, suggesting that only about half of the actual variability in chemical shift is captured in simple empirical models (Ösapay and Case 1991; Neal et al. 2003). Correlations between empirical models and quantum calculations on small model systems are often much better than this (Sitkoff and Case 1997; Barfield 2002; Parker et al. 2006), but it has not been easy to translate information garnered from small systems into effective predictions for proteins. One clear problem is that some amide protons can be hydrogen-bonded to other amide groups, and some to water, and the differences between these acceptors are large.

In this work, we have carried out a systematic study of hydrogen bonding effects on shifts in model systems, looking primarily at the three geometric variables shown in Fig. 1, and considering both water and amide hydrogen bonding partners. This focus must certainly represent to some extent and over-simplification, since other variables have been found to be important in similar studies (Barfield 2002; Parker et al. 2006). In particular, we know from some additional preliminary studies that allowing different values for the NH...O and H...OC angles can affect shielding at short distances. Nevertheless, the current models show good agreement with both model quantum data (see Fig. 7), as one would expect from the training used, but also for amide protons that are internally hydrogen-bonded in a reasonable protein data set (see Fig. 8). Predictions for amide-water interactions in proteins are still poor if only X-ray structures (and their associated ''crystallographic'' waters are used), but it looks like better results should be possible if snapshots from molecular dynamics can be used to provide a more realistic representation of the solvent environment (see Fig. 10). Of course, such simulations are time-consuming, and we report here only results for a single protein, thioredoxin. Further studies are clearly needed to know how general these results will be.

We have incorporated the empirical models used here into the latest version of the SHIFTS program (http://www.scripps.edu/case), where it can be used in conjunction with existing predictions for shifts of other protons, and for C and N shifts for backbone and Cβ atoms. It is worth noting that for some statistical applications (such as generating restraints or quality indices for low resolution structures) even the modest improvements shown in Fig. 8

are worthwhile. Better correlations would clearly be desirable in principle, and may need to rely on improvements in the structures used as input for the calculations as well as in the models that connect structure and chemical shift. The fact (seen here and elsewhere) that molecular dynamics results seem to provide broad improvements is encouraging, and we believe that the current results will provide a useful foundation for future efforts to understand how amide proton shifts respond to their microscopic environment.

Acknowledgments This work was supported by NIH grant GM45811. We thank Jan Ziegler, Stephan Schwarzinger and Jan Jensen for helpful discussions.

References

Asakura T, Taoka K, Demura M, Williamson MP (1995) The relationship between amide proton chemical shifts and secondary structure. J Biomol NMR 6:227–236

Barfield M (2002) Structural dependencies of interresidue scalar coupling ^{h3}JNC' and donor ^{1}H chemical shifts in the hydrogen bonding regions of proteins. J Am Chem Soc 124:4158–4168

Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98:5648–5652

Beger RD, Bolton PH (1997) Protein φ and ψ dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. J Biomol NMR 10:129–142

Bohmann JA, Weinhold F, Farrar TC (1997) Natural chemical shielding analysis of nuclear magnetic resonance shielding tensors from gauge-including atomic orbital calculations. J Chem Phys 107:1173–1184

Buckingham AD, Schaefer T, Schneider WG (1960) Solvent effects in nuclear magnetic resonance spectra. J Chem Phys 32:1227–1233

Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods R (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Cui Q, Karplus M (2000) Molecular properties from combined QM/MM methods 2. Chemical shifts in large molecules. J Phys Chem B 104:3721–3743

de Dios AC, Pearson JG, Oldfield E (1993) Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. Science 260:1491–1496

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012

Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb JA, Cheeseman JR, Keith TA, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, Pople JA (1998) Gaussian 98 (Revision A9). Gaussian Inc, Pittsburgh PA

Haigh CW, Mallion RB (1980) Ring current theories in nuclear magnetic resonance. Prog NMR Spectr 13:303–344

Iwadate M, Asakura T, Williamson MP (1999) Cα and Cβ carbon-13 chemical shifts in proteins from an empirical database. J Biomol NMR 13:199–211

Jeng M-F, Campbell AP, Begley T, Holmgren A, Case DA, Wright PE, Dyson HJ (1994) High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. Structure 2:853–868

Jorgensen WL, Chandrasekhar J, Madura J, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

Le H, Oldfield E (1994) Correlation between N NMR chemical shifts in proteins and secondary structure. J Biomol NMR 4:341–348

Le H, Oldfield E (1996) Ab initio studies of amide-N chemical shifts in dipeptides: applications to protein NMR spectroscopy. J Phys Chem 100:16423–16428

McConnell HM (1957) Theory of nuclear magnetic shielding in molecules I Long-range dipolar shielding of protons. J Chem Phys 27:226–229

Meiler J (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. J Biomol NMR 26:25–37

Meiler J, Maier W, Will M, Meusinger R (2002) Using neural networks for 13C NMR chemical shift prediction-comparison with traditional methods. J Magn Reson 157:242–252

Moon S, Case DA (2006) A comparison of quantum chemical models for calculating NMR shielding parameters in peptides: mixed basis set and ONIOM methods combined with a complete basis set extrapolation. J Comput Chem 27:825–836

Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. J Biomol NMR 26:215–240

Ösapay K, Case DA (1991) A new analysis of proton chemical shifts in proteins. J Am Chem Soc 113:9436–9444

Parker LL, Houk AR, Jensen JH (2006) Cooperative hydrogen bonding effects are key determinants of backbone amide proton chemical shifts in proteins. J Am Chem Soc 128:9863–9872

Pearson JG, Le H, Sanders LK, Godbout N, Havlin RH, Oldfield E (1997) Predicting chemical shifts in proteins: structure refinement of valine residues by using ab initio and empirical geometry optimizations. J Am Chem Soc 119:11941–11950

Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B 45:13244–13249

Polshakov VI, Birdsall B, Feeney J (1999) Characterization of rates of ring-flipping in trimethoprim in its ternary complexes with *Lactobacillus casei* dihydrofolate reductase and coenzyme analogues. Biochemistry 38:15962–15969

Redfield C, Dobson CM (1990) H NMR studies of human lysozyme: spectral assignment and comparison with hen lysozyme. Biochemistry 29:7201–7214

Rumelhart DE, McClelland J (1986) Parallel distrbuted processing. MIT Press, Boston

Sharma Y, Kwon OY, Brooks B, Tjandra N (2002) An ab initio study of amide proton shift tensor dependence on local protein structure. J Am Chem Soc 124:327–335

Sitkoff D, Case DA (1997) Density functional calculations of proton chemical shifts in model peptides. J Am Chem Soc 119:12262–12273

Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and Cα and Cβ C nuclear magnetic resonance chemical shifts. J Am Chem Soc 113:5490–5492

Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Meth Enzymol 338:3–34

Xu XP, Case DA (2001) Automated prediction of N, Cα, Cβ and C' chemical shifts in proteins using a density functional database. J Biomol NMR 21:321–333

Xu XP, Case DA (2002) Probing multiple effects on N, Cα, Cβ and C' chemical shifts in peptides using density functional theory. Biopolymers 65:408–423

Zupan J, Gasteiger J (1993) Neural networks for chemists. VCH Verlagsgesellschaft mbH, Weinheim