J Biomol NMR (2015) 62:403–409
DOI 10.1007/s10858-015-9958-z

ARTICLE

PPM_One: a static protein structure based chemical shift predictor

Dawei Li[1] · **Rafael Brüschweiler**[1,2]

Received: 13 December 2014 / Accepted: 12 June 2015 / Published online: 20 June 2015

Abstract We mined the most recent editions of the BioMagResDataBank and the protein data bank to parametrize a new empirical knowledge-based chemical shift predictor of protein backbone atoms using either a linear or an artificial neural network model. The resulting chemical shift predictor PPM_One accepts a single static 3D structure as input and emulates the effect of local protein dynamics via interatomic steric contacts. Furthermore, the chemical shift prediction was extended to most side-chain protons and it is found that the prediction accuracy is at a level allowing an independent assessment of stereospecific assignments. For a previously established set of test proteins some overall improvement was achieved over current top-performing chemical shift prediction programs.

Keyword Protein chemical shift prediction · Physics-based predictor · Backbone and side-chain chemical shifts · Neural network · Database analysis

Introduction

Chemical shifts represent the most ubiquitous NMR information of proteins. Chemical shifts of thousands of proteins have been deposited in the BioMagResDataBank (BMRB) (Ulrich et al. 2008). In addition to assignment purposes, chemical shift information has found wide-spread use in several protein NMR areas. These include the assessment of the propensity of protein segments to adopt various types of secondary structure (Shen and Bax 2013; Shen et al. 2009a; Wang and Jardetzky 2002; Wishart and Case 2001; Wishart and Sykes 1994), the determination or refinement of 3D protein structures (Cavalli et al. 2007; Rosato et al. 2012; Shen et al. 2008, 2009b; Wishart et al. 2008), the extraction of site-specific order parameters as measures of local dynamics (Berjanskii and Wishart 2006), the validation and improvement of molecular dynamics simulations and protein force fields (Li and Brüschweiler 2010, 2011), the optimization of enhanced conformational sampling techniques (Markwick et al. 2010), the comparison of protein structure and dynamics in solution and in crystals (Robustelli et al. 2012), and the prediction of rotating frame relaxation data (Xue et al. 2012). Most of these applications require the accurate prediction of chemical shifts from protein structures. Because of the quantum-chemical origin of chemical shifts, they have a rather complex dependence on molecular structure, which makes their accurate prediction a formidable challenge, especially for biomacromolecules such as proteins.

One of the early approaches to chemical shift prediction is based on approximate analytical relationships derived from quantum-chemical calculations, which forms the basis of the SHIFTS program (Xu and Case 2001, 2002). With the rapid expansions of both the BMRB (Ulrich et al. 2008) database for protein chemical shifts and the protein data bank (PDB) (Berman et al. 2000), empirical knowledge-based approaches have been developed in recent years that parameterize chemical shift hyper-surfaces as a function of atomic coordinates (Kohlhoff et al. 2009; Lehtivarjo et al. 2009, 2012; Li and Brüschweiler 2012; Neal et al. 2003; Sahakyan et al. 2011a, b; Shen and Bax 2007, 2010; Xu and

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9958-z) contains supplementary material, which is available to authorized users.

✉ Rafael Brüschweiler
bruschweiler.1@osu.edu

[1] Campus Chemical Instrument Center, The Ohio State University, Columbus, OH 43210, USA

[2] Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

Case 2001, 2002). At room temperature, the experimental chemical shift of a given nucleus reflects the Boltzmann-weighted average of the 'instantaneous' chemical shifts of a large number of conformational substates. Therefore, chemical shift predictors that are parametrized based on experimental chemical shifts of proteins in solution against static crystal structures implicitly include some degree of dynamics averaging (Lehtivarjo et al. 2009, 2012; Li and Brüschweiler 2012). In addition, some predictors, such as SPARTA+, include S^2 order parameters that are predicted from local interatomic contacts to account for some dynamics averaging effects (Shen and Bax 2010; Zhang and Brüschweiler 2002). When a realistic ensemble of a protein is available, one can calculate chemical shifts of each static structure and take the ensemble average to obtain the predicted values. By fitting the ensemble averaged, back-calculated chemical shifts against experimental values, a set of 'static' model parameters can be obtained. Both 4DSPOT (Lehtivarjo et al. 2009, 2012) and PPM (Li and Brüschweiler 2012) follow this strategy by parametrizing chemical shifts using conformational ensembles generated by MD instead of static protein structures. In other cases, structure-based prediction has been augmented with sequence-based chemical shift information (Han et al. 2011; Wishart et al. 1997).

While most of the developments in protein chemical shift prediction have focused on backbone nuclei, amino-acid side-chain chemical shifts are also informative reporters on protein structure and dynamics (Sahakyan et al. 2011b). Several software packages are currently available for the prediction of at least some of the side-chain ^{1}H chemical shifts (Han et al. 2011; Lehtivarjo et al. 2009; Neal et al. 2003; Sahakyan et al. 2011a, b; Xu and Case 2001, 2002).

The steady expansions of both the BMRB and PDB databases in recent years provide the means to continuously improve empirical knowledge-based chemical shift prediction by using a larger number of descriptors of protein structure, e.g. higher order Fourier series expansions in dihedral angle space. The present work makes extensive use of this possibility. Moreover, the prediction has been expanded to most protein side-chain protons. To distinguish this new chemical shift predictor from our ensemble based predictor PPM (Li and Brüschweiler 2012), the new predictor and its underlying software is called PPM_One.

Method

BMRB and PDB input file lists

An up-to-date list of protein chemical shifts of the BMRB was obtained by downloading all chemical shift files in the NMR-STAR 2.1 format (Ulrich et al. 2008). For each file, we checked the BMRB "System_physical_state" and the "Mol_system_component_name" information to ensure that the protein was assigned in its native apo state, i.e. in the absence of ligands. All entries that did not fulfill these properties were removed. Intrinsically disordered proteins have not been excluded at this stage, but they are eliminated in subsequent steps. Next, we checked for each BMRB entry all PDB files from the PDB database (Berman et al. 2000) whether they matched the sequence of a corresponding BMRB chemical shift file, and selected only those proteins solved by X-ray crystallography with a resolution of 2.0 Å or better. In this process, PDB files with missing, mutated, or otherwise modified residues (except for residues in the two terminal regions) were excluded. PDB files with bound ligands or in complex with other proteins were kept after the ligands and interacting proteins were removed. PDB files that contained multiple asymmetric chains were split into multiple files. For most BMRB entries, there was usually more than one matching PDB file. We then employed the SHIFTS (Xu and Case 2001, 2002) program to predict chemical shifts of all these PDB files and selected the one that had the lowest chemical shift root-mean-square deviation (RMSD), which is the average RMSD value of all predicted Cα, Cβ, and C′ chemical shifts with respect to experiment. In rare cases, the lowest chemical shift RMSD was larger than 3 ppm, which may indicate an unusual protein geometry, miscalibration, or misassignment. These BMRB–PDB pairs were removed. We then did sequence alignment for all remaining pairs and removed all homologs whose sequence similarity was larger than 50 %. This left us with 405 BMRB–PDB pairs, which are listed in Supporting Information Table S1 and were utilized to train PPM_One. Among these, eight proteins that had been selected for testing in a previous study (Shen and Bax 2010) were used for subsequent tests. These eight proteins were not used during any of the training stages. The sequence similarity between proteins in the training set and the test set was less than 20 %. To correct for possible referencing errors, we applied referencing corrections to carbon chemical shifts of all PDB BMRB pairs by comparing the deposited data with the predictions from SPARTA+. For each protein, the experimental chemical shifts were subjected to an offset to optimally match the SPARTA+ prediction. In this process, all carbon chemical shifts of the same protein were uniformly shifted. Using PPM_One for re-referencing (in an iterative way) or using shifts, instead of SPARTA+, did not affect the results. Nitrogen and hydrogen chemical shifts were not corrected because the referencing corrections were found to be very small.

Linear model for backbone chemical shift prediction

When a part of a protein is highly flexible adopting many different conformations, the experimental chemical shifts tend to approach the so-called random coil value. It has been shown that the explicit inclusion of local flexibility into chemical shift prediction improves the prediction accuracy (Shen and Bax 2010). To account for this effect, we expressed the predicted chemical shifts δ_k^{pre} as follows:

$$\delta_k^{pre} = \delta_k^{pre'} f(C_k) + \delta_k^0 (1 - f(C_k)) \quad (1)$$

$$\delta_k^{pre'} = p_k^{ring-current} \delta_k^{ring-current} + p_k^{magn-aniso} \delta_k^{magn-aniso} + p_k^{dihedral} \delta_k^{dihedral} + p_k^{h-bond} \delta_k^{h-bond} + p_k^{seq} \delta_k^{seq} \quad (2)$$

$$f(x) = \left(1 - \exp\left(-\frac{2(x - x_{\min,k})}{x_{\max,k} - x_{\min,k}}\right)\right) \Big/ \left(1 + \exp\left(-\frac{2(x - x_{\min,k})}{x_{\max,k} - x_{\min,k}}\right)\right) \quad (3)$$

$\delta_k^{pre'}$ is obtained by expressing the chemical shift as a linear combination of different structural descriptors with weights p_k. These terms will be described below. Subscript k accounts for different types of nuclei, namely Cα, Cβ, C′, Hα, N, and H^N. C_k is the contact sum, which has been shown to be able to predict residue flexibility from an average structure quite well (Li and Brüschweiler 2009; Zhang and Brüschweiler 2002). The sigmoid function f(x) gradually converts $\delta_k^{pre'}$ to the "chemical shift offset" δ_k^0 when the contact sum C_k approaches $x_{min,k}$.

Chemical shifts are sensitive to chemical substituent effects, which can be accounted for by the inclusion of a correction term that depends on the type of the amino acid of interest and the amino acids that directly precede and follow. With 20 different amino acid types, the sequence-dependent chemical shift parametrization contains 60 parameters (descriptors).

Backbone and side-chain dihedral angles of consecutive residue triples, i.e. the "previous", "current", and the "following" residue, provide a detailed description of the local chemical environment and they are usually important for chemical shift predictions (Kohlhoff et al. 2009; Neal et al. 2003; Shen and Bax 2010). Fourier series expansion up to 3rd order in the dihedral angles of the current residue were employed both for cosine and sine functions with the six Fourier coefficients (prefactors) serving as fit parameters. This leads to six descriptors for each dihedral angle. To optimize the fit stability, the first two side-chain dihedral angles were included for amino acids with two or more side-chain dihedral angles. Because the different chemical properties of the 20 amino-acid types affect their chemical shifts differently, the dihedral angle descriptors were treated for each amino-acid type separately. Together, there is a total of up to 480 descriptors for the dihedral angle contribution of the current residue (20 amino-acid types, up to 4 dihedral angles (φ, ψ, χ_1, and χ_2) with 6 Fourier coefficients per dihedral angle). In addition, Fourier series expansions up to 2rd order of the backbone dihedral angles and the first side chain dihedral angle of the previous and following residues were also employed. Following a similar strategy as for the current amino acid, the dihedral angle descriptors of the previous (following) residue were treated differently according to amino-acid type of previous (following) residue. Note that these dihedral angle descriptors did not depend on the amino-acid type of the current residue. The dihedral angle contribution for the previous and following residue each contains 240 descriptors.

The effect of backbone hydrogen bonds on chemical shifts was parameterized in terms of the inverse of the distance r_{OH} between the acceptor O and donor H^N atoms. These contributions were set to zero if $r_{OH} > 3$ Å or either of the angles formed by the NH^NO or H^NOC' atom triples was smaller than 120°. To improve the stability of the fits, the hydrogen bond descriptors were treated as amino-acid type independent. Because each residue contained both a donor and an acceptor group, hydrogen bond terms were parametrized by six descriptors.

Ring current and magnetic anisotropic effect were only applied to proton chemical shifts. Ring current effects $\delta_{ring-current}^{(k)}$ could arise from five different aromatic amino acid rings, namely the ones in Phe, Tyr, His, and the 5-ring and 6-ring of Trp (Trp-5, Trp-6). This geometric descriptor is defined as (Haigh and Mallion 1972, 1979; Osapay and Case 1991; Sahakyan et al. 2011a)

$$f_{ring\ current} = \sum_{p,q} S_{pq} \left(\frac{1}{r_p^3} + \frac{1}{r_q^3}\right) \quad (4)$$

where the sum includes all adjacent atom pairs in the ring. r_p and r_q are the distances between neighboring ring atoms p and q to the proton and S_{ij} is the area of the triangle formed by atom i, atom j and the projection of the proton on the aromatic ring (denoted below as "o"). The sign of S_{ij} is determined whether the vector product $\mathbf{t}_{oi} \times \mathbf{t}_{ij}$ is parallel (positive sign) or antiparallel (negative sign) to the ring normal defined by $\mathbf{t}_{12} \times \mathbf{t}_{23}$, where $\mathbf{t}_{oi}$ is the vector pointing from o to atom i and and $\mathbf{t}_{ij}$ as the vector pointing from atom i to atom j. Magnetic anisotropy effects $\delta_{magn-aniso}^{(k)}$ were calculated using the axially symmetric model by Osapay and Case (1991) following McConnell's (1957) formulation of anisotropy effects of peptide groups:

$$f_{magn-aniso} = \frac{1}{r^3}\left(3\cos^2\theta - 1\right) \quad (5)$$

where r is the distance from the proton to the peptide amide group (formed by the OC′N atoms) and θ is the angle between the vector joining the proton to the amide group and the amide group normal. Similar to Sahakyan et al. (2011a), an analogous treatment is employed for the OCN side-chain groups of residues Asn and Gln, for the OCO side-chain groups of Glu and Asp, and for the NCN side-chain group of Arg. Ring current and magnetic anisotropy effects totally contributed 9 descriptors.

The parameters $x_{min,k}$ and $x_{max,k}$ of Eq. (3) enter $f(C_k)$ and $1-f(C_k)$ in Eq. (1) in a non-linear way and, hence, they cannot be optimized by linear regression. Therefore, to obtain optimal x_{min} and x_{max} values we first did a grid search for each atom type by minimizing the RMSD between δ_k^{pre} and δ_k^{exp}. Next, all prefactors [p_k in Eq. (2)] together with δ_k^0 were determined by linear regression with respect to δ_k^{exp} and the RMSD between δ_k^{exp} and δ_k^{pre} was determined. In our linear regression procedure, 10 % of the data points were randomly excluded and used only for validation. To make the predictor more robust, this process was repeated 1000 times and the model parameters obtained from all fits were averaged.

Artificial neural network based prediction

A single hidden layer feed forward artificial neural network (ANN) was applied to predict chemical shifts of all backbone sites, following ideas behind PROSHIFT (Meiler 2003) and SPARTA+ (Shen and Bax 2010). Our input layer included 113 nodes for non-proton atoms and 122 nodes for protons. The first 60 nodes accounted for the triple residue sequence. As in SPARTA+ we utilized an amino acid similarity table, which helped improve the prediction accuracy of the ANN slightly. The ψ, χ_1, χ_2 dihedral angles of the previous residue, ϕ, ψ, χ_1, χ_2 dihedral angles of the current residue, and ϕ, χ_1, χ_2 dihedral angles of the following residue were included to account for the local structural information. More specifically, Fourier series expansions up to second order of each of these dihedral angles were included as elements in the input layer. By contrast, for SPARTA+ only the first order Fourier coefficients are included. Hydrogen bond information of the CO group of the previous residue, the NH group of the following residue and both groups of the current residue were also included in the input layer. Each of the four groups contains three parameters, namely the inverse of the hydrogen bond length (H…O distance) and the cosines of the NHO and HOC angles. For the prediction of Cα, Cβ, and C′ chemical shifts, the contact sum of the corresponding atom was included as a parameter; for the prediction of N, H^N, Hα chemical shifts, the Cα contact sum of the same residue was used instead. For the Hα and HN atoms, nine additional parameters were included for the ring current and magnetic anisotropic effects.

The hidden layer of our ANN included 25 neurons. The transfer function from the input to the hidden layer was $f(x) = (1-e^{-2x})/(1+e^{-2x})$. And the transfer function from the hidden layer to the output layer was $f(x) = x$. A flowchart of the ANN is depicted in Fig. 1b of the work by Shen and Bax (2010). During our ANN training process, 20 % of randomly picked data points were used for validation to prevent over-training (overfitting). The Levenberg–Marquardt optimization algorithm was used during training. To reduce the statistical uncertainties in the model parameters and improve the prediction performance, the ANN was trained 12 times using different initial parameters and the averages of the output model parameters were taken for chemical shift prediction. Averaging over a larger number of training calculations provides no further improvement.

Side-chain proton chemical shifts prediction

Our side chain proton chemical shift predictor mainly depends on the linear combination of ring current effects, magnetic anisotropic effects, and an overall chemical shift offset. For all side chain Hβ atoms, Fourier series expansions up to 2rd order of the backbone dihedral angles of the same residue were also included. The prediction is based on

$$\delta_m^{pre} = p_m^{ring-current}\delta_m^{ring-current} + p_m^{magn-aniso}\delta_m^{magn-aniso} + p_m^{dihedral}\delta_m^{dihedral} + \delta_m^0 \quad (6)$$

where index m reflects each different proton type. The dihedral angle terms are used only for Hβ proton chemical shifts.

For many geminal proton pairs, such as Hβ2 and Hβ3, the BMRB indicates that they were stereospecifically assigned (i.e. they were annotated with ambiguity code 1). To check the consistency of these specific assignments with the chemical shift prediction according to Eq. (6), we performed two sets of calculations. In the first set, the mean chemical shift of geminal proton pairs was predicted and compared with the experimental mean chemical shift, while for the second set individual protons were predicted and compared with the corresponding experimental chemical shifts based on their stereospecific assignments. Based on statistical arguments, the first set should lead to a smaller RMSD than the 2nd set. For the purpose of performance comparison we correct for this fact by using a "relative prediction error", which is given by the RMSD of the prediction with respect to experiment divided by the standard deviation of the experimental chemical shifts themselves. If the prediction of the first set improved merely because of statistical averaging effects, the relative prediction error of the two sets should be similar.

For each proton type, the corresponding fitting parameters p_k in Eq. (6) were determined by linear regression.

For many proton types, the limited availability of experimental data caused the fitting to become unstable. This issue was addressed by treating the fit parameters $p^{ring-current}$ and $p^{magn-aniso}$ independently of the proton-type.

Result and discussion

Prediction accuracies of backbone atoms from both our linear model and the ANN model are displayed in Fig. 1 (the raw data are listed in Table S2 of the Supporting Information) in comparison with SPARTA+ and SHIFTX2 (without using its sequence-based predictor). The reported chemical shift RMSDs correspond to averages from the eight test proteins used previously (Shen and Bax 2010). Overall, our *linear model* of PPM_One achieved an accuracy that is comparable with the one of SPARTA+ for Cα, Cβ, and Hα atoms. However, the prediction accuracies of other atoms were slightly worse than SPARTA+, possibly because of non-linear contributions to their chemical shifts from the hydrogen bond. The linear model without contact sum correction is slightly worse. By contrast, our *ANN model* achieved consistently better performance than SPARTA+ for all atoms except the C′ carbonyl carbons. However, since C′ chemical shifts are only available for four of the eight proteins in the test set, the statistical uncertainty for these atoms is larger than for the other backbone atoms. Our ANN model also provides a slight improvement over SHIFTX2.

The histogram in Fig. 2 shows the relative prediction errors for the average chemical shift of geminal protons (set 1, blue bars) and for individual protons using the stereospecific assignment information from the BMRB (set 2, red bars). The significantly reduced relative prediction error of set 1 over set 2 suggests that a non-negligible number of the geminal protons reported in the BMRB have been stereospecifically misassigned (despite the fact that they are annotated with ambiguity code 1), consistent with previous findings (Borowski 2012; Williamson and Asakura 1992). The same behavior has been observed when chemical shifts were predicted using the program SHIFTS (Xu and Case 2001, 2002). Based on these results, it is recommended to use the average chemical shift of geminal proton pairs, rather than individual chemical shifts, for back-calculation purposes to prevent the inflation of side-chain chemical RMSDs due to wrong stereospecific assignments. These results indicate that the prediction accuracy of PPM_One of these side-chain proton chemical shifts is at a level where it can be directly used to independently confirm stereospecific assignments for proteins with known structure (Nilges et al. 1990).

The prediction accuracy of side-chain protons is depicted in Fig. 3 (raw data are listed in Table S3), in comparison with CH3Shifts (Sahakyan et al. 2011a) (methyl protons only), ArShifts (Sahakyan et al. 2011b) (aromatic protons only), SHIFTS (Xu and Case 2001, 2002), and the RMSD of the experimental data alone. Because the eight protein test set did not include a sufficiently large number of data points for many of the proton types to achieve reliable statistics, the reported RMSD of each proton type was calculated from all proteins of the training set. To avoid overfitting, special attention was given to the stability of the fits during the training process and was shown not to be a problem because of the limited number of terms included for the side-chain chemical shift prediction. A list of protons that were trained globally rather than

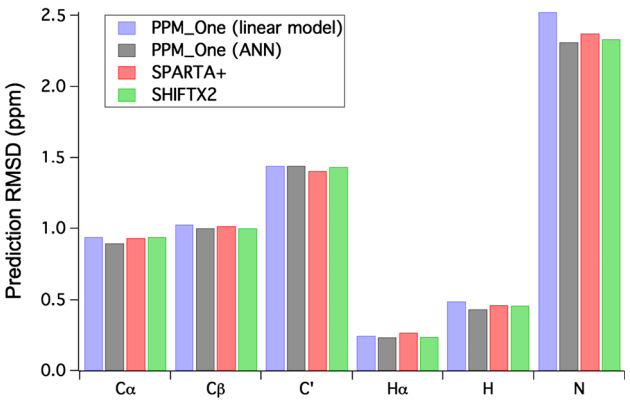

Fig. 1 Performance of the prediction of backbone chemical shifts by PPM_One (*blue* and *black bars* for liner and ANN models, respectively) compared with SPARTA+ (*red bars*) and SHIFTX2 (*green bars*). The chemical shifts of eight test proteins were used (see text)

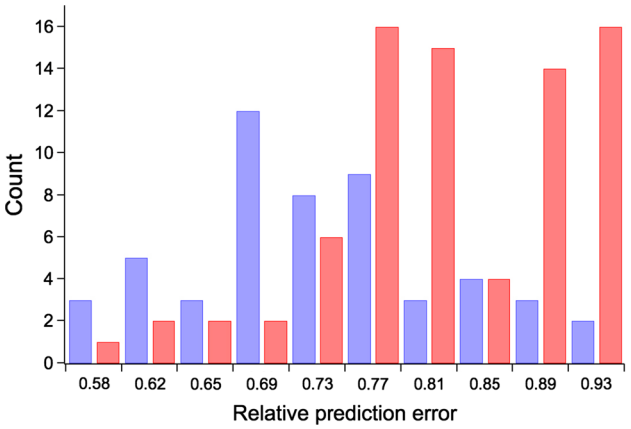

Fig. 2 Histogram of relative chemical shift prediction error of model 1 (*blue bars*) and model 2 (*red bars*) for 52 (model 1) or 78 (model 2) types of geminal protons, respectively, of all training set proteins. Model 1 predicts the mean chemical shift of geminal protons while model 2 predicts the chemical shifts of individual protons. The relative prediction error is defined as the RMSD of the prediction with respect to experiment divided by the standard deviation of the experimental data set itself

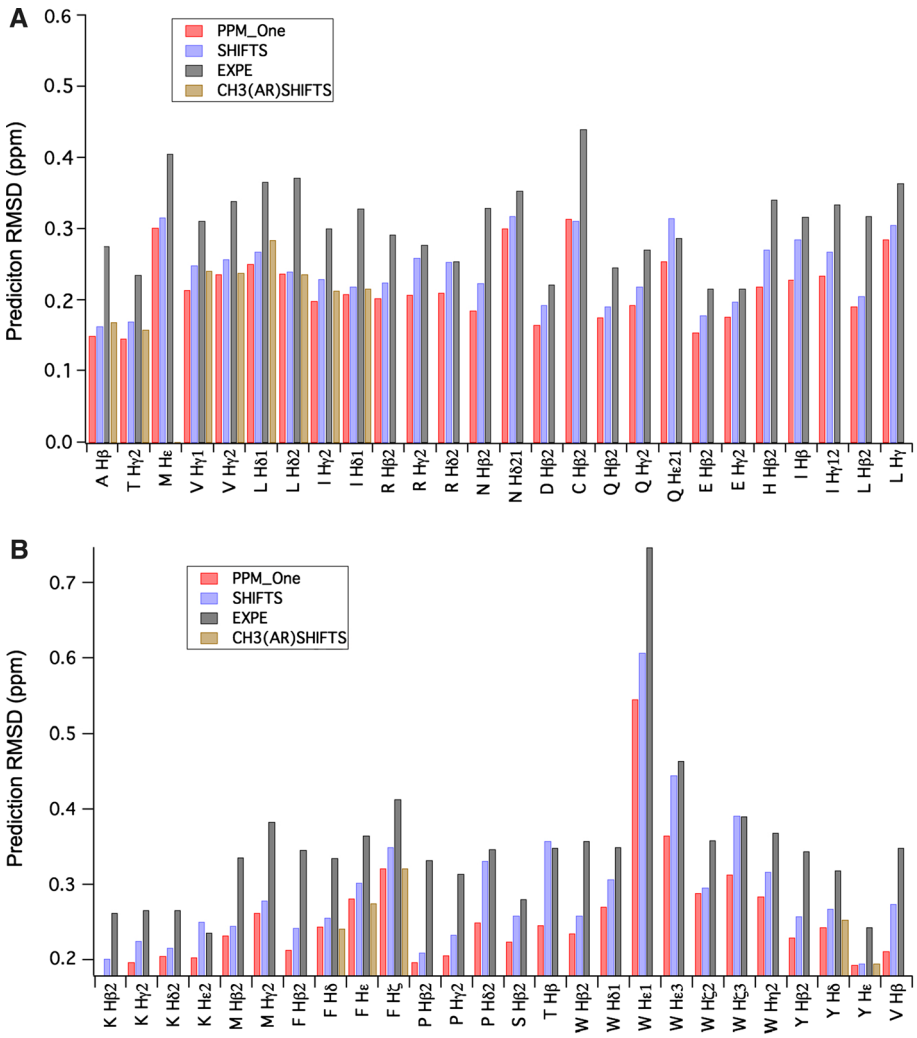

Fig. 3 Performance of the prediction of side-chain proton chemical shifts by PPM_One (*red*) compared with SHIFTS (*blue*), CH3SHIFTS & ARSHIFTS (*brown*), and the RMSD of only the experimental values (*gray*)

individually is given in the Supporting Information. The prediction accuracy varies for different side-chain proton types, whereby Hβ protons were predicted with higher accuracy than the other side-chain protons. In principle, local electric fields can also have an effect on proton chemical shifts. However, inclusion of such effects in our prediction showed no improvement (Li and Brüschweiler 2012). Certain protons have a very limited number of experimental values and proved to be hard to predict using our model, hence they were excluded. These include all side-chain O–H hydroxyl and carboxyl protons, Hγ of Cysteine, Hε of Glutamine, Hζ and Hε of Lysine, Hε of Arginine and Hε1, Hε2, Hδ1, Hδ2 ring protons of Histidine. The large prediction errors of O–H protons might be related to their participation in possible hydrogen-bonding effects. For Cysteines, disulfide bonds cannot be inferred from the PDB structures with good reliability, which introduces a significant uncertainty in the prediction of Cys Hγ chemical shifts. For the three often positively charged amino acids (Lys, Arg, His), our model did not include any protonation state information and possible water mediated interactions, which might have contributed to the low prediction quality for the protons listed above.

In summary, the new PPM_One chemical shift predictor provides an improved empirical chemical shift prediction of backbone atoms and side-chain protons from single static protein structures. PPM_One is readily applicable to a wide range of biophysical NMR applications that rely on chemical shifts predicted at the presently highest achievable accuracy. PPM_One also offers a useful reference for the assessment of the quality of predicted chemical shifts for comparison with chemical shifts predicted from conformational ensembles. The ability of PPM_One to predict chemical shifts of most side-chain protons is expected to further broaden the scope of chemical shift applications. The integrated PPM/PPM_One web server, the source code including all fitting parameters, and a downloadable executable program are available at http://spin.ccic.ohio-state.edu/index.php/ppm.

Acknowledgments This work was supported by the NSF (Grant MCB 1360966). R.B. is an Ohio Research Scholar.

References

Berjanskii M, Wishart DS (2006) NMR: prediction of protein flexibility. Nat Protoc 1:683–688

Berman HM et al (2000) The protein data bank. Nucleic Acids Res 28:235–242

Borowski P (2012) Conformational analysis of the chemical shifts for molecules containing diastereotopic methylene protons. J Magn Reson 214:1–9

Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci USA 104:9615–9620

Haigh CW, Mallion RB (1972) New tables of ring current shielding in proton magnetic-resonance. Org Magn Reson 4:203

Haigh CW, Mallion RB (1979) Ring current theories in nuclear magnetic-resonance. Prog Nucl Magn Reson Spectrosc 13:303–344

Han B, Liu YF, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43–57

Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. J Am Chem Soc 131:13894

Lehtivarjo J, Hassinen T, Korhonen SP, Perakyla M, Laatikainen R (2009) 4D prediction of protein H-1 chemical shifts. J Biomol NMR 45:413–426

Lehtivarjo J, Tuppurainen K, Hassinen T, Laatikainen R, Perakyla M (2012) Combining NMR ensembles and molecular dynamics simulations provides more realistic models of protein structures in solution and leads to better chemical shift prediction. J Biomol NMR 52:257–267

Li DW, Brüschweiler R (2010) NMR-based protein potentials. Angew Chem Int Ed 49:6778–6780

Li DW, Brüschweiler R (2011) Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. J Chem Theory Comput 7:1773–1782

Li DW, Brüschweiler R (2009) All-atom contact model for understanding protein dynamics from crystallographic B-factors. Biophys J 96:3074–3081

Li DW, Brüschweiler R (2012) PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. J Biomol NMR 54:257–265

Markwick PRL, Cervantes CF, Abel BL, Komives EA, Blackledge M, McCammon JA (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. J Am Chem Soc 132:1220

McConnell HM (1957) Theory of nuclear magnetic shielding in molecules. 1. 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

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

Nilges M, Clore GM, Gronenborn AM (1990) 1H-NMR stereospecific assignments by conformational data-base searches. Biopolymers 29:813–822

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

Robustelli P, Stafford KA, Palmer AG (2012) Interpreting protein structural dynamics from NMR chemical shifts. J Am Chem Soc 134:6365–6374

Rosato A et al (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20:227–236

Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M (2011a) Structure-based prediction of methyl chemical shifts in proteins. J Biomol NMR 50:331–346

Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M (2011b) Using side-chain aromatic proton chemical shifts for a quantitative analysis of protein structures. Angew Chem Int Ed 50:9620–9623

Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 38:289–302

Shen Y, Bax A (2010) SPARTA plus: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR 48:13–22

Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR 56:227–241

Shen Y et al (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690

Shen Y, Delaglio F, Cornilescu G, Bax A (2009a) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223

Shen Y, Vernon R, Baker D, Bax A (2009b) De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR 43:63–78

Ulrich EL et al (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408

Wang YJ, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci 11:852–861

Williamson MP, Asakura T (1992) The application of 1H NMR chemical shift calculations to diastereotopic groups in proteins. FEBS Lett 302:185–188

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

Wishart DS, Sykes BD (1994) The C-13 chemical-shift index—a simple method for the identification of protein secondary structure using C-13 chemical-shift data. J Biomol NMR 4:171–180

Wishart DS, Watson MS, Boyko RF, Sykes BD (1997) Automated H-1 and C-13 chemical shift prediction using the BioMagResBank. J Biomol NMR 10:329–336

Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:W496–W502

Xu XP, Case DA (2001) Automated prediction of 15 N, 13Calpha, 13Cbeta and 13C′ chemical shifts in proteins using a density functional database. J Biomol NMR 21:321–333

Xu XP, Case DA (2002) Probing multiple effects on 15 N, 13C alpha, 13C beta, and 13C′ chemical shifts in peptides using density functional theory. Biopolymers 65:408–423

Xue Y, Ward JM, Yuwen TR, Podkorytov IS, Skrynnikov NR (2012) Microsecond time-scale conformational exchange in proteins: using Long molecular dynamics trajectory to simulate NMR relaxation dispersion data. J Am Chem Soc 134:2555–2562

Zhang F, Brüschweiler R (2002) Contact model for the prediction of NMR N–H order parameters in globular proteins. J Am Chem Soc 124:12654–12655