Protein structure validation and refinement using amide proton chemical shifts derived from quantum mechanics

Christensen, Anders Steen; Linnet, Troels Emtekær; Borg, Mikael; Boomsma, Wouter Krogh; Lindorff-Larsen, Kresten; Hamelryck, Thomas Wim; Jensen, Jan Halborg

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0084123

Publication date:
2013

Document version
Publisher's PDF, also known as Version of record

OPEN ACCESS Freely available online

PLOS ONE

Protein Structure Validation and Refinement Using Amide Proton Chemical Shifts Derived from Quantum Mechanics

Anders S. Christensen[1]*, Troels E. Linnet[2], Mikael Borg[3], Wouter Boomsma[2], Kresten Lindorff-Larsen[2], Thomas Hamelryck[3], Jan H. Jensen[1]

1 Department of Chemistry, University of Copenhagen, Copenhagen, Denmark, **2** Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Copenhagen, Denmark, **3** Structural Bioinformatics Group, Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark

Abstract

We present the ProCS method for the rapid and accurate prediction of protein backbone amide proton chemical shifts - sensitive probes of the geometry of key hydrogen bonds that determine protein structure. ProCS is parameterized against quantum mechanical (QM) calculations and reproduces high level QM results obtained for a small protein with an RMSD of 0.25 ppm ($r = 0.94$). ProCS is interfaced with the PHAISTOS protein simulation program and is used to infer statistical protein ensembles that reflect experimentally measured amide proton chemical shift values. Such chemical shift-based structural refinements, starting from high-resolution X-ray structures of Protein G, ubiquitin, and SMN Tudor Domain, result in average chemical shifts, hydrogen bond geometries, and trans-hydrogen bond ($^{h3}J_{NC'}$) spin-spin coupling constants that are in excellent agreement with experiment. We show that the structural sensitivity of the QM-based amide proton chemical shift predictions is needed to obtain this agreement. The ProCS method thus offers a powerful new tool for refining the structures of hydrogen bonding networks to high accuracy with many potential applications such as protein flexibility in ligand binding.

Citation: Christensen AS, Linnet TE, Borg M, Boomsma W, Lindorff-Larsen K, et al. (2013) Protein Structure Validation and Refinement Using Amide Proton Chemical Shifts Derived from Quantum Mechanics. PLoS ONE 8(12): e84123. doi:10.1371/journal.pone.0084123

Editor: Freddie Salsbury, Wake Forest University, United States of America

Received July 24, 2013; **Accepted** November 11, 2013; **Published** December 31, 2013

Funding: ASC is funded by the Novo Nordisk STAR PhD program. MB is funded by the Danish Council for Independent Research (FTP, 09-066546). WB and KL-L are supported by a Hallas-Møller stipend (to KL-L) from the Novo Nordisk Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have read the journal's policy and have the following conflicts: The authors declare funding from a commercial source, Novo Nordisk. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: andersx@nano.ku.dk

Introduction

Chemical shifts hold valuable structural information that is being used increasingly in the determination of protein structure and dynamics[1]. This is made possible primarily by empirical chemical shift predictors such as SHIFTS, SPARTA, SHIFTX, PROSHIFT, and CamShift [2–7]. While these methods generally offer quite accurate predictions, the predicted chemical shifts of backbone amide protons (δ_H) tend to be significantly less accurate than, for example, the proton on the α-carbon [8,9]. This is unfortunate since ^{15}N-HSQC forms a large fraction of all protein NMR studies and δ_H holds valuable information about the hydrogen bond geometry of the ubiquitous amide-amide hydrogen bonds that are key to protein secondary structure. Parker, Houk and Jensen [10] have proposed a δ_H-predictor that was shown to offer significantly more accurate predictions, although this was only demonstrated for 13 δ_H-values. The method suggests that there is an exponential dependence of δ_H in the NH··O = C bond length (as suggested by Barfield [11] and Cornilescu *et al.* [12]) as well as a non-negligible contribution from cooperative effects in hydrogen bonding networks. This exponential dependence makes empirical parameterizations of δ_H-predictors challenging since even small discrepancies between the structure used in the parameterization (usually an X-ray structure without explicitly represented hydrogens) and the solution-phase structural ensemble that gives rise to the experimentally observed δ_H-values can have a significant effect. The method by Parker *et al.* addresses this problem by parameterization against δ_H-values obtained by quantum mechanical (QM) calculations, and is similar in spirit to the QM-based α-carbon chemical shift predictor CheShift developed by Vila *et al.* [13,14]. Both studies noted that the QM-based chemical shift predictors tend to be more sensitive to small structural changes compared to popular empirical chemical shift predictors and therefore promises to be valuable tools in protein structure validation and refinement. Here we present several key advances in the use of backbone amide proton chemical shifts to refine and validate the geometry of the amide-amide hydrogen bonding network in proteins. First we present and validate the ProCS method which extends the QM-based backbone amide proton chemical shift predictor proposed by Parker *et al.* [10]. Second we present a computational methodology for using ProCS and experimental δ_H-values to refine the hydrogen bond-geometries of proteins. This is accomplished by implementing ProCS in the Markov chain Monte Carlo (MCMC) protein simulation framework PHAISTOS [15], and using this in

combination with a molecular mechanics (MM) force field. Third, we show for a number of small proteins that structural refinement against experimental δ_H values using ProCS leads to hydrogen bond geometries that are in closer agreement with high-resolution X-ray structures and experimental trans-hydrogen bond spin-spin coupling constants ($^{h3}J_{NC'}$) compared to using an energy function based on the empirical chemical shift predictor CamShift [7] or solely using a force field (OPLS-AA/L [16] with the GB/SA continuum solvent model [17]).

Results and Discussion

The ProCS method

The ProCS program uses a modified implementation of the formula developed by Parker *et al.*[10] where the amide proton chemical shift is approximated by a sum of additive terms:

$$\delta_H = \delta_{BB} + \Delta\delta_{1^{\circ}HB} + \Delta\delta_{2^{\circ}HB} + \Delta\delta_{3^{\circ}HB} + \Delta\delta_{RC} \quad (1)$$

Here, δ_{BB} is a backbone term that depends on the (ϕ,ψ) torsion angles of the residue, $\Delta\delta_{1^{\circ}HB}$ is due to a primary hydrogen bond directly to the amide proton in question, $\Delta\delta_{2^{\circ}HB}$ is due to a secondary hydrogen bond to the carbonyl oxygen in the amide group, $\Delta\delta_{3^{\circ}HB}$ is a small term that incorporates further polarization due to hydrogen bonding at the primary and/or secondary bonding partner and $rDelta\delta_{RC}$ describes magnetic perturbations due to ring currents in nearby aromatic side chains. ProCS calculates amide proton chemical shift values referenced to dimethyl-silapentane-sulfonate (DSS).

We have replaced the original δ_{BB} term, which was a crude 3-step function, by a scaled version of the (ϕ,ψ) backbone torsion angle hypersurface parametrized by Czinki and Császár [18]. The δ_{BB} term is given as

$$\delta_{BB} = 0.828\cdot(\mathrm{ICS}(\phi,\psi) + 0.77\ \mathrm{ppm}) \quad (2)$$

where $\mathrm{ICS}(\phi,\psi)$ is the n-th order cosine series given in reference [18]. The scaling is necessary to account for differences in choice of basis set and molecular geometry optimization [19].

In the cases described by Parker *et al.*, $\Delta\delta_{RC}$-values are obtained through the SHIFTS web-interface[3]. Since this would be impractical, we implemented the point-dipole [20,21] approximation given by:

$$\Delta\delta_{RC} = i\ B\ \frac{1-3\cos^2(\theta)}{|\vec{r}|^3} \quad (3)$$

where i is an intensity parameter which depends on the type of aromatic ring, B is a constant of 30.42 ppm Å^3, $\vec{r}$ is the vector between the amide proton and the center of the aromatic ring and θ is the angle between $\vec{r}$ and the normal to the plane of the aromatic ring located on its center. The values of i and B are obtained from the parameter set by Christensen *et al.* [22].

The following expression for $\Delta\delta_{1^{\circ}HB}$ was implemented for primary bonds to backbone amide carbonyl oxygen atoms:

$$\Delta\delta_{1^{\circ}HB} = [4.81\cos^2(\theta) + \sin^2(\theta)\{3.10\cos^2(\rho)$$

$$-0.84\cos(\rho) + 1.75\}]e^{-2.0\ \mathrm{A}^{-1}(r_{OH} - 212.760\ 2A)}\cdot 1\ \mathrm{ppm} \quad (4)$$

This formula originates from the works of Barfield[11] and is fitted to chemical shifts computed for model systems of hydrogen bonding between two formamide molecules. In order to treat hydrogen bonding to other oxygen atom types (carboxylic acids and alcohols as found in side chains and C-terminal), we carried out similar scans (see Section S2 and Fig. S4 in Supporting Information S1) over bond angles and lengths and stored these in lookup-tables from which the chemical shift perturbation due to any hydrogen bonding geometry can be interpolated. Hydrogen bonding to carboxylic acid oxygen atoms interaction were modeled by *N*-methylacetamide/acetate dimers, while bonds to alcohols oxygen atoms were modeled by *N*-methylacetamide/methanol dimers.

For non-hydrogen bonding amide protons, which are found primarily on the protein surface, $\Delta\delta_{1^{\circ}HB}$ is approximated as the interaction between a water molecule and an *N*-methylacetamide molecule. In this case, $\Delta\delta_{1^{\circ}HB}$ is equal to 2.07 ppm for an energy minimized bonding geometry (see Section S3 and Fig. S5 in Supporting Information S1). The functional forms of $\Delta\delta_{2^{\circ}HB}$ and $\Delta\delta_{3^{\circ}HB}$ were kept as described in reference [10].

Reproducing QM chemical shifts

ProCS predictions result from several terms [Eq. 1] that are assumed to be additive. To test this additivity assumption we use density functional theory (DFT) and compute chemical shielding values (at the B3LYP/cc-pVTZ/PCM level) for the crystal structure of human parathyroid hormone, residues 1–34 at 0.9 Å resolution, PDB-code 1ET1 [23]. Chemical shift values for amide protons at the termini are excluded from the statistics presented in this section, since they do not participate in any hydrogen bonds in the crystal structure. Using the linear scaling method due to Jain *et al.* [24] similar DFT calculations reproduce experimental proton chemical shifts of a test set of 80 small to medium sized molecules to an RMSD of 0.13 ppm. [24]

ProCS reproduces the QM calculation with an RMSD of 0.25 ppm (Table 1) based on the same structure. ProCS is parameterized based on a number of DFT calculations (see Methods section) which have been shown to yield proton chemical shifts within 0.16 ppm of experimental values for small organic molecules [19]. Thus, the error from non-additivity is roughly the same as the expected deviation from experiment.

The chemical shifts predicted by empirical methods do not agree well with the DFT results, with RMSD values ranging from 0.56 to 0.70 ppm (see Table 1 and Fig. 1). The DFT chemical shifts span a relatively large range (5.8–9.3 ppm) while the empirically predicted chemical shifts span a very narrow range (up to 6.9–8.9 ppm for SPARTA+) - see Fig. 1. This indicates that the empirical methods are less sensitive to small differences in hydrogen bond geometry found in the X-ray structure.

Reproducing experimental chemical shifts from X-ray structures

The QM method used here reproduces small molecule ^{1}H chemical shifts with an RMSD of 0.13 ppm [24]. The RMSD between the chemical shifts calculated by QM using the static X-Ray structure and the experimental data obtained in solution is 0.66 ppm. The main sources of this discrepancy are likely inaccuracies in the hydrogen bond lengths in the X-ray structure compared to solution, since there is an exponential dependence of the proton chemical shifts on this distance [Eq. 4], and/or the use of a single structure rather than a structural ensemble.

The corresponding RMSD to experimental data for ProCS (0.63 ppm) is similar to the QM RMSD and significantly larger than the 0.25 ppm RMSD between QM and ProCS, indicating

Table 1. Correlation coefficients and RMSD between five chemical shift predictors, chemical shifts derived from quantum mechanics (B3LYP/cc-pVTZ/PCM) chemical shifts and experimental values.

Data source[a]	Exp'tl	Exp'tl	QM	QM
	r	RMSD	*r*	RMSD
ProCS	0.54	0.63	0.94	0.25
SHIFTS[2]	0.64	0.37	0.59	0.70
SHIFTX[5]	0.69	0.37	0.71	0.62
SPARTA+[40]	0.69	0.42	0.68	0.56
CamShift[7]	0.64	0.32	0.59	0.66

[a]The crystal structure of human parathyroid hormone, residues 1–34 at 0.9 Å resolution (PDB-code 1ET1[23]) is used as input structure in all chemical shift calculations.
doi:10.1371/journal.pone.0084123.t001

Table 2. Reproduction of experimental amide proton chemical shift values based on 13 X-ray structures with a crystallographic resolution of 1.35 Å or less.

Method	$\langle r \rangle$[a]	$\langle$RMSD$\rangle$[b]
ProCS	0.58	1.13 ppm
SHIFTS[2]	0.56	0.64 ppm
SHIFTX[5]	0.71[c]	0.51 ppmc
SPARTA+[40]	0.79	0.40 ppm
CamShift[7]	0.74	0.46 ppm

$\langle r \rangle$">denotes the average correlation coefficient over the 13 structure.
$\langle$">RMSD$\rangle$ denotes the average root mean square deviation over the 13 structure.
[c]For SHIFTX, three structures displayed over fitting behavior with $r \approx 0.99$. These structures are excluded from the average values.
doi:10.1371/journal.pone.0084123.t002

that ProCS is sufficiently accurate to identify inaccuracies in the X-ray structure, and/or the effect of using a single structure rather than a structural ensemble. A similar comparison to experiment for 13 other proteins is given in Table 2 (PDB-codes: 1BRF, 1CEX, 1CY5, 1ET1, 1I27, 1IFC, 1IGD, 1OGW, 1PLC, 1RGE, 1RUV, 3LZT, 5PTI). The deviation from experiment for the empirical methods are significantly smaller than for ProCS with RMSD values ranging from 0.46 to 0.64 ppm (Table 2). A likely explanation for this is that the empirical methods are parameterized using X-ray structures. In order for these methods to produce low RMSD values relative to experiment they need to be insensitive to errors in protein structure.

Refining protein structures based on chemical shifts

If indeed the difference in experimental and computed chemical shifts reports on inaccuracies in the protein structure, then minimizing this difference can be used for structural refinement. To test this hypothesis we generate structural ensembles that minimizes the difference in computed and observed chemical shifts to the specified uncertainty in the chemical shift model and determine the quality of these structures by comparison to experimental structures and coupling constants (next section).

Refinement is accomplished using a Markov chain Monte Carlo (MCMC) technique described in detail in the Methods section. In short, the method involves Monte Carlo sampling of structural changes using a posterior distribution constructed using the OPLS-AA/L fore field [16] with the GB/SA implicit solvent model [17] (referred to hereafter simply as "OPLS") and amide proton chemical shifts differences from experiment computed using either CamShift or ProCS. We note that the resulting ensemble is not a dynamic ensemble but an ensemble that reflects experimentally measured amide proton chemical shifts. The simulation lengths are roughly equivalent to 6–10 ns of molecular dynamics simulations [25]. We refine the structure of ubiquitin, Protein G, and SMN Tudor domain each based on three energy functions: OPLS alone, OPLS+ProCS and OPLS+CamShift. Each MC refinement results in an ensemble of 24,000 structural samples for Ubiquitin and 40,000 for Protein G and SMN Tudor Domain, from which average chemical shifts for each amide proton are computed. The results are summarized in Table 3.

The average ProCS chemical shifts are in better agreement with experiment (RMSD 0.81 ppm) compared to using X-ray structures (RMSD 1.10 ppm). The respective RMSD values for amide protons hydrogen bonded to backbone amide groups, other hydrogen bonds, and no hydrogens bonds are 0.31 ppm, 0.78 ppm and 1.09, respectively. These RMSD values reflect the uncertainties defined for each kind of hydrogen bonding situation in the ProCS model (see Methods section) meaning that the simulations have indeed converged to a distribution of structures reflecting the experimental chemical shifts within the accuracy of the ProCS model at the given temperature. A corresponding structural ensemble generated solely from the OPLS force field increases the RMSD from experiment to 1.52 ppm, indicating more inaccurate hydrogen bond geometries (more on this in the next section).

An MC-based structural refinement based on OPLS and chemical shifts derived from CamShift has no substantial effect

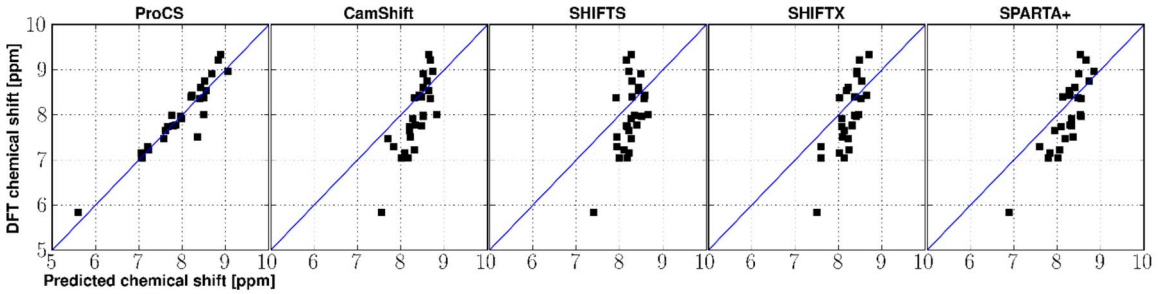

Figure 1. Correlation between chemical shift predictions from five different NMR prediction methods and quantum mechanical chemical shifts for human parathyroid hormone, residues 1–37 (PDB code: 1ET1). Blue lines represent a 1-to-1 correlation.
doi:10.1371/journal.pone.0084123.g001

Table 3. Statistics for three different types of protein simulations.

	ProCS	CamShift	⟨Bond length	
Structures[a]	^{1}H RMSD	^{1}H RMSD	deviation⟩[b]	$^{h3}J_{NC'}$ RMSD
Ubiquitin Ensembles: CamShift + OPLS	0.79 ppm	-	0.03 Å	0.17 Hz
Ubiquitin Ensembles: CamShift + OPLS	-	0.50 ppm	0.37 Å	0.17 Hz
Ubiquitin Ensembles: OPLS (no chemical shifts)	1.56 ppm	0.60 ppm	0.41 Å	0.18 Hz
1UBQ X-ray starting structure	1.22 ppm	0.51 ppm	-	0.22 Hz
SMN Tudor Domain Ensembles: ProCS + OPLS	0.93 ppm	-	0.09 Å	0.24 Hz
SMN Tudor Domain Ensembles: CamShift + OPLS	-	0.46 ppm	0.17 Å	0.23 Hz
SMN Tudor Domain Ensembles: OPLS (no chemical shifts)	1.47 ppm	0.61 ppm	0.22 Å	0.23 Hz
1MHN X-ray starting structure	1.09 ppm	0.65 ppm	-	0.24 Hz
Protein G Ensembles: ProCS + OPLS	0.69 ppm	-	0.06 Å	0.14 Hz
Protein G Ensembles: CamShift + OPLS	-	0.52 ppm	0.38 Å	0.18 Hz
Protein G Ensembles: OPLS (no chemical shifts)	1.54 ppm	0.68 ppm	0.37 Å	0.20 Hz
1PGB X-ray starting structure	1.21 ppm	0.55 ppm	-	0.17 Hz

[a]The ensembles are obtained from MCMC simulations using either OPLS-AA/L with the GB/SA solvent model (OPLS) force field energy or OPLS energy plus a chemical shift energy term from from either ProCS or CamShift. Values are calculated over four runs on each of three protein structures, Ubiquitin, Protein G and SMN Tudor Domain, or their static X-ray structure.

[b]The mean bond length deviation denotes the mean absolute difference between the mean hydrogen bond length observed in the sampled structures to the mean hydrogen bond length observed in the corresponding X-ray structure noted below.

doi:10.1371/journal.pone.0084123.t003

on the chemical shift RMSD compared to the X-ray structure (0.50 vs 0.46 ppm). Using the OPLS-derived structural ensemble increases the RMSD by 0.1 ppm compared to using X-ray structures when CamShift is used to calculate chemical shifts. This indicates that an OPLS-based refinement does not improve the hydrogen bonding geometry and that CamShift is less sensitive to a change in structure compared to ProCS.

Hydrogen bond geometries

The H··O distances and H··O = C angles of the backbone amide-amide hydrogen bonds for which $^{h3}J_{NC'}$ coupling constants have been measured (see next section) are extracted from the ensembles and compared to the corresponding values found in the experimental X-ray structures with hydrogens added from PDB2PQR [26,27]. The result are shown in Table 3 and Figures 2 and 3.

Fig. 2 shows the distributions of H··O distances from the ensembles computed using the three energy terms described in the previous section. Structural refinement using OPLS and ProCS for ubiquitin results in ensembles with average H··O distances that have an RMSD within 0.02 Å of those found in the X-ray structures 1UBQ and 1UBI (both 1.80 Å X-ray resolution) and 0.04 Å from the ubiquitin structure 1OGW (1.30 Å X-ray resolution) in which the leucine residues 50 and 67 have been replaced by fluoro leucine. For Protein G we note that the resulting ensemble does not have an average H··O distance that agrees well (0.07 Å difference) with the starting structure 1PGB (1.92 Å X-ray resolution). However the difference from the 1PGA structure (2.07 Å X-ray resolution) and the more accurate 1IGD structure (X-ray resolution of 1.1 Å) is much less, 0.02 Å and 0.00 Å, respectively. The 1IGD structure is a close homologue which has 89% sequence identity score and 95% sequence similarity. In the case of the SMN Tudor Domain, ProCS-based refinement results in slightly longer amide-amide hydrogen bond lengths (0.02 Å on average) compared to the X-ray structure 1MHN.

In contrast, structural refinement using CamShift and OPLS or just OPLS leads to increases in average H··O bond lengths of up to 0.15 Å, with a standard deviation 2–3 times larger than that found in the OPLS+ProCS simulation. In all cases use of CamShift has relatively little effect on the ensemble average H··O distance compared to just using OPLS.

In all cases, the use of ProCS leads to a significantly smaller standard deviation in H··O bond lengths: 0.017 Å compared to 0.045 and 0.041 Å for CamShift+OPLS and OPLS, respectively (Fig. 3A). The H··O = C bond angles observed in the ProCS+OPLS simulations are on average within -2.0° of corresponding value observed in the X-ray structures. The same bond angle differences are -6.7° and -7.4° observed in the CamShift+OPLS and OPLS simulations, respectively (Fig. 3B).

Trans-hydrogen bond coupling constants

Better agreement with X-ray structures does not necessarily imply better solution-phase structures. In order to compare the resulting ensembles to solution-phase data we compute average trans-hydrogen bond coupling constants and compare these to experimental values. Experimental trans-hydrogen bond $^{h3}J_{NC'}$ spin-spin coupling constants represent a very sensitive measure for solution-phase hydrogen bonding conformations and are known to correlate with amide proton chemical shifts [28]. The coupling constants depend exponentially on the hydrogen bonding distance and on bond angles [11]. Data from ensemble back-calculated $^{h3}J_{NC'}$ spin-spin coupling constants are summarized in Fig. 4 and Table 3.

In the ubiquitin simulations, the OPLS force field on its own does not yield ensemble $^{h3}J_{NC'}$ averages in good agreement with experimental data. In this simulation, several hydrogen bonds were eventually broken. Calculated $^{h3}J_{NC'}$-values for these partly unfolded hydrogen bonds show up close to 0 Hz (see Fig. 4A). The RMSD to experimental values is here 0.18 Hz. Adding the energy term from amide proton chemical shifts via CamShift does not help keeping these hydrogen bonds fixed, but results in a minor improvement in RMSD to 0.17 Hz. Adding the amide proton

chemical shifts energy term via ProCS to the OPLS force field stabilized the hydrogen bonds and also gave an improvement in the RMSD values to 0.14 Hz, which is close to that of the most accurate structural NMR ensembles of ubiquitin (see Table 4). For Protein G we obtained similar RMSD values: 0.20 Hz, 0.14 Hz and 0.18 Hz for the OPLS alone, OPLS+ProCS and the OPLS+CamShift simulations, respectively. In the SMN Tudor Domain simulation, the average $^{h3}J_{NC'}$ value of all three types of simulations were comparably close to experimental values 0.24, 0.24 and 0.23 Hz for OPLS alone, OPLS+ProCS and the OPLS+CamShift simulations, respectively. Thus, overall the coupling constants based on the ProCS refined ensembles are indeed in better agreement with experimental values indicating the refinement led to improved hydrogen bond geometries compared to using OPLS or OPLS+CamShift.

Impact on Q-factor

In this section we investigate how amide proton chemical shifts restraints affect back-calculated $^{1}D_{NH}$ residual dipolar couplings (RDCs) compared to experimental values for ubiquitin. RDCs are attractive in this regard since they report on structural features that are not related to hydrogen bonding conformations as studied intensively in the previous sections. The Q-factor is a qualitative measure for the agreement between back-calculated RDCs and the corresponding experimentally observed values [29].

We find, that for our Ubiquitin ensemble generated using the OPLS force field alone has a Q-factor of 0.29 while inclusion of chemical shifts only gives a very modest improvement of this figure to 0.27 for both CamShift and ProCS as chemical shift model. The same value calculated for the three X-ray structures 1UBQ, 1UBI and 1OGW are 0.22, 0.25 and 0.26, respectively. For six NMR-based ensembles the Q-factor is in the range 0.04–0.38, though in some cases the ensembles were refined against the RDCs (see Table 4). We observe no significant correlation ($P<0.05$) between RMSDs for predicted chemical shifts or spin-spin couplings constant to their experimental values and the calculated Q-factor for the 12 cases presented in Table 4.

While amide proton chemical shifts have some dependence on the dihedral angles of the backbone, the dependence on the particular hydrogen bonding conformations is much larger in comparison. This is due to an exponential dependence on the hydrogen bond length.

The distribution from which we sample chemical shifts is constructed from a prior distribution based on the OPLS force field and a likelihood which contains information from experimental chemical shifts. We expect that structural features of the resulting ensemble, which are not local to the hydrogen bond geometry, will largely reflect the prior distribution, i.e. in this our case, the OPLS force field.

Computational efficiency

Executing the simulations on one core of a Intel Xeon X5560 running at 2.80 GHz with the 1UBQ structure, the average evaluation time of the three different energy-terms were OPLS-AA/L: 27 ms, CamShift 1.35: 4.7 ms, ProCS: 0.74 ms. Similar evaluation times were observed for the 1MHN and 1PGB simulations. Note that, in our implementation, the CamShift term calculates chemical shifts for six atoms per residue, even if those chemical shifts are not a used to evaluate the corresponding energy term. The OPLS and CamShift terms were implemented with a caching algorithm, so only the subset of parts of the chemical shift terms that change after a local Monte Carlo move were recomputed. This approach was not implemented for ProCS since the OPLS force field energy evaluation is by far the most computationally expensive step. Running on four cores, we obtained between 10 to 16 mio Monte Carlo iteration steps total *per* day, depending on the protein size and combination of energy terms.

Methods

Monte Carlo refinement of protein structure

We employ Markov chain Monte Carlo sampling from a Bayesian posterior distribution to perform protein structure refinements and simulations. MCMC simulations are attractive because no gradient expressions need to be derived for ProCS. Bayesian inference[30] provides a rigorous mathematical framework for the inference of protein structure from experimental data. It involves the construction of a posterior distribution, which consists of a prior distribution and a likelihood. The former brings in general information on protein structure, and in our case is

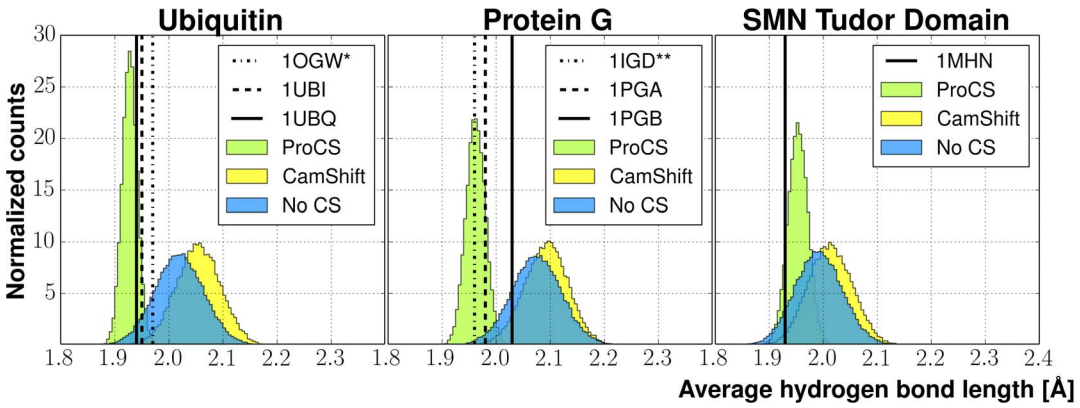

Figure 2. Distribution of average hydrogen bond lengths throughout Monte Carlo simulations on Ubiquitin, Protein G and SMN Tudor Domain. Histograms are normalized (to an area of 1) to fit identical axes. Vertical lines indicate average values obtained from experimental X-ray structures (PDB-codes are noted in the figure legends). The blue histogram represents the simulation with only the molecular mechanics energy from the OPLS-AA/L force field with the GB/SA solvent model (but no chemical shift energy term). Green and yellow histograms indicate the use of OPLS force field plus an additional chemical shift energy term from ProCS or CamShift, respectively. *1OGW contains fluoro leucine at residues 50 and 67. **1IGD is a closely related homologue (see text).
doi:10.1371/journal.pone.0084123.g002

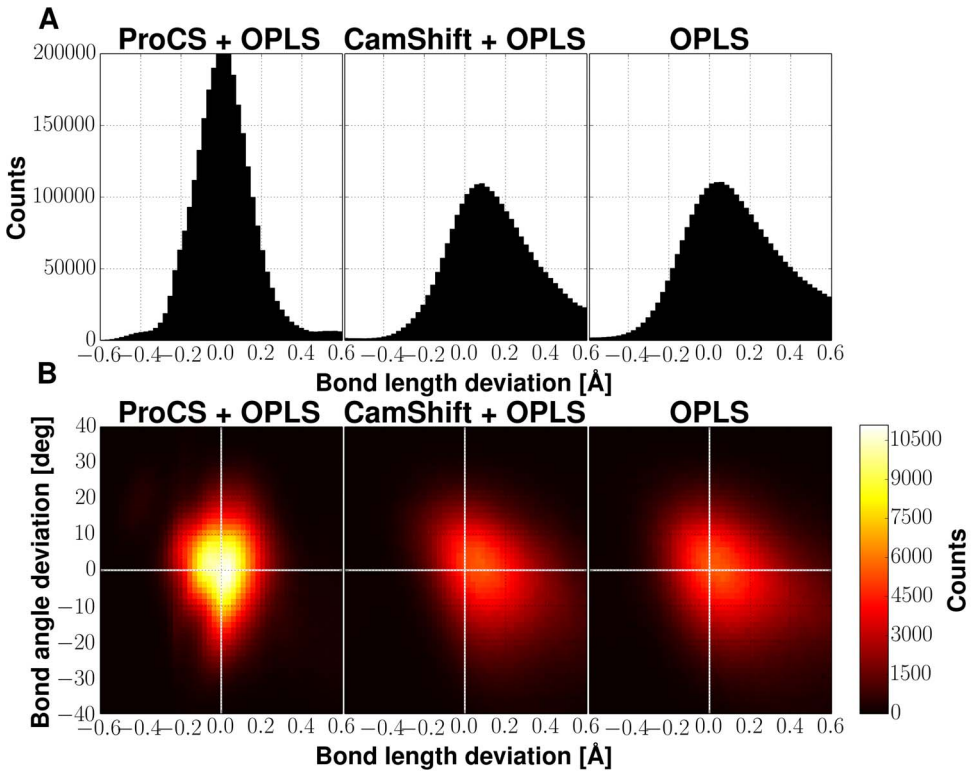

Figure 3. Deviation in hydrogen bonding geometries between the experimental X-ray structure and samples obtained from Markov Chain Monte Carlo (MCMC) simulations using the OPLS-AA/L force field with the GB/SA solvent model with either no chemical shift energy term or a chemical shift energy from either ProCS or CamShift. Data is calculated over all amide-amide bonding pairs for which experimental $^{h3}J_{NC'}$ spin-spin coupling constants were present. (A) shows the distribution of the deviations found in the MCMC ensembles from the experimental hydrogen bond length found in the X-ray structure. (B) shows the correlation of deviations in hydrogen bond lengths and H··O = C bond angles from the experimental X-ray structures.
doi:10.1371/journal.pone.0084123.g003

based on the OPLS energy function. The latter brings in the experimental data, and is based on the difference between the back-calculated data from a simulated structure and the experimental data. Using PHAISTOS, we draw samples from the joint probability distribution, which is given by:

$$p(X|\{\delta_i^{\exp}\},I) \propto p(\{\delta_i^{\exp}\}|X,I)p(X|I) \quad (5)$$

where X represents a protein structure, $\{\delta_i^{\exp}\}$ is experimental chemical shift data and I denotes prior information, such as sequence and knowledge about the uncertainties in the prediction model. The prior distribution $p(X|I)$ is proportional to $\exp(-\beta E_{FF})$, where E_{FF} is the molecular mechanics force field potential energy and $\beta = 1/k_B T$. $p(\{\delta_i^{\exp}\}|X,I)$ denotes the probability of observing experimental data given a trial structure. Under the assumption that the error in the chemical shift prediction model follows a Gaussian distribution with some set of standard deviations $\{\sigma_i\}$, the expression for $p(\{\delta_i^{\exp}\}|X,I)$ is:

$$\mathrm{p2}(\{\delta_{\mathrm{i}}^{2exp}\}\Big|\mathrm{X},\{\sigma_{\mathrm{i}}\}) = \Pi_{\mathrm{i}=1}^{\mathrm{n}}\left[\sqrt{\frac{1}{2\pi\sigma_{\mathrm{i}}^2}}\exp\left\{-\frac{(\Delta\delta_{\mathrm{i}})^2}{2\sigma_{\mathrm{i}}^2}\right\}\right] \quad (6)$$

where $\Delta\delta_i$ is the discrepancy between predicted and experimental data for the i-th nucleus of the data set in the trial structure, X. This formulation of the posterior distribution assumes that the prior distribution on X is also a good prior distribution for the chemical shift differences, $\Delta\delta_i$, otherwise an additional term would be required[31]. The set of standard deviations, $\{\sigma_i\}$ was assigned based on the primary bond type, since, for instance, the model for solvent exposed amide protons is much cruder than the amide-amide bonding model. σ_i was set to 0.3 ppm, for primary bonds to another backbone amide, 0.5 ppm to a side chain amide group, 0.8 ppm to a side chain alcohol or carboxylic acid group and 1.2 ppm for solvent exposed amide protons and other types of bond not included in the prediction model.

Protein Structures and NMR data

All protein structures used in this study were downloaded from the RCSB Protein Data Bank[32] (PDB) and protonated using PDB2PQR 1.5, [26,27] with PROPKA[33] to determine protonation states at the pH at which NMR data was recorded. Chemical shift data were obtained from the RefDB[34] or the Biological Magnetic Resonance Bank[35], and subsequently re-referenced through Shiftcor[34]. $^{h3}J_{NC'}$ spin-spin coupling constants for 1PGB, 1UBQ and 1MHN were obtained from references [28], [12] and [36], respectively.

MCMC simulations

MCMC simulations were carried out in PHAISTOS v1.0-rc1 (rev. 335) using the Metropolis-Hastings algorithm at 300 K. The simulations are initialized from the experimental crystal structures. Four independent trajectories were simulated for each protein structure. A total of 100 mio MC steps were taken for each trajectory for Protein G and the SMN Tudor Domain simulation

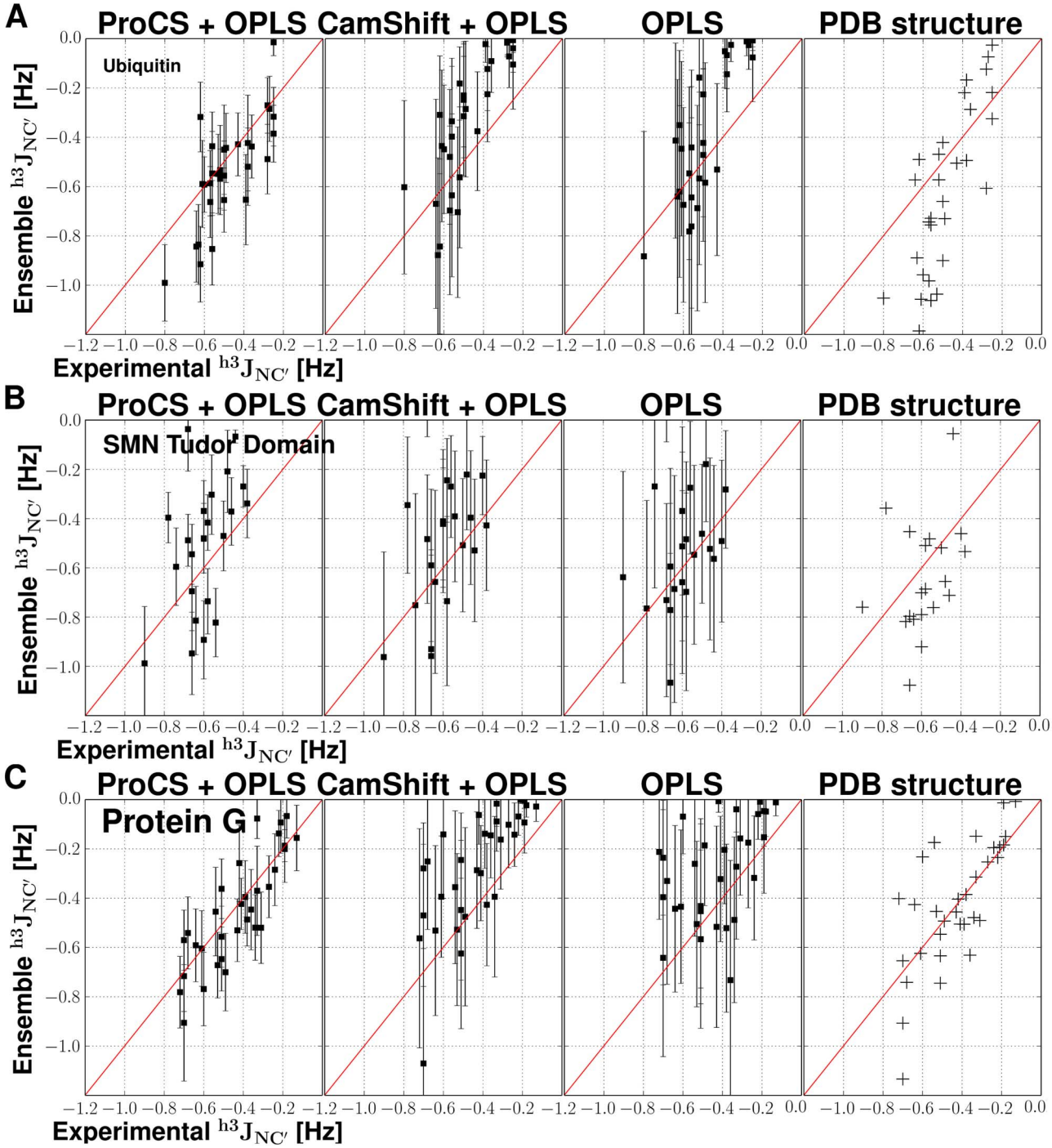

Figure 4. Reproducing experimental $^{h3}J_{NC'}$ spin-spin coupling constants via different structural ensembles and experimental X-ray structures. Squares denote the average coupling constant observed for that hydrogen bond in the ensemble and error bars represent the standard deviation observed throughout the simulations. Crosses represent the spin-spin coupling constants calculated using the static experimental X-ray structure. Results from simulations on ubiquitin is displayed in A, SMN Tudor domain in B and Protein G in C. Left column displays simulations only the OPLS-AA/L force field with the GB/SA solvent model (OPLS) and the ProCS energy term; second column is from OPLS plus the CamShift energy term; thrid column is for the simulation with only the OPLS force field energy. In the rightmost column $^{h3}J_{NC'}$ are computed from the corresponding X-ray structure.
doi:10.1371/journal.pone.0084123.g004

and 85 mio MC steps for the Ubiquitin simulation. Structures were saved every 10,000 Monte Carlo step. The Monte Carlo move-set was composed of 25% CRISP backbone moves[25] and 75% uniform side chain moves. The force field energy was calculated using the OPLS-AA/L force field [16] with the GB/SA continuum solvent model [17]. The following crystal structures obtained from the PDB were used as starting structures in the simulations: 1PGB (Protein G), 1UBQ (Ubiquitin) and 1MHN

Table 4. Statistics for selected ubiquitin ensembles and X-ray structures.[a]

	(CamShift)	(CamShift)	(ProCS)	(ProCS)	$^{h3}J_{NC'}$	
PDB-ID	^{1}H RMSD	*r*	^{1}H RMSD	*r*	RMSD	Q-factor
[b]2KOX	0.29	0.84	0.68	0.86	0.12	0.04
[c]2K39	0.34	0.82	0.98	0.77	0.13	0.07
[d]2KN5	0.23	0.91	0.71	0.82	0.12	0.22
[e]2NR2	0.44	0.74	1.35	0.64	0.14	0.25
[f]1XQQ	0.38	0.81	0.92	0.77	0.14	0.38
[g]1D3Z	0.41	0.79	1.00	0.71	0.30	0.06
[h]1UBQ	0.40	0.77	0.92	0.72	0.22	0.22
[i]1UBI	0.40	0.77	0.97	0.73	0.33	0.25
[j]1OGW	0.36	0.73	0.84	0.73	0.17	0.26
[k]OPLS + ProCS	0.32	0.79	0.17	0.98	0.14	0.27
[k]OPLS + CamShift	0.32	0.90	1.15	0.86	0.17	0.27
[k]OPLS	0.48	0.78	1.11	0.78	0.18	0.29

[a]Chemical shifts RMSD and *r* values are calculated for the residues for which $^{h3}J_{NC'}$ spin-spin coupling constants have been measured. [12]
[b]ERNST method/CHARMM27 + NOE + RDC [41]
[c]OPLS/AA-L + NOE + RDC [42]
[d]Backrub method/Rosetta all-atom energy + RDC [42]
[e]MUMO method/CHARMM22 + NOE + RDC [43]
[f]DER method/CHARMM22 + NOE + S^2 [44]
[g]NOE + RDC [45]
[h]X-ray 1.80 Å structure [46]
[i]X-ray 1.80 Å structure [47]
[j]X-ray 1.32 Å structure (synthetic protein with fluoro-LEU at residues 50 and 67) [48]
[k]The methods presented here
doi:10.1371/journal.pone.0084123.t004

(SMN Tudor Domain). Time evolution of Monte Carlo energy and chemical shift RMSDs are available in the Supplementary Information (Section S1, Figures S1–S3 of Supporting Information S1).

Back calculation of spin-spin coupling constants

$^{h3}J_{NC'}$ spin-spin coupling constants were calculated using the approximation by Barfield[11].

$$^{h3}\mathrm{J}_{\mathrm{NC}'}(\theta,\rho,\gamma_{OH}) = [-1.31\cos^2(\theta) + \{0.62\cos^2(\rho) + 0.92\cos(\rho) + 0.14\}\sin^2(\theta)]\ \mathrm{e}^{-3.2\mathrm{A}} - 1(\mathrm{r}_{2OH} - 1.760\mathrm{A}).1\ Hz \quad (7)$$

Here, the coupling depend on the $\angle$N-H$\cdot\cdot$O = C angle, ρ, $\angle$H$\cdot\cdot$O = C, θ, and the hydrogen bonding distance, r_{OH}. From the MCMC ensembles, the mean $^{h3}J_{NC'}$ spin-spin coupling constant was calculated via Eqn. 7 and the standard deviation was calculated as the root mean square deviation from the mean. The $^{h3}J_{NC'}$ RMSD to experiment is then given as

$$^{h3}\mathrm{J}_{NC'}RMSD = \sqrt{\frac{\sum_{\mathrm{i}}\left(^{h3}\mathrm{J}^{\exp,\mathrm{i}}_{NC'} - \langle ^{h3}\mathrm{J}^{calc,\mathrm{i}}_{NC'}\rangle\right)^2}{\mathrm{N}}} \quad (8)$$

where $\langle ^{h3}J^{calc,i}_{NC'}\rangle$ is the average value over the ensemble for the *i*'th coupling constant.

QM NMR calculations

All density functional theory (DFT) calculations of NMR isotropic shielding constants involved in the parametrization of ProCS were carried out in Gaussian 03[37]. Data was obtained at the GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory using the scaling technique by Rablen *et al.* [19].

The NMR calculation on the 1ET1 protein structure was carried out at the B3LYP/cc-pVTZ/PCM level of theory with a water-like dielectric constant of 78.3553. In this case shielding constants were converted to chemical shifts using the scaling factor obtained by Jain *et al.* [24], assuming that the value of the dielectric constant has a negligible contribution to the scaling factors.

Calculation of ubiquitin Residual Dipolar Couplings

Residual dipolar couplings were back-calculated from the structural ensembles using singular value decomposition to fit the alignment tensor [38]. Ensemble averaging was taken into account so that all structures simultaneously were fitted to a single alignment tensor [39]. The agreement to experimental values was calculated via the Q-factor: [29]

$$\mathrm{Q} = \frac{\sqrt{\Sigma(RDC^{\exp} - RDC^{calc})^2}}{\sqrt{\Sigma(RDC^{calc})^2}} \quad (9)$$

Conclusions

ProCS is a QM-based backbone amide proton chemical shift (δ_{H}) predictor that can deliver QM quality chemical shift predictions for a protein structure in a millisecond. δ_{H}-values predicted using X-ray structures are in worse agreement with experiment, compared to those of the popular empirical chemical shift-predictors CamShift, SHIFTS, SHIFTX, and SPARTA+.

However the agreement with experiment can be significantly improved by refining the protein structures using an energy function that includes a force field and a solvation term (OPLS-AA/L with the GB/SA continuum solvent model) and a chemical shift term in the program PHAISTOS. This refinement also results in structures with predicted trans-hydrogen bond coupling constants ($^{h3}J_{NC'}$) in good agreement with experiment indicating that the refined protein structures reflect the structures in solution. Comparison of average hydrogen bond geometries to those of high-resolution (<1.35 Å) X-ray structures reveals that the structural refinement improves the predicted δ_H-values through relatively small changes in the hydrogen bond geometry distribution.

Structural refinement without chemical shifts (i.e. using only the OPLS-AA/L + Generalized Born solvation energy) or combined with CamShift has relatively little effect on the predicted δ_H-values, while the predicted $^{h3}J_{NC'}$ values are in slightly worse agreement with experiment compared to using X-ray structures or ProCS-refined structures. This is not surprising given the fact that CamShift and similar empirical methods were designed to be insensitive to relatively small changes in protein structure in order to offer robust chemical shift predictions based on X-ray structures of varying accuracy. Structural refinement based on other empirical shift predictors, such as SHIFTS, SHIFTX, and SPARTA+, were not tested mainly because an efficient interface to PHAISTOS requires a complete re-implementation of the method. However, based on our comparison to the QM-calculations (Table 1 and Fig. 1) we do not think the conclusions will be substantially different. Our data, and that of Vila *et al.* [14], suggests that QM-derived chemical shift predictors are sufficiently accurate to extract small changes in structure and dynamics from experimentally measured protein chemical shifts.

We are currently working on implementing a QM-based chemical shift prediction method for the remaining H, C, and N nuclei in a protein in ProCS (unfortunately, the source code of the CheShift method developed by Vila *et al.* for QM-based C chemical shift prediction is not available). The resulting ProCS/PHAISTOS interface should provide a powerful tool for chemical shift-based protein structure refinement.

The ensembles resulting from the simulations can be downloaded from DOI: http://dx.doi.org/10.5879/BILS/p000001

Implementations of ProCS and CamShift can be downloaded as separate modules for PHAISTOS under the terms of the GNU General Public License v3 from: http://github.com/jensengroup/

Supporting Information

Supporting Information S1 Section S1: Time evolution of energies and chemical shift RMSDs during MCMC simulation. Figures S1–S3: Details of Monte Carlo energies and chemical shift RMSDs over time for the presented simulations. **Section S2: Parametrization of chemical shift contributions due to hydrogen bonding interactions to carboxylic acids and alcohols.** Figure S4: Sketches showing the geometric parameters and the systems used in the modeling of chemical shift contributions due to hydrogen bonding. **Section S3: Model for solvent exposed amide protons.** Table S1: Chemical shift contributions due to hydrogen bonding to water molecules. Figure S5: Local minima of NMA-water dimer.
(PDF)

Author Contributions

Conceived and designed the experiments: ASC KLL TH JHJ. Performed the experiments: ASC TEL MB WB. Analyzed the data: ASC TEL JHJ. Wrote the paper: ASC JHJ.

References

1. Mulder FAA, Filatov M (2010) Ab initio NMR chemical shift data and shielding calculations: Emerging tools for protein structure determination. Chem Soc Rev 395: 578–590.
2. Moon S, Case DA (2001) A new model for chemical shifts of amide hydrogens in proteins. J Biomol NMR 38: 139–150.
3. Xu XP, Case DA (2001) Automated prediction of 15N, $13C^{\alpha}$, $13C^{\beta}$ and 13C chemical shifts in proteins using a density functional database. J Biomol NMR 21: 321–333.
4. 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.
5. Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein 1h and 13C and 15N chemical shifts. J Biomol NMR 26: 215–240.
6. Meiler J (2003) PROSHIFT: Protein chemical shift prediction using artificial neural networks. J Biomol NMR 26: 25–37.
7. Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate pre-dictions of protein NMR chemical shifts from interatomic distances. J Am Chem Soc 131: 13894–13895.
8. Wishart D, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Methods Enzymol 338: 3–34.
9. Case DA (2013) Chemical shifts in biomolecules. Curr Opin Struct Biol 23: 172–176.
10. 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.
11. Barfield M (2002) Structural dependencies of interresidue scalar coupling ^{h3}J nc' and donor 1H chemical shifts in the hydrogen bonding regions of proteins. J Am Chem Soc 124: 4158–4168.
12. Cornilescu G, Ramirez BE, Frank MK, Clore MG, Gronenborn AM, et al. (1999) Correlation between ^{h3}J nc' and hydrogen bond length in proteins. J Am Chem Soc 121: 6275–6279.
13. Vila JA, Scheraga HA (2009) Assessing the accuracy of protein structures by quantum mechanical computations of 13C(alpha) chemical shifts. Acc Chem Res 42: 1545–1553.
14. Vila JA, Arnautova YA, Martin OA, Scheraga HA (2009) Quantum-mechanics-derived 13Ca chemical shift server (cheshift) for protein structure validation. Proc Natl Acad Sci 106: 16972–16977.
15. Boomsma W, Frellsen J, Harder T, Bottaro S, Johansson KE, et al. (2013) PHAISTOS: a framework for markov chain monte carlo simulation and inference of protein structure. J of Comp Chem 00: 000-000, DOI: 10.1002/jcc.23292.
16. Kaminski GA, Friesner RA (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105: 6474–6487.
17. Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation: A fast analytical method for the calculation of approximate born radii. J Phys Chem A 101: 3005–3014.
18. Czinki E, Császár AG (2004) On NMR isotropic chemical shift surfaces of peptide models. J Mol Struct (THEOCHEM) 675: 107–116.
19. Rablen PR, Pearlman SA, Finkbiner J (1999) A comparison of density functional methods for the estimation of proton chemical shifts with chemical accuracy. J Phys Chem A 103: 7357–7363.
20. Pople JA (1956) Proton magnetic resonance of hydrocarbons. J Chem Phys 24: 1111.
21. Pople JA (1958) Molecular orbital theory of aromatic ring currents. Mol Phys 1: 175–180.
22. Christensen AS, Sauer SPA, Jensen JH (2011) Definitive benchmark study of ring current effects on amide proton chemical shifts. J Chem Theory Comput 7: 2078–2084.
23. Jin L, Briggs SL, Chandrasekhar S, Chirgadze NY, Clawson DK, et al. (2000) Crystal structure of human parathyroid hormone 1-34 at 0.9 å resolution. J Biol Chem 275: 27238–27244.
24. Jain R, Bally T, Rablen PR (2009) Calculating accurate proton chemical shifts of organic molecules with density functional methods and modest basis sets. J Org Chem 74: 4017–4023.
25. Bottaro S, Boomsma W, Johansson KE, Andreetta C, Hamleryck TW, et al. (2011) Subtle monte carlo updates in dense molecular systems. J Chem Theory Comput 8: 695–702.
26. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup, execution, and analysis of poisson-boltzmann electrostatics calculations. Nucl Acids Res 32: W665–W667.
27. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, et al. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucl Acids Res 35: W522–W525.

28. Cordier F, Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue 3hJNC' scalar couplings. J Am Chem Soc 121: 1601–1602.
29. Bax A (2003) Weak alignment offiers new nmr opportunities to study protein structure and dynamcs. Prot Sci 12: 1–16.
30. Rieping W, Habeck M, Nilges M (2005) Inferential structure determination. Science 308: 303–306.
31. Hamelryck T, Borg M, Paluszewski M, Paulsen J, Frellsen J, et al. (2010) Potentials of mean force for protein structure prediction vindicated, formalized and generalized. PLoS ONE 5: e13714.
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The protein data bank. Nucl Acids Res 28: 235–242.
33. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins 61: 704–721.
34. Zhang H, Neal S, Wishart D (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25: 173–195.
35. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, et al. (2008) Biomagresbank. Nucl Acids Res 36: 402–408.
36. Markwick PRL, Sprangers R, Sattler M (2003) Dynamic effects on j-couplings across hydrogen bonds in proteins. J Am Chem Soc 125: 644–645.
37. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. (2004) Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT.
38. Losonczi JA, Andrec M, Fischer MW, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson 138: 334342.
39. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433: 128–132.
40. Shen Y, Bax A (2010) SPARTA+: a modest improvement in empirical NMR chemical shift pre- diction by means of an artificial neural network. J Biomol NMR 48: 13–22.
41. Fenwick RB, Esteban-Martn S, Richter B, Lee D, Walter KFA, et al. (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. J Am Chem Soc 133: 10336–10339.
42. Lange OF, Lakomek NA, Fars C, Schrder GF, Walter KFA, et al. (2008) Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. Science 320: 1471–1475.
43. Richter B, Gsponer J, Várnai P, Salvatella X, Vendruscolo M (2007) The mumo (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. J Biomol NMR 37: 117–135.
44. Lindorff-Larsen K, Best R, DePristo M, Dobson C, Vendruscolo M (2004) Simultaneous determination of protein structure and dynamics. Nature 433: 128–132.
45. Cornilescu G, Marquardt J, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120: 6836–6837.
46. Vijay-Kumar S, Bugg C, Cook W (1987) Structure of ubiquitin refined at 1.8 a resolution. J Mol Biol 194: 531–544.
47. Ramage R, Green J, Muir T, Ogunjobi O, Love S, et al. (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. Biochem J 299: 151–158.
48. Alexeev D, Barlow PN, Bury SM, Charrier JD, Cooper A, et al. (2003) Synthesis, structural and biological studies of ubiquitin mutants containing (2s, 4s)-5-uoroleucine residues strategically placed in the hydrophobic core. ChemBioChem 4: 894–896.