J Biomol NMR (2015) 63:299–307
DOI 10.1007/s10858-015-9990-z

ARTICLE

Improved validation of IDP ensembles by one-bond Cα–Hα scalar couplings

Vytautas Gapsys[1] · Raghavendran L. Narayanan[2] · ShengQi Xiang[2] · Bert L. de Groot[1] · Markus Zweckstetter[2,3,4]

Received: 26 June 2015 / Accepted: 28 September 2015 / Published online: 3 October 2015

Abstract Intrinsically disordered proteins (IDPs) are best described by ensembles of conformations and a variety of approaches have been developed to determine IDP ensembles. Because of the large number of conformations, however, cross-validation of the determined ensembles by independent experimental data is crucial. The $^1J_{C\alpha H\alpha}$ coupling constant is particularly suited for cross-validation, because it has a large magnitude and mostly depends on the often less accessible dihedral angle ψ. Here, we reinvestigated the connection between $^1J_{C\alpha H\alpha}$ values and protein backbone dihedral angles. We show that accurate amino-acid specific random coil values of the $^1J_{C\alpha H\alpha}$ coupling constant, in combination with a reparameterized empirical Karplus-type equation, allow for reliable cross-validation of molecular ensembles of IDPs.

Keywords NMR · Intrinsically disordered protein · Scalar coupling · Ensemble

Introduction

Intrinsically disordered proteins (IDPs) play an important role in a wide range of biological and pathological processes in eukaryotic organisms (Iakoucheva et al. 2002; Tompa 2002; Uversky 2002; Wright and Dyson 1999). Because of their involvement in devastating diseases such as cancer, cardiovascular disease and Alzheimer's disease, IDPs are increasingly recognized as potential drug targets (Uversky 2011; Uversky et al. 2008). Targeting IDPs by small molecules is, however, difficult because IDPs do not fold into a single, rigid structure in solution, but exchange between a large number of conformations. IDPs are therefore best described by ensemble of conformations (Fisher and Stultz 2011; Jensen et al. 2014; Marsh et al. 2012; Mittag et al. 2010).

Several different approaches have been proposed to determine molecular ensembles of IDPs. These include pure computational approaches, such as long time-scale MD simulations on dedicated computers (Lindorff-Larsen et al. 2012). Comparison with experimental data, however, often shows that the choice of force field can strongly influence the resulting ensemble (Piana et al. 2014). Therefore, sample-and-select approaches were developed where sub-ensembles, which are best in agreement with experimental data, are derived from a broader distribution (Allison et al. 2009; Choy and Forman-Kay 2001; Fisher and Stultz 2011; Jensen et al. 2014; Marsh et al. 2012; Mittag et al. 2010; Xiang et al. 2013). The data used for selection can include small angle X-ray scattering curves (Bernado et al. 2007) and potentially other biophysical parameters. The most powerful data, however, come from NMR spectroscopy, because chemical shifts, scalar couplings, residual dipolar couplings and NOE contacts are highly sensitive probes for the conformations sampled by

Vytautas Gapsys and Raghavendran L. Narayanan have contributed equally to this work.

✉ Markus Zweckstetter
markus.zweckstetter@dzne.de

[1] Computational Biomolecular Dynamics Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

[2] Department for NMR-Based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

[3] German Center for Neurodegenerative Diseases (DZNE), 37077 Göttingen, Germany

[4] Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medical Center Göttingen, 37075 Göttingen, Germany

individual residues (Marsh et al. 2012; Rezaei-Ghaleh et al. 2012; Schwalbe et al. 2014; Uversky 2002).

Because of the large number of conformations sampled by IDPs in solution, molecular ensembles of IDPs are inherently underdetermined by the available experimental data (Allison et al. 2009; Fisher and Stultz 2011; Jensen et al. 2014; Mantsyzov et al. 2014; Marsh et al. 2012; Mittag and Forman-Kay 2007; Xiang et al. 2013). In order to achieve the most representative ensembles, it is therefore important to use the maximum available experimental data for ensemble selection. At the same time, some experimental data should be left out during the selection process and instead should be used for cross-validation—that is for comparison of the experimental values with those back-calculated from the ensemble, which was selected without the use of these data (Jensen et al. 2014; Schwalbe et al. 2014). Often chemical shifts and residual dipolar couplings (Allison et al. 2009; Fisher and Stultz 2011; Jensen et al. 2014; Marsh et al. 2012; Mittag and Forman-Kay 2007; Xiang et al. 2013), and in some cases NOEs (Ball et al. 2011; Fawzi et al. 2008; Mantsyzov et al. 2014; Marsh and Forman-Kay 2012; Schwalbe et al. 2015), are used for ensemble selection. In this case, ${}^{3}J_{HnH\alpha}$ and ${}^{1}J_{C\alpha H\alpha}$ scalar couplings might then be used for cross-validation. Because ${}^{3}J_{HnH\alpha}$ and ${}^{1}J_{C\alpha H\alpha}$ scalar couplings depend in distinct ways on the conformation of the protein backbone—the ${}^{3}J_{HnH\alpha}$ coupling is mostly influenced by the dihedral angle φ while ${}^{1}J_{C\alpha H\alpha}$ most strongly depends on ψ (Billeter et al. 1992; Edison et al. 1994a, b; Kopple et al. 1978; Vuister and Bax 1993; Vuister et al. 1993)—it is best to use them both for cross-validation.

In order to be able to use ${}^{1}J_{C\alpha H\alpha}$ for the determination and analysis of ensembles of IDPs, a quantitative correlation between the protein backbone conformation and the value of ${}^{1}J_{C\alpha H\alpha}$ is required. Such correlations were investigated by molecular orbital calculations and experimental data on cyclic peptides (Egli and Vonphilipsborn 1981), as well as by ab initio calculations in an alanine derivative (Edison et al. 1994a). In addition, an empirical correlation between ${}^{1}J_{C\alpha H\alpha}$ couplings and protein backbone angles was derived from a set of proteins, for which both experimental ${}^{1}J_{C\alpha H\alpha}$ couplings and high-resolution X-ray structures were available (Vuister and Bax 1993; Vuister et al. 1993). This and other analyses showed that ${}^{1}J_{C\alpha H\alpha}$ couplings vary in proteins from ~132 to ~150 Hz with residues in extended conformations, such as a β-sheet, having average ${}^{1}J_{C\alpha H\alpha}$ values around 140.5 ± 1.8 Hz. For residues in α-helices values of 146.5 ± 1.8 Hz were found, while non-glycine residues with positive φ angles have ${}^{1}J_{C\alpha H\alpha}$ values below 137 Hz (Schmidt et al. 2009; Vuister and Bax 1993).

When we previously analyzed molecular ensembles of IDPs, which were selected on the basis of a large number of ${}^{3}J_{HnH\alpha}$ scalar couplings and residual dipolar couplings (Xiang et al. 2013), we found larger than expected deviations between experimental and back-calculated ${}^{1}J_{C\alpha H\alpha}$ values. To obtain insight into this problem and provide a validation for the sampling of ψ-angles in IDPs, we here reinvestigated the connection between ${}^{1}J_{C\alpha H\alpha}$ values and protein backbone dihedral angles. We determined an improved set of ${}^{1}J_{C\alpha H\alpha}$ random coil values and showed that on the basis of these values an empirical relation can be obtained, which allows robust ${}^{1}J_{C\alpha H\alpha}$-based cross-validation of IDP ensembles.

Materials and methods

NMR spectroscopy

NMR samples contained 1 mM of ${}^{13}C/{}^{15}N$-labeled Tau (the longest isoform with 441 residues) in 50 mM phosphate buffer, pH 6.0. NMR spectra were recorded at 278 K on a Bruker Avance 900 MHz spectrometer equipped with a cryogenic probe. For determination of ${}^{1}J_{C\alpha H\alpha}$ couplings in Tau, a 3D (HA)CANH experiment was recorded, in which no decoupling was applied during the ${}^{13}C_{\alpha}$ evolution time. The ${}^{1}J_{C\alpha H\alpha}$ coupling thus remained active during a 28 ms constant time evolution period, resulting in two highly resolved ${}^{13}C$ doublet components, which are split by the ${}^{1}J_{C\alpha H\alpha}$ coupling (Zweckstetter and Bax 2001). Taking into account the signal-to-noise ratio and the duration of the constant time evolution period (Kontaxis et al. 2000), the experimental error in the ${}^{1}J_{C\alpha H\alpha}$ couplings was estimated to be below 0.5 Hz. To avoid increased crowding relative to a regular (HA)CANH spectrum, the two doublet components were separated into two separate spectra by calculating the sum and difference of an in-phase and an anti-phase (HA)CANH spectrum (Zweckstetter and Bax 2001).

Experimental ${}^{1}J_{C\alpha H\alpha}$ couplings in a peptide comprising residues 201–219 of the splicing factor SRSF1 were previously measured using the same (HA)CANH experiment (Xiang et al. 2013; Zweckstetter and Bax 2001).

Determination of conformational ensembles

Conformations of residues 201–219 of the splicing factor SRSF1 in the non-phosphorylated and phosphorylated state were determined previously (Xiang et al. 2013).

Results

Determination of ${}^{1}J_{C\alpha H\alpha}$ random coil values

${}^{1}J_{C\alpha H\alpha}$ coupling constants are influenced by a variety of factors including solvent and related electric field effects (Barfield and Johnston 1973), substituent effects (Hansen

1981), lone electron pairs, and dihedral angle orientation (Egli and Vonphilipsborn 1981). Therefore, not only the backbone conformation but also the amino acid type plays an important role (Vuister et al. 1993). To determine a quantitative correlation between $^1J_{C\alpha H\alpha}$ couplings and protein backbone angles, it is therefore important to correct the experimental $^1J_{C\alpha H\alpha}$ values for the inherent variation, which is caused by differences in the amino acid type. To this end, $^1J_{C\alpha H\alpha}$ values observed for residues in short peptides or in flexible loops of globular proteins might be used. Here we measured $^1J_{C\alpha H\alpha}$ values in the 441-residue protein Tau. Tau is an IDP (Cleveland et al. 1977) and a variety of NMR investigations have shown that Tau contains only transient secondary structure (Mukrasch et al. 2009). A total of 295 $^1J_{C\alpha H\alpha}$ values were obtained for non-overlapping residues. $^1J_{C\alpha H\alpha}$ values, which belong to the same amino acid type, were grouped and both mean and median $^1J_{C\alpha H\alpha}$ values were calculated (Fig. 1). No values were obtained for tryptophan, which is not present in the primary sequence of Tau. In addition, coupling constants from glycine residues could not be analyzed due to severe signal overlap.

We took care in order to use only well-separated cross-peaks. Nevertheless some outliers, i.e. where the $^1J_{C\alpha H\alpha}$ value of a residue significantly deviates from the values observed for other residues of the same amino acid type, were present (Fig. 1). We currently do not know the reason for these more unusual values, but it is important to remember that even using 3D experiments partial signal overlap cannot be excluded in an IDP with 441 residues. Notably, the contribution of these values to an "average", amino acid-specific random coil value is small, as a comparison of mean and median values shows (Table 1). The largest difference between the mean and median value was observed for phenylalanine, where only three experimental data points were available. The second largest difference between the mean and median value was 1.8 Hz and was observed for proline. The analysis further showed that threonine has the smallest median $^1J_{C\alpha H\alpha}$ of 142.3 Hz, while most other amino acids have $^1J_{C\alpha H\alpha}$ values of approximately 143.5 Hz. For cysteine, the median value was 144.8 Hz. Proline has the largest value (147.7 Hz), which might be due to the influence of the C_α substituent on $^1J_{C\alpha H\alpha}$ (Schmidt et al. 2009; Vuister et al. 1993).

Table 1 also lists the random coil values, which were previously determined from $^1J_{C\alpha H\alpha}$ measurements in angiotensin II, a peptide comprising the 'central helix' of calmodulin and from the unstructured tails of staphylococcal nuclease (SNase) (Vuister et al. 1993). These values can be compared with the median values of the current work for 18 amino acids, because Tau does not contain tryptophan and no values were reported for cysteine in Vuister et al. (1993). For alanine, histidine aspartate and proline the previous random coil values and the current mean values were similar, with a maximum difference of 0.7 Hz for proline (148.4 Hz based on nine data points reported in Vuister et al. (1993); 147.7 Hz based on 25 data points reported in the current work). However, in case of arginine, asparagine, glutamine, glutamate, leucine and isoleucine the previous 'random coil' values were smaller by approximately 2 Hz.

To provide further support for the use of the median values reported in Table 1 as new 'random coil values', we analyzed the $^1J_{C\alpha H\alpha}$ couplings in a short peptide, which comprises residues 225–246 of Tau. An extensive set of chemical shifts, residual dipolar couplings and NOEs showed that Tau(225–246) is highly dynamic and has at best very little secondary structure (Schwalbe et al. 2015). Compared to full-length Tau, Tau(225–246) has the advantage that signal overlap is strongly reduced. $^1J_{C\alpha H\alpha}$ couplings were obtained using a J-modulated constant-time HSQC (Tjandra and Bax 1997) and not the 3D (HA)CANH, to potentially assess the influence of systematic errors in the

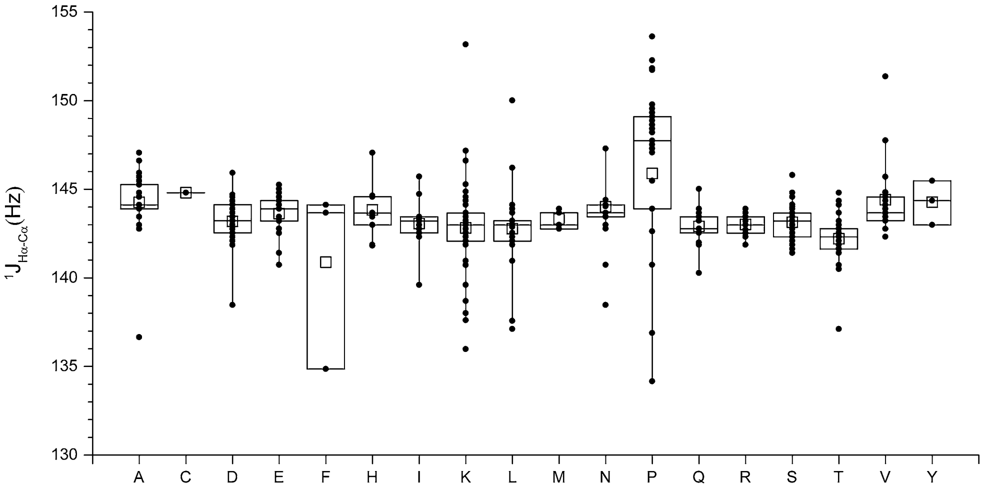

Fig. 1 $^1J_{C\alpha H\alpha}$ spin–spin coupling constants observed in the intrinsically disordered protein Tau. $^1J_{C\alpha H\alpha}$ were grouped according to amino acid type and subjected to a *box plot* analysis. The *line* inside the *box* indicates the median value for each residue type, while *bottom* and *top* correspond to 25th and 75th percentile, respectively. The mean values are marked as *square*. *Vertical lines* indicate the 5–95 % range

Table 1 Amino-acid specific random coil values of the $^{1}J_{C\alpha H\alpha}$ spin–spin coupling constant

	Number of $^{1}J_{C\alpha H\alpha}$ values in Tau(1–441)	Mean in Tau(1–441)	Median in Tau(1–441)	Number of $^{1}J_{C\alpha H\alpha}$ values in Tau(225–246)	Mean in Tau(225–246)	Random coil values from Vuister et al. (1993)
A	24	144.3	144.1	2	145.0	143.7
R	10	143.0	143.1	2	143.4	141.5
N	10	144.0	143.7			141.5
D	19	143.2	143.2			142.5
C	2	144.8	144.8			
Q	14	142.9	142.8	1	143.9	141.1
E	18	143.6	143.9			141.9
G						
H	10	143.8	143.7			143.8
I	13	143.1	143.2			141.3
L	23	142.8	143.0	1	143.3	141.1
K	38	142.8	143.1	3	143.6	141.5
M	4	143.3	143.3			142.2
F	3	140.9	143.7			142.9
P	25	145.9	147.7			148.4
S	28	143.1	143.2	4	143.2	142.1
T	27	142.2	142.3	2	142.7	141.4
W						143.0
Y	3	144.3	144.4			143.0
V	24	144.4	143.7	3	143.6	141.3

measurement. Although the number of values available for each amino acid type is small, the average values in Tau(225–246) were close to the median values obtained with full-length Tau (Table 1). The spread around the average values in Tau(225–246) was small. Notably, for most amino acids the mean value was around ~143.5 Hz, in agreement with the results from full-length Tau. This includes arginine, lysine, valine, glutamine and leucine, indicating that the new 'random coil' values are more representative.

Reparameterization of the Karplus-type equation for $^{1}J_{C\alpha H\alpha}$ couplings

To select $^{1}J_{C\alpha H\alpha}$ scalar couplings to be used in parameterization of an empirical Karplus-type relation we considered three independently measured data sets: (a) calmodulin, bovine pancreatic trypsin inhibitor (BPTI) and SNase by Vuister et al. (1993); (b) calmodulin and ubiquitin by Zweckstetter and Bax (2001); (c) flavodoxin, RNase T1, frataxin, ubiquitin, xylanase and DFPase by Schmidt et al. (2009). As the sets contained measurements of the same proteins (calmodulin, ubiquitin), we were able to assess internal consistency between the experimental data. Comparison of the couplings for ubiquitin from Zweckstetter and Bax (2001) with the values from Schmidt et al. (2009) revealed almost perfect correlation with an RMSD below 0.5 Hz (Fig. 2a). In contrast, agreement for the calmodulin couplings from Zweckstetter and Bax (2001) and Vuister et al. (1993) was worse with a correlation coefficient of 0.69 and a RMSD of 2.88 Hz (Fig. 2b). The analysis of internal consistency between the measured scalar couplings prompted us to use a merged data set from Zweckstetter and Bax (2001) and Schmidt et al. (2009) for the further parameterization. To avoid redundancy only one set of values for ubiquitin from Zweckstetter and Bax (2001) was included. In total, the final data set contained 931 values from 7 proteins.

The $^{1}J_{C\alpha H\alpha}$ couplings observed in Tau were subsequently used for parameterization. To this end, we used the median values reported in Table 1, because they are less affected by outliers. In case of phenylalanine, where the difference between the mean and median value was large, we also used the median value, which is close to the random coil value previously determined by Vuister et al. (1993). These amino-acid specific 'random coil values' were then subtracted from the 931 $^{1}J_{C\alpha H\alpha}$ scalar couplings. Backbone dihedral angles φ and ψ were extracted from the X-ray structures of ubiquitin (PDB id: 1UBQ; Vijaykumar et al. 1987) and calmodulin (PDB id: 1CLL; Chattopadhyaya et al. 1992). For flavodoxin, RNase T1, frataxin, xylanase and DFPase the dihedral angles collected from X-ray structures by Schmidt et al. (2009) were used.

A singular value decomposition (SVD) fit was performed to parameterize the equation

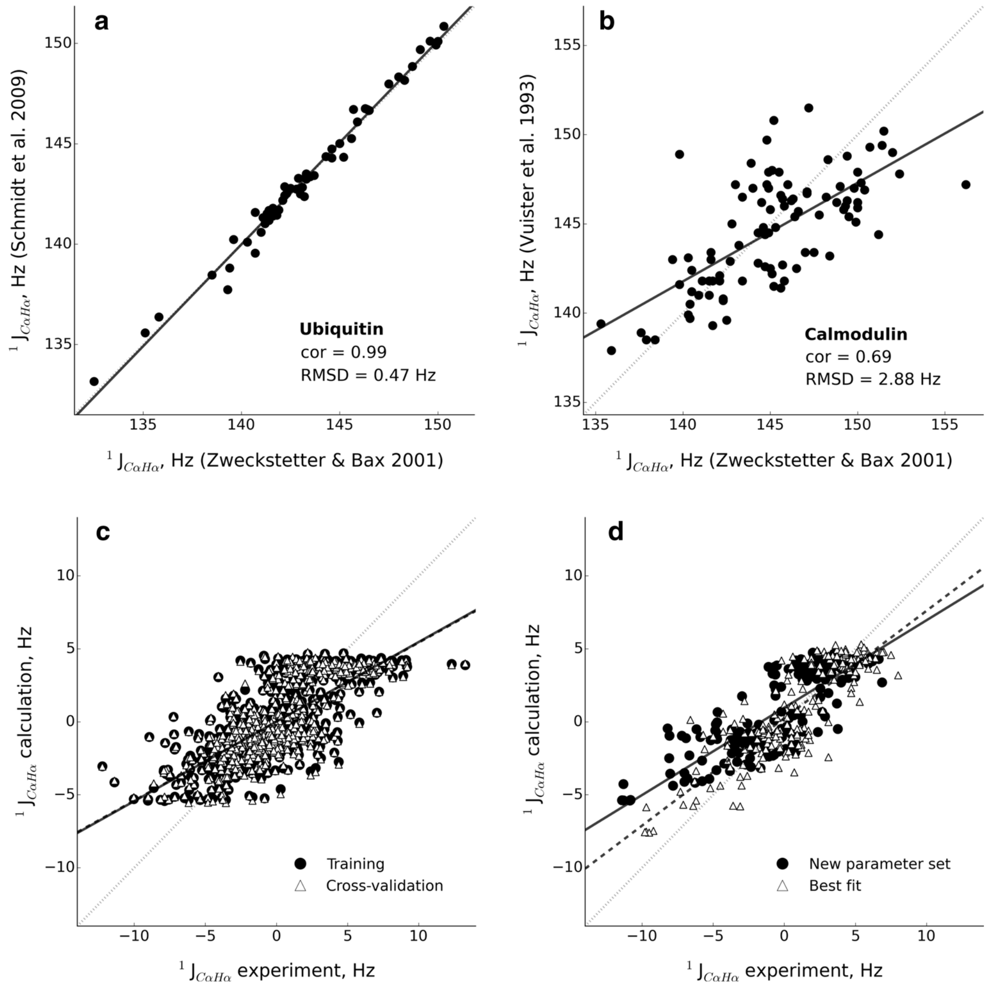

Fig. 2 **a** Experimentally measured $^1J_{C\alpha H\alpha}$ scalar couplings of ubiquitin from Zweckstetter and Bax (2001) plotted against the values from Schmidt et al. (2009). **b** Experimentally measured $^1J_{C\alpha H\alpha}$ scalar couplings of calmodulin from Zweckstetter and Bax (2001) plotted against the values from Vuister et al. (1993). **c** Experimentally measured $^1J_{C\alpha H\alpha}$ scalar coupling constants plotted against the values calculated using the new Karplus-type equation parameters and random coil values. *Circles* denote the training set, *triangles* correspond to the cross-validation set. The measured scalar couplings comprise the values for calmodulin and ubiquitin (Zweckstetter and Bax 2001). **d** Experimentally measured $^1J_{C\alpha H\alpha}$ scalar couplings in calmodulin, staphylococcal nuclease and BPTI (Vuister et al. 1993) plotted against the calculated coupling values. Calculation using the new set of parameters is marked in *circles*. *Triangles* denote the best SVD fit of the Karplus-type equation to the given data. In both figures (**c**, **d**) the random coil values as derived from the Tau protein (Table 1) were subtracted from the measured and calculated coupling values. In figure (**d**) for the best fit data the random coil values from Vuister et al. (1993) were subtracted from the measured and calculated coupling values

$$\Delta^1J_{C\alpha H\alpha} = A + B\sin(\psi + \psi') + C\cos(2(\psi + \psi')) \\ + D\cos(2(\varphi + \varphi'))$$

by finding a set of parameters A, B, C and D which minimize the root-mean-square-deviation (RMSD) to the experimentally derived $\Delta^1J_{C\alpha H\alpha}$ values. The SVD fit was combined with an exhaustive scan of the ψ′ and φ′ angles allowing for a non-linear optimization of the whole parameter set. To properly assess the predictive power of the model ten-fold cross-validation was carried out: the data set was divided into ten parts of randomly selected points without repetition. Ten independent fitting procedures (combining exhaustive angle scan with SVD) were performed, by leaving 10 % of the data for cross-validation and using the rest of the data for model building. After all the data was once used for cross-validation, all the predicted cross-validation values were concatenated and the RMSD, as well as correlation coefficient (*cor*) between the measured and predicted couplings were estimated (Fig. 2c). A summary of the different parameterization variants is provided in Table 2. From the steady cross-validation results (*cor* = 0.73, RMSD = 2.31 Hz) it appears that optimizing the φ' dihedral angle has no significant impact on the prediction accuracy. In addition, optimization of ψ' has only minor effect for the training part of the model. Taken all

Table 2 Summary of the parameter optimization procedures for the Karplus-type equation

Data used for parameterization	Optimized parameters	*cor* training	RMSD training, Hz	*cor* crossval.	RMSD crossval., Hz	*cor* (Vuister et al. 1993)	RMSD (Vuister et al. 1993), Hz
Couplings from Zweckstetter and Bax (2001) and Schmidt et al. (2009); random coil values from Tau protein	A, B, C, D, φ′ = 30°, ψ’ = 150°	0.73	2.30	0.73	2.31	0.80	2.60
	A, B, C, D, φ′ = 30°, ψ′ = 138°	0.74	2.29	0.73	2.31	0.84	2.41
	A, B, C, D, φ′ = 30°, ψ′	0.74	2.28	0.73	2.31	0.83	2.46
	A, B, C, D, φ′, ψ′	0.74	2.28	0.73	2.31	0.82	2.48
Couplings and random coil values from Vuister et al. (1993)	A, B, C, D, φ’ = 30°, ψ′ = 138°			0.70	2.75	0.86	1.95

together, we have identified the parameter set (A = −1.06, B = 1.61, C = −2.94, D = 1.32, ψ’ = 142° and φ’ = 30°) to best predict $^1J_{C\alpha H\alpha}$ scalar coupling values. The estimated parameters are close to those found previously (A = 1.7, B = 1.4, C = -4.1, D = 1.7, ψ’ = 138° and φ’ = 30°; Vuister et al. 1993) with an exception of the parameter A, for which the value provided in Vuister et al. (1993) may be lacking a minus sign.

In addition, we performed a reparameterization of the equation using $^1J_{C\alpha H\alpha}$ couplings and dihedral backbone angles for BPTI, SNase and calmodulin as reported in the supporting information of Vuister et al. (1993). In total, the data set of the three proteins comprised 203 values. These 203 $^1J_{C\alpha H\alpha}$ couplings together with the corresponding backbone dihedral angles and the random coil values reported by Vuister et al. (1993) were used to find the best fit in terms of the A, B, C and D parameters, while keeping ψ’ and φ’ fixed at 138° and 30°, respectively (Fig. 2d). The resulting RMSD of 1.95 Hz is lower than the RMSD of 2.46 Hz for the new parameter set described above (Table 2). However, the parameters optimized against the data from Vuister et al. (1993) perform worse with respect to the 931 couplings from Zweckstetter and Bax (2001) and Schmidt et al. (2009), resulting in a RMSD of 2.75 Hz. This indicates that the parameter sets obtained using the (Vuister et al. 1993) data lack predictive power for the rest of the proteins. Based on these findings we conclude that the parameter set A = −1.06, B = 1.61, C = −2.94, D = 1.32, ψ’ = 142° and φ’ = 30°, in combination with the new set of random coil values (Table 1) is suitable for the accurate prediction of $^1J_{C\alpha H\alpha}$ couplings.

Cross-validation of molecular ensembles of the RS-rich region of SRSF1

Next we tested the power of the new random coil values and the new parameter set for cross-validation of molecular ensembles of IDPs. To this end we selected the serine/arginine-rich region (residues 201–219) of the arginine/serine (RS)-rich splicing factor 1 (SRSF1) (also known as ASF/SF2). Serine/arginine-rich proteins are important players in RNA metabolism and are extensively phosphorylated at serine residues in RS repeats (Lin and Fu 2007). Residues 201–219 of SRSF1, further on called SRSF1(RS1), contain a sequence of eight RS dipeptides (Fig. 3a). NMR spectroscopy showed that SRSF1(RS1) is intrinsically disordered and phosphorylation of the RS repeats leads to a conformational switch (Xiang et al. 2013). We then determined representative ensembles of SRSF1(RS1) in the non-phosphorylated and phosphorylated state by selecting sub-ensembles from trajectories of unbiased MD simulations using a Monte-Carlo search in combination with exhaustive scanning. Ensembles containing 30 conformers were selected, which were best in agreement with experimental N–H, Cα–Hα, Cα–CO residual dipolar couplings, as well as $^3J_{HN-H\alpha}$ scalar couplings (Xiang et al. 2013).

Experimental $^1J_{C\alpha H\alpha}$ couplings were compared to values back-calculated from the ensembles of SRSF1(RS1) in the non-phosphorylated and phosphorylated state using the new parameter set for the Karplus-type equation, as well as the previous parameter set (Vuister et al. 1993) and the parameter set obtained by reparameterization using the random coil values and the $^1J_{C\alpha H\alpha}$ couplings reported in Vuister et al. (1993). For all three data sets, a good correlation between experimental and back-calculated $^1J_{C\alpha H\alpha}$ values for residues in non-phosphorylated SRSF1(RS1) was observed (Fig. 3b, c). The newly determined parameter set, however, had an RMSD = 0.84 Hz, while the other two parameterizations resulted in RMSD values of 1.8 and 1.9 Hz. In case of phosphorylated SRSF1(RS1), the correlation between experimental and back-calculated values was lower, but again comparable between the different sets of parameters (Fig. 3d, e). In agreement with the results for non-phosphorylated SRSF1(RS1), the lowest RMSD was found for the new parameterization (A = −1.06, B = 1.61, C = −2.94, D = 1.32, ψ’ = 142° and φ’ = 30°). Although, in the analysis the six phosphorylated serine residues (S207, S209,

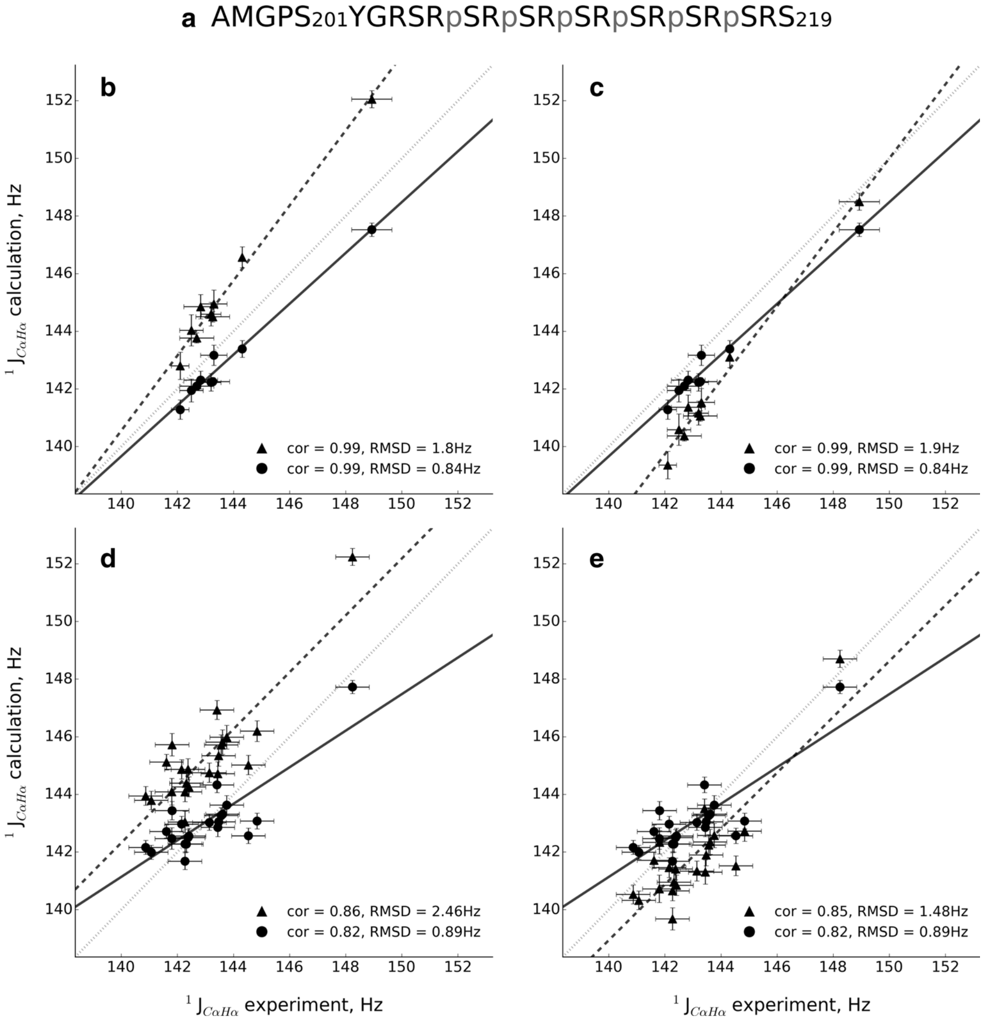

Fig. 3 **a** Primary sequence of SRSF1(RS1). Residues, which can be efficiently phosphorylated by SRPK1 (Xiang et al. 2013), are marked. **b–e** Experimentally measured $^1J_{C\alpha H\alpha}$ couplings in non-phosphorylated (**b**, **c**) and phosphorylated (**d**, **e**) SRSF1(RS1) plotted against the calculated coupling values from the ensembles obtained in Xiang et al. (2013). In case of phosphorylated SRSF1(RS1), phosphoserine residues were treated as serines. *Filled circles* and *solid lines* correspond to the values calculated using the Karplus-type equation with the new set of random coil values and the newly obtained parameters. *Filled triangles* and *broken lines* correspond to the calculations using random coil values from Vuister et al. (1993) with the equation parameters reported in Vuister et al. (1993) (**b**, **d**) or the reparameterized equation [using the random coil values and $^1J_{C\alpha H\alpha}$ coupling values reported in Vuister et al. (1993)] (**c**, **e**)

S211, S213, S215, S217) were treated as non-phosphorylated serines, high correlation and low RMSD values were obtained (Fig. 3d, e).

Table 3 Comparison of $^1J_{C\alpha H\alpha}$ couplings back-calculated from the non-phosphorylated (WT) and phosphorylated (RpS) SRSF1(RS1) ensembles

	Correlation	RMSD, Hz
WT expt. versus WT random	0.75 ± 0.02	0.79 ± 0.01
WT expt. versus WT selected	0.89 ± 0.01	0.84 ± 0.01
RpS expt. versus RpS random	0.14 ± 0.02	1.05 ± 0.01
RpS expt. versus RpS selected	0.42 ± 0.01	0.89 ± 0.01
WT expt. versus RpS expt.	0.55 ± 0.03	0.92 ± 0.03
WT random versus RpS random	0.71 ± 0.03	0.35 ± 0.02
WT selected versus RpS selected	0.57 ± 0.03	0.52 ± 0.01

Statistical errors were bootstrapped from the pool of 100 ensembles. Statistical errors for the experimental measurements were bootstrapped from the pool of 100 ensembles distributed normally around the experimentally measured value with the standard deviation given by the experimental uncertainty

To highlight the strengths of using the $^1J_{C\alpha H\alpha}$ couplings for cross-validation of IDP ensembles, we compared the SRSF1(RS1) ensembles selected using NMR observables with those generated randomly. Random ensembles were selected from the non-phosphorylated and phosphorylated pool of structures for SRSF1(RS1) (denoted WT in Table 3) and SRSF1(RpS1) (denoted RpS in Table 3), respectively. Since residue specific random coil values are large in magnitude and may bias correlation estimates, we have subtracted for the current analysis the random coil contributions from the experimental and back-calculated couplings. The sub-selected and randomly generated WT SRSF1(RS1) ensembles both match the experimentally measured $^1J_{C\alpha H\alpha}$ coupling values: we find a correlation of 0.89 and 0.75, respectively. It corresponds well with the previous findings, which showed that non-phosphorylated SRSF1(RS1) is disordered (Xiang et al. 2013). For the phosphorylated SRSF1(RS1) the selected ensembles correlate better with the experimental measurements than the random selections: correlations of 0.42 and 0.14. Although the correlation in this case is low, it allows distinguishing ensembles of interest from randomly chosen structures.

We have also investigated whether the $^{1}J_{C\alpha H\alpha}$ couplings are able to discriminate between the non-phosphorylated and phosphorylated SRSF1(RS1) ensembles. To allow for a proper comparison, the coupling values in the Arg/Ser region of the phosphorylated peptide were averaged, because their cross-peaks strongly overlap in case of the non-phosphorylated SRSF1(RS1). A moderate correlation of 0.55 between the WT and RpS ensembles is expected from the experimental measurements of the $^{1}J_{C\alpha H\alpha}$ couplings (Table 3). The random ensembles, even though selected from different pools of structures, show significant similarity ($cor = 0.71$) in comparison to the selected ensembles ($cor = 0.57$), thus showing that indeed the $^{1}J_{C\alpha H\alpha}$ couplings can successfully discriminate WT from RpS ensembles.

Discussion

We determined an improved set of amino-acid specific random coil values of $^{1}J_{C\alpha H\alpha}$ coupling constants, based on experimental $^{1}J_{C\alpha H\alpha}$ couplings observed in the 441-residue IDP Tau, as well as a short Tau peptide. The analysis showed that the $^{1}J_{C\alpha H\alpha}$ random coil values vary from 142.3 Hz in threonine to 144.1 Hz in alanine (excluding cysteine and proline; Fig. 1). For cysteine, the median value in Tau was 144.8 Hz and for proline 147.7 Hz (Table 1). The observed variation is well outside the experimental measurement error, because the magnitude of the $^{1}J_{C\alpha H\alpha}$ coupling constant is large and its experimental value can be determined with high accuracy. At the same time, the amino-acid specific variation is very small compared to the magnitude of $^{1}J_{C\alpha H\alpha}$, requiring accurate random coil values of $^{1}J_{C\alpha H\alpha}$. This is particularly important for the cross-validation of ensembles of IDPs, where rigid secondary structures are not formed and therefore the deviations from the random coil values are small.

The new set of random coil values formed the basis for a reparameterization of the Karplus-type equation of $^{1}J_{C\alpha H\alpha}$ (Fig. 2). Using the reparameterized Karplus-type equation a good correlation between experimental $^{1}J_{C\alpha H\alpha}$ couplings and values back-calculated from the 3D structure of seven globular proteins were obtained, indicating that the new parameter set can be used for structural analysis. We then applied the amino-acid specific random coil values and the reparameterized Karplus-type equation for cross-validation of the conformational ensembles of an intrinsically disordered peptide, the RS region of SRSF1, which was previously determined using a large number of residual dipolar couplings as well as $^{3}J_{HnH\alpha}$ scalar couplings (Xiang et al. 2013). The cross-validation resulted in a high correlation between experimental and back-calculated $^{1}J_{C\alpha H\alpha}$ values (Fig. 3). In addition, low RMSD values of 0.84 and 0.89 Hz were obtained, indicating that the new amino-acid specific random coil values in combination with the reparameterized Karplus-type equation are highly useful for analysis of conformational ensembles.

$^{1}J_{C\alpha H\alpha}$ coupling constants are particularly useful for analysis of molecular ensembles of IDPs, because they can be measured efficiently and with high accuracy. In addition, they are most sensitive for the dihedral angle ψ. Comparison of experimental and back-calculated values in non-phosphorylated and phosphorylated SRSF1(RS1) further suggested that phosphorylation of serine residues does not impair the use of $^{1}J_{C\alpha H\alpha}$ coupling constants for structural analysis. This is an important finding, because IDPs are highly regulated by post-translational modifications (Oldfield and Dunker 2014) and there is great interest in the structural consequences of phosphorylation and other post-translational modifications in IDPs.

Conclusion

Accurate amino-acid specific random coil values of the $^{1}J_{C\alpha H\alpha}$ coupling constant, in combination with reparameterized parameters for a Karplus-type equation, allow reliable cross-validation of molecular ensembles of IDPs.

Acknowledgments We thank Eckhard Mandelkow and Jacek Biernat for the Tau sample. This work was in part supported by the DFG through ZW71/8-1.

References

Allison JR, Varnai P, Dobson CM, Vendruscolo M (2009) Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements. J Am Chem Soc 131:18314–18326. doi:10.1021/Ja904716h

Ball KA, Phillips AH, Nerenberg PS, Fawzi NL, Wemmer DE, Head-Gordon T (2011) Homogeneous and heterogeneous tertiary structure ensembles of amyloid-β peptides. Biochemistry 50:7612–7628. doi:10.1021/bi200732x

Barfield M, Johnston MD (1973) Solvent dependence of nuclear spin–spin coupling-constants. Chem Rev 73:53–73. doi:10.1021/Cr60281a004

Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664. doi:10.1021/ja069124n

Billeter M, Neri D, Otting G, Qian YQ, Wuthrich K (1992) Precise vicinal coupling constants 3JHN α in proteins from nonlinear fits of J-modulated [15N, 1H]-COSY experiments. J Biomol NMR 2:257–274

Chattopadhyaya R, Meador WE, Means AR, Quiocho FA (1992) Calmodulin structure refined at 1.7 A resolution. J Mol Biol 228:1177–1192

Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 308:1011–1032. doi:10.1006/jmbi.2001.4750

Cleveland DW, Hwo SY, Kirschner MW (1977) Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. J Mol Biol 116:227–247

Edison AS, Markley JL, Weinhold F (1994a) Calculations of one-, two- and three-bond nuclear spin–spin couplings in a model peptide and correlations with experimental data. J Biomol NMR 4:519–542

Edison AS, Weinhold F, Westler WM, Markley JL (1994b) Estimates of phi and psi torsion angles in proteins from one-, two- and three-bond nuclear spin–spin couplings: application to staphylococcal nuclease. J Biomol NMR 4:543–551

Egli H, Vonphilipsborn W (1981) C-13-Nmr Spectroscopy.29. Conformational dependence of one-bond C-α, H spin coupling in cyclic-peptides. Helv Chim Acta 64:976–988. doi:10.1002/hlca.19810640404

Fawzi NL, Phillips AH, Ruscio JZ, Doucleff M, Wemmer DE, Head-Gordon T (2008) Structure and dynamics of the Aβ(21–30) peptide from the interplay of NMR experiments and molecular simulations. J Am Chem Soc 130:6145–6158. doi:10.1021/ja710366c

Fisher CK, Stultz CM (2011) Constructing ensembles for intrinsically disordered proteins. Curr Opin Struct Biol 21:426–431. doi:10.1016/j.sbi.2011.04.001

Hansen PE (1981) Carbon–hydrogen spin–spin coupling-constants. Prog Nucl Magn Reson Spectrosc 14:175–296. doi:10.1016/0079-6565(81)80001-5

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323:573–584

Jensen MR, Zweckstetter M, Huang JR, Blackledge M (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem Rev 114:6632–6660. doi:10.1021/cr400688u

Kontaxis G, Clore GM, Bax A (2000) Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. J Magn Reson 143:184–196. doi:10.1006/jmre.1999.1979

Kopple KD, Ahsan A, Barfield M (1978) Regarding H–C–C(O)–15-N coupling as an indicator of peptide torsional angle. Tetrahedron Lett 3519–3522

Lin S, Fu XD (2007) SR proteins and related factors in alternative splicing. Adv Exp Med Biol 623:107–122

Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE (2012) Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. J Am Chem Soc 134:3787–3791. doi:10.1021/ja209931w

Mantsyzov AB, Maltsev AS, Ying J, Shen Y, Hummer G, Bax A (2014) A maximum entropy approach to the study of residue-specific backbone angle distributions in α-synuclein, an intrinsically disordered protein. Protein Sci 23:1275–1290. doi:10.1002/pro.2511

Marsh JA, Forman-Kay JD (2012) Ensemble modeling of protein disordered states: experimental restraint contributions and validation. Proteins 80:556–572. doi:10.1002/prot.23220

Marsh JA, Teichmann SA, Forman-Kay JD (2012) Probing the diverse landscape of protein flexibility and binding. Curr Opin Struct Biol 22:643–650. doi:10.1016/j.sbi.2012.08.008

Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 17:3–14. doi:10.1016/j.sbi.2007.01.009

Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit 23:105–116. doi:10.1002/jmr.961

Mukrasch MD et al (2009) Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol 7:e34

Oldfield CJ, Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem 83:553–584. doi:10.1146/annurev-biochem-072711-164947

Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr Opin Struct Biol 24:98–105. doi:10.1016/j.sbi.2013.12.006

Rezaei-Ghaleh N, Blackledge M, Zweckstetter M (2012) Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. ChemBioChem 13:930–950. doi:10.1002/cbic.201200093

Schmidt JM, Howard MJ, Maestre-Martinez M, Perez CS, Lohr F (2009) Variation in protein C(α)-related one-bond J couplings. Magn Reson Chem 47:16–30. doi:10.1002/mrc.2337

Schwalbe M et al (2014) Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-synuclein in solution from NMR and small angle scattering. Structure 22:238–249. doi:10.1016/j.str.2013.10.020

Schwalbe M, Kadavath H, Biernat J, Ozenne V, Blackledge M, Mandelkow E, Zweckstetter M (2015) Structural impact of tau phosphorylation at threonine 231. Structure 23(8):1448–1458. doi:10.1016/j.str.2015.06.002

Tjandra N, Bax A (1997) Measurement of dipolar contributions to 1 J CH splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra. J Magn Reson 124:512–515. doi:10.1006/jmre.1996.1088

Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756

Uversky VN (2011) Flexible nets of malleable guardians: intrinsically disordered chaperones in neurodegenerative diseases. Chem Rev 111:1134–1166. doi:10.1021/cr100186d

Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246. doi:10.1146/annurev.biophys.37.032807.125924

Vijaykumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 a resolution. J Mol Biol 194:531–544. doi:10.1016/0022-2836(87)90679-6

Vuister GW, Bax A (1993) Quantitative J correlation—a new approach for measuring homonuclear 3-bond J(H(N)H(α) coupling-constants in N-15-enriched proteins. J Am Chem Soc 115:7772–7777

Vuister GW, Delaglio F, Bax A (1993) The use of 1JCαHα coupling constants as a probe for protein backbone conformation. J Biomol NMR 3:67–80

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure- function paradigm. J Mol Biol 293:321–331

Xiang S et al (2013) Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. Structure 21:2162–2174. doi:10.1016/j.str.2013.09.014

Zweckstetter M, Bax A (2001) Single-step determination of protein substructures using dipolar couplings: aid to structural genomics. J Am Chem Soc 123:9490–9491