J Biomol NMR (2011) 51:369–378
DOI 10.1007/s10858-011-9567-4

ARTICLE

Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins

Rodolfo M. Rasia · Ewen Lescop · Javier F. Palatnik · Jérôme Boisbouvier · Bernhard Brutscher

Received: 24 June 2011 / Accepted: 25 August 2011 / Published online: 14 September 2011

Abstract It has been demonstrated that protein folds can be determined using appropriate computational protocols with NMR chemical shifts as the sole source of experimental restraints. While such approaches are very promising they still suffer from low convergence resulting in long computation times to achieve accurate results. Here we present a suite of time- and sensitivity optimized NMR experiments for rapid measurement of up to six RDCs per residue. Including such an RDC data set, measured in less than 24 h on a single aligned protein sample, greatly improves convergence of the Rosetta-NMR protocol, allowing for overnight fold calculation of small proteins. We demonstrate the performance of our fast fold calculation approach for ubiquitin as a test case, and for two RNA-binding domains of the plant protein HYL1. Structure calculations based on simulated RDC data highlight the importance of an accurate and precise set of several complementary RDCs as additional input restraints for high-quality de novo structure determination.

Keywords Fast NMR · RDC · Rosetta · Protein structure calculation

Introduction

Recent years have witnessed an increasing interest in the development of NMR methods that provide atom-resolved structural and dynamic information on proteins from a reduced experimental data set that can be recorded in short overall time using sensitive NMR pulse schemes combined with fast multidimensional data acquisition techniques (Kazimierczuk et al. 2010; Coggins et al. 2010). Reduced NMR data acquisition times become a crucial issue whenever the molecular system under investigation has a short lifetime in the NMR sample tube. It is also important for "high-throughput" NMR applications in the context of structural genomics initiatives. In addition, focused NMR methods requiring only a limited amount of data are attractive whenever the high-resolution structure is not the main objective, but one wants to define the fold of a protein for the subsequent characterization of molecular interfaces in complexes with other protein or nucleic-acid partners, or potential drugs.

Residual dipolar couplings (RDCs), provide a particular rich source of structural information as they are sensitive to

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9567-4) contains supplementary material, which is available to authorized users.

R. M. Rasia · E. Lescop · J. Boisbouvier (✉) · B. Brutscher (✉)
Institut de Biologie Structurale, Jean-Pierre Ebel CNRS/CEA/UJF, 41 rue Jules Horowitz, 38027 Grenoble Cedex, France
e-mail: jerome.boisbouvier@ibs.fr

B. Brutscher
e-mail: Bernhard.brutscher@ibs.fr

R. M. Rasia · J. F. Palatnik
Instituto de Biología Molecular y Celular de Rosario, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

Present Address:
E. Lescop
Institut de Chimie des Substances Naturelles, CNRS, Gif-sur-Yvette, Grenoble, France

J. Boisbouvier · B. Brutscher
CEA, IBS-Grenoble, Grenoble, France

J. Boisbouvier · B. Brutscher
Université de Grenoble 1, Grenoble, France

both the local and global molecular geometry (Tjandra and Bax 1997; Prestegard et al. 2004; Blackledge 2005). Furthermore, RDCs can be measured, and assigned to nuclear spin pairs in the protein, once sequential backbone resonance assignment has been completed, without the need of tedious and time-consuming side chain resonance and ^{1}H-^{1}H NOE assignments. Spin coupling constants involving backbone nuclei are therefore, after chemical shifts, the second most easily, and accurately measurable NMR observables. The measured RDCs can be translated into structural restraints of high accuracy that can be used for NMR assignment of proteins of known structure (Jung and Zweckstetter 2004; Korukottu et al. 2007; Wang et al. 2011), or the recognition of protein folds (Valafar and Prestegard 2003). RDCs can also yield crucial information on the relative orientations of individual domains within macromolecular assemblies (Garrett et al. 1999; Koenig et al. 2002; Jain et al. 2004), even in the case of weak binding affinities (Ortega-Roldan et al. 2009). The de novo determination of protein structure using RDC data as the sole geometric constraints was also achieved using different strategies (Fowler et al. 2000; Hus et al. 2001; Kontaxis et al. 2005). However, due to the degeneracy of the orientations defined by the RDC values these methods rely on the acquisition of RDC data sets with respect to independent alignment tensors, and need very complete RDC data sets, thus limiting their application. The combination of RDC data with ab initio protein structure prediction methods such as Rosetta allowed the calculation of protein folds using restricted RDC data sets from a single alignment medium (Rohl and Baker 2002). More recently, strategies for fold calculation using chemical shift as sole restraints were designed and shown to yield correct folds for moderately sized proteins (Cavalli et al. 2007; Shen et al. 2008). Inclusion of a limited set of NOEs and RDC data further extended the applicability of this method to monomeric proteins of up to 25 kDa (Raman et al. 2010), transient protein-folding intermediates (Korzhnev et al. 2010), and molecular complexes (Montalvao et al. 2008; Das et al. 2009; Sgourakis et al. 2011). A drawback of these methods, however, is that they are computationally time consuming (Shen et al. 2008; Raman et al. 2010), typically few days of calculation are required using a computer cluster of moderate size.

Accurate RDC measurements for a large number of sites along the protein backbone require the recording of high-resolution high-dimensional ($\geq$3D) NMR correlation spectra. The acquisition time required for indirect time-domain sampling in 3D experiments is often prohibitive with respect to the intrinsic sensitivity obtained at high magnetic field NMR spectrometers equipped with cryogenically cooled high-Q probes. This observation has triggered the development of alternative, fast NMR data acquisition schemes yielding reduced acquisition times without a significant loss of spectral information. Various conceptually different tools have been proposed for the acquisition and processing of sparsely sampled NMR data (Freeman and Kupce 2003; Felli and Brutscher 2009). An alternative complementary approach to sparse data sampling is the use of short inter-scan delays to speed up data acquisition. Recently, we have introduced SOFAST-HMQC (Schanda and Brutscher 2005), and BEST-type (Schanda et al. 2006a; Lescop et al. 2007) correlation experiments that allow recording of protein correlation spectra within a few seconds (2D) or minutes (3D) for NMR assignment purposes. Combining fast-pulsing NMR experiments with advanced data analysis tools, as implemented for example in the BATCH protocol, has proven to yield reliable backbone resonance assignment of small proteins within a few hours (Lescop and Brutscher 2009).

Here, we introduce a suite of BEST-optimized correlation experiments designed for the accurate measurement of six different RDCs within bi-peptides along the protein backbone. The whole set of four high-resolution 3D spectra can be recorded in typically less than 24 h on a $\sim$1 mM protein sample using a modern NMR spectrometer equipped with a cryogenic probe. We further show that the use of such an extended set of RDC data greatly improves the convergence of Rosetta structure calculations toward the correct protein fold, allowing the calculation of backbone protein structures with limited computational resources in a short time (typically overnight).

Materials and methods

Sample preparation

Weakly aligned samples were prepared as follows. Lyophilized ^{13}C, ^{15}N labeled ubiquitin was dissolved in 5% DMPC:DHPC (3:1) bicelles in 10 mM phosphate buffer, 15 mM NaN_3, pH 6.0 at 1.5 mM concentration (Ottiger and Bax 1998). For HYL1 domains the lyophilized protein samples were dissolved in a preformed 5% C12E5/Hexanol (C12E5:Hexanol ratio 0.96) liquid crystalline phase (Rückert and Otting 2000).

NMR data acquisition

All experiments were performed at 25°C on a Varian 600 MHz spectrometer equipped with a triple-resonance coldprobe. Overall experimental times for the ubiquitin sample were: 1 h for BEST HNCO-JNH, 6 h for BEST HNCO-JCH, 4 h for BEST HNCO-JCC, and 3 h for BEST HN(CO)CA-JCH. The spectra were recorded with 50/23/602, 80/23/602, 160/23/602, 55/23/602 complex points in

the ^{13}C (t_1)/^{15}N (t_2)/^{1}H (t_3) dimensions, respectively, with spectral widths set to 7,530 Hz (^{1}H), 1,318 Hz (^{15}N), 1,200 Hz (^{13}CO in BEST HNCO-JNH), and 2,000 Hz (^{13}CA/^{13}CO). The recycle delay was set to 300 ms, and two transients per increment were acquired, except for BEST HN(CO)CA-JCH, where four transients were acquired to increase the signal to noise ratio. For both HYL1 domains, four transients were acquired per increment in each experiment, except for BEST HNCO-JCH where two transients were acquired. The recycle delay was set to 500 ms in order to maximize the signal to noise per unit time. The experimental times for these samples were as follows: BEST HNCO-JNH, 3h30′; BEST HNCO-JCH, 9h20′; BEST HNCO-JCC, 6 h; BEST HN(CO)CA-JCH, 4 h, for a total experimental time of 22h50′.

BEST-type pulse sequences

The BEST-type HNC correlation experiments proposed here for accurate measurement of six scalar and dipolar spin coupling constants per bi-peptide are shown in Fig. 1. The BEST (Band-selective Excitation Short-Transient) scheme results in minimal perturbation of aliphatic and water ^{1}H spins by applying exclusively band-selective pulses, PC9(Kupce and Freeman 1993), E-BURP2, and RE-BURP (Geen and Freeman 1991), or pairs of broadband inversion pulses (BIP) (Smith et al. 2001) on the ^{1}H channel. The large amount of aliphatic and water ^{1}H spin polarization present at the end of the BEST pulse sequence then enhances longitudinal (spin–lattice) relaxation of amide hydrogen spins via dipole–dipole interactions (NOE effects) and hydrogen exchange. More details about the BEST concept can be found elsewhere (Schanda et al. 2006a; Lescop et al. 2007). Sensitivity-enhanced and INEPT-based versions of these experiments are shown in Fig. 1a, b, respectively. Generally the sensitivity-enhanced version (a) yields higher sensitivity for proteins in isotropic solution, whereas sequence (b) is preferred for aligned protein samples. All experiments are based on sequential correlation of the amide ^{1}H and ^{15}N of residue i with the $^{13}C'$ or $^{13}C^{\alpha}$ of residue (i − 1). The particular features of each pulse sequence are described in the following paragraphs.

3D BEST HNCO-JNH for the measurement of H^N–N coupling constants

The pulse sequence of the BEST HNCO-JNH experiment is obtained by combining the insert of Fig. 1c with the BEST schemes 1a or 1b. The one-bond ^{1}H–^{15}N coupling constant is obtained from the line splitting measured along the $^{13}C'$ dimension. This line splitting is artificially enhanced by means of accordion spectroscopy (Bodenhausen and Ernst 1982). The ^{1}H–^{15}N coupling evolution is active during part of the initial ^{15}N–$^{13}C'$ constant-time (CT) transfer step as well as during the $^{13}C'$ frequency labeling period. This results in adjustable scaling of the apparent line splitting by a factor $(1 + \lambda)$ with respect to the actual spin-coupling constant $^{1}(J + D)_{HN-N}$. Typical values of λ are in the range $0.5 < \lambda < 1.0$.

3D BEST HNCO-JCH for the measurement of H^N–C′ coupling constants

The pulse sequence of the BEST HNCO-JCH experiment is obtained by combining the insert of Fig. 1d with the BEST schemes 1a or 1b. The two-bond $^{1}H^N$–$^{13}C'$ coupling constant is again obtained from the line splitting measured along the $^{13}C'$ dimension. Because of the magnitude of the $^{2}(J + D)_{HN-C'}$ coupling that is small compared to the line width, we use a *J*-mismatch compensated DIPSAP spin-state selection filter (Brutscher 2001) to separate the doublet lines in two sub-spectra. A total of three data sets need to be recorded using different settings of the delay ε and phase φ_4 (see caption Fig. 1). As DIPSAP filtering is achieved during the ^{15}N–$^{13}C'$ transfer step, it does not induce any signal loss due to additional spin relaxation.

3D BEST HNCO-JCC for the measurement of C′–C^{α} and H^N–C^{α} coupling constants

The pulse sequence of the BEST HNCO-JCC experiment is obtained by combining the insert of Fig. 1e with the BEST schemes 1a or 1b. This experiment exploits the E-COSY principle (Griesinger et al. 1985) for simultaneous measurement of two coupling constants from the line splitting observed along orthogonal dimensions. The $^{1}(J + D)_{C'-CA}$ coupling is measured along the $^{13}C'$ dimension while $^{3}(J + D)_{HN-CA}$ coupling information is obtained in the ^{1}H dimension of the 3D spectrum.

3D BEST CT-HN(CO)CA-JCH for the measurement of C^{α}–H^{α} and H^N–H^{α} coupling constants

The pulse sequence of the BEST HN(CO)CA-JCH experiment is obtained by combining the insert of Fig. 1f with the BEST schemes 1a or 1b. This experiments results in a sequential H–N–CA instead of a H–N–CO correlation spectrum. The $^{13}C^{\alpha}$ frequency labeling is performed in a CT manner to avoid additional line splitting due to homonuclear C^{α}–C^{β} couplings. Again, an E-COSY-type peak pattern is obtained for each correlation (residue) with a $^{1}(J + D)_{CA-HA}$ line splitting along the ^{13}C dimension, and a $^{4}(J + D)_{HN-HA}$ line separation in the ^{1}H dimension. The E-COSY splitting may be affected if the $^{1}H^{\alpha}$

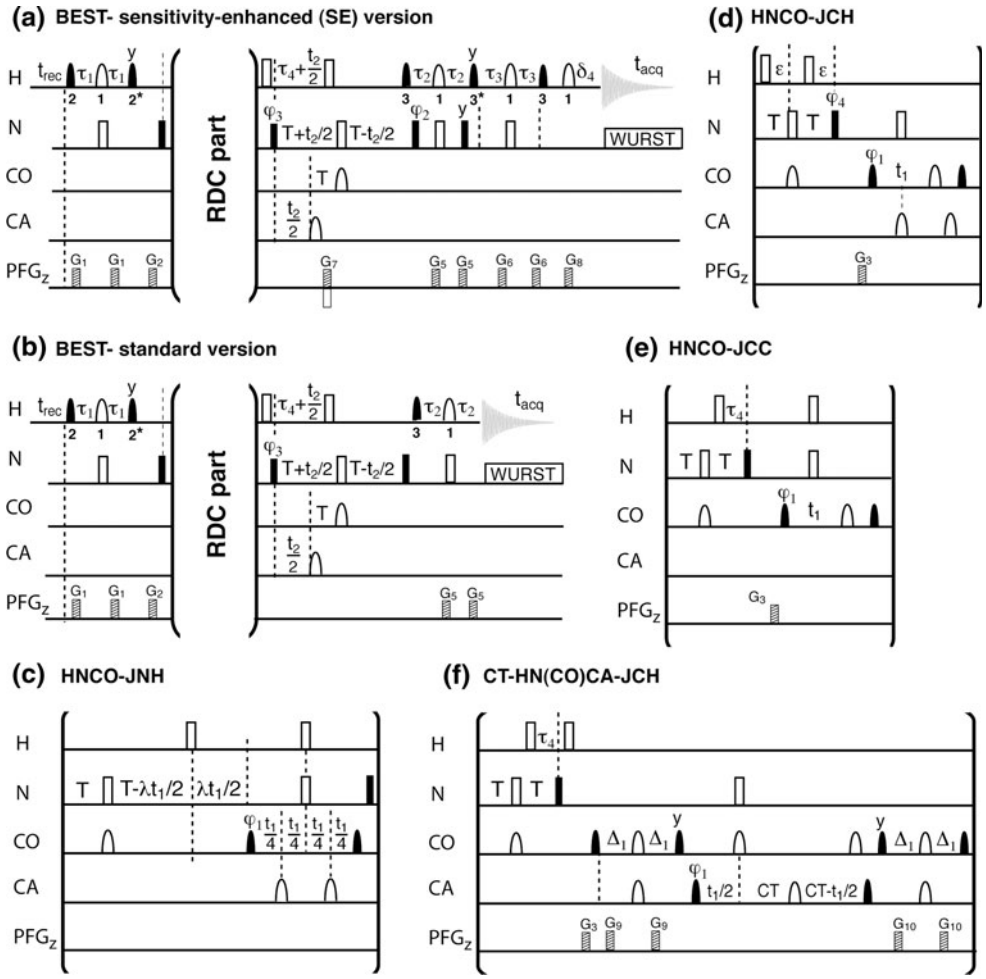

Fig. 1 BEST-type experiments for the measurement of spin coupling constants in the protein backbone. The common parts of the four experiments are shown in (**a**) sensitivity-enhanced version, and (**b**) standard INEPT-based version. The *different inserts* correspond to the following experiments (allowing measurement of the following coupling constants): **c** BEST HNCO-JNH ($^1(J + D)_{HN-N}$), **d** BEST HNCO-JCH ($^2(J + D)_{HN-C'}$), **e** BEST HNCO-JCC ($^1(J + D)_{C'-CA}$ and $^3(J + D)_{HN-CA}$), **f** BEST HN(CO)CA-JCH ($^1(J + D)_{CA-HA}$ and $^4(J + D)_{HN-HA}$). *Filled* and *open* pulse *symbols* indicate 90° and 180° rf pulses. Unless indicated, all pulses are applied with phase x. All selective ^{1}H pulses are centered at 8.5 ppm, covering a bandwidth of 4.0 ppm, with the following shapes: [1] REBURP (Geen and Freeman 1991), [2] PC9 (Kupce and Freeman 1993) and [3] E-BURP2 (Geen and Freeman 1991). The durations of these pulses are respectively 2 ms, 3 ms and 1.92 ms at 600 MHz ^{1}H frequency. A *star* indicates a flip *back* pulse obtained by time inversion of the excitation pulse shape. *Open squares* on ^{1}H indicate BIP-720-50-20 pulses (Smith et al. 2001). CO pulses have the shape of the center lobe of a sinx/x function, whereas CA and CA/CB pulses are applied with a *rectangular shape* and zero excitation at the CO frequency. The transfer delays are set to: $\tau_1 = 2.4$ ms—0.5 $(\delta_1 + \delta_2)$, $\tau_2 = 2.4$ ms—0.5 δ_1, $\tau_3 = 2.4$ ms—0.5 δ_1—0.65 δ_3, $\tau_4 = 2.7$ ms, T = 14.5 ms, and $\Delta_1 = 4.5$ ms, with δ_1, δ_2, δ_3, and δ_4 corresponding to the pulse lengths of the REBURP, PC9, E-BURP2, and G_8 gradient, respectively. These settings take into account spin coupling (and chemical shift) evolution during the various shaped ^{1}H pulses (Lescop et al. 2010). For the DIPSAP filter (Brutscher 2001) in sequence (**d**) three repetitions of the experiments are recorded with the following settings: (I) $\varepsilon = 0$, $\varphi_4 = y$, (II) $\varepsilon = 1.35$ ms, $\varphi_4 = x$, and (III) $\varepsilon = 2.7$ ms, $\varphi_4 = -x$. The two spin-state-separated sub spectra are then generated from these data by computing the linear combinations: 0.73 (I) + 0.27 (III) ± (II). Pulsed field gradients, G1–G8 are applied along the z-axis (PFGz) with durations of 200 ms to 2 ms and field strengths ranging from 5 to 40 G/cm. The phase cycling is: $\varphi_1 = x,-x$, $\varphi_2 = x$, $\varphi_3 = x$, $\varphi_4 = y$, and the receiver $\varphi_{rec} = -x$, x. Quadrature detection in t1 is obtained by time-proportional phase incrementation of φ_1 according to TPPI-States. For sensitivity-enhanced quadrature detection in t_2 (sequence a), echo-antiecho data are recorded by inverting the sign of gradient G_7 and phase φ_2. For quadrature detection in t_2 in the INEPT-based version (**b**), of φ_3 is incremented according to TPPI-States. All pulse sequences (in Varian pulse program language) are available from the authors upon request

resonances are within the excitation bandwidth of the $^1H^N$ selective pulses. This was indeed the case for the F4$^1H^\alpha$–V5$^1H^N$ coupling (δ F4$^1H^\alpha = 5.67$ ppm), which was excluded from the data analysis.

Results and discussion

In order to evaluate the accuracy of the BEST-optimized RDC experiments (Fig. 1), we have recorded spectra for a

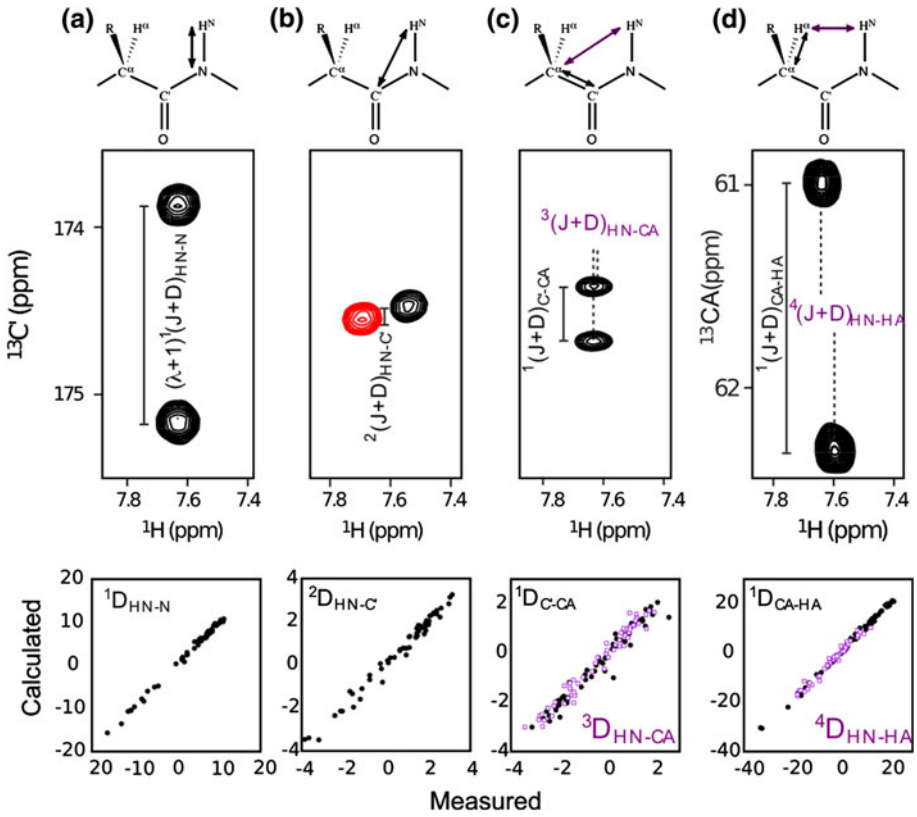

Fig. 2 Part of ^{1}H–^{13}C planes extracted from 3D data sets of BEST-HNCO-JNH (**a**), BEST HNCO-JCH (**b**), BEST HNCO-JCC (**c**), and BEST HN(CO)CA-JCH (**d**), showing *cross-peak* patterns of ubiquitin residue Gln62. The spin couplings that can be extracted from the different spectra are indicated on *top*. RDC correlation *plots* for each coupling measured on ubiquitin aligned in a bicelle medium are shown at the *bottom* of each spectrum [calculated RDC values are predicted from the high resolution solution structure of ubiquitin (PDB code : 1D3Z)] (Cornillescu et al. 1998)

sample of ubiquitin dissolved in a weak alignment medium containing 5% phospholipid bicelles. Figure 2 illustrates the quality of the measured RDC data, and shows the coupling-induced cross peak patterns observed in the different NMR spectra. The alignment tensor for ubiquitin in this medium was obtained by singular value decomposition using all measured RDC values and the expected RDC values were back calculated and compared to the measured values. The Q factor for the data obtained with respect to the RDC-refined solution structure of ubiquitin (1D3Z) was 0.158. A sensitivity comparison of the BEST sequences with comparable standard hard-pulse schemes is shown in figure S1. The maximum of the sensitivity curve obtained using a BEST-type experiment is found at an inter-scan delay of $\sim$500 ms, as compared to $\sim$1.5 s for the reference experiment; the sensitivity gain is $\sim$30% if both experiments are performed under their respective optimal sensitivity conditions. A BEST-type experiment run with an inter-scan delay of 300 ms yields the same sensitivity as a standard experiment using a 1.5 s recycle delay, and is roughly 2-times more sensitive than a standard experiment using the same short inter-scan delay. On one hand, the high repetition rate in BEST-type experiments allows for sampling of longer evolution times in the indirect dimensions yielding increased spectral resolution for a given experimental time that is crucial for accurate determination of peak positions as required for spin coupling (and RDC) measurements. On the other hand, if sample stability is a problem, the BEST-optimized RDC experiments can be performed in an overall experimental time reduced by a factor of 3–5 with respect to standard methods, without compromising experimental sensitivity.

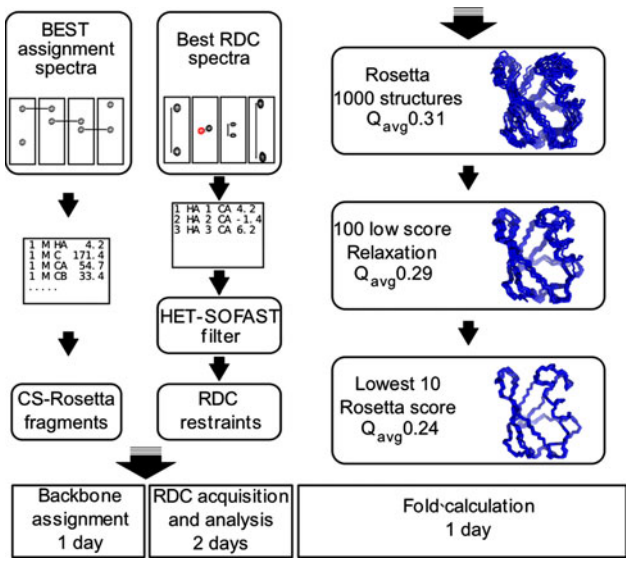

Fig. 3 Schematic representation of the fold calculation protocol. Time estimates for the RDC data collection are based on the 1.5 mM ubiquitin sample used in this work. The *Q* factors shown are the average over the whole set of structures. The backbone of 10 randomly chosen conformers of each ensemble are shown on the *right*

For NMR-restraint driven protein fold calculation, NMR backbone chemical shifts (HN, N, CO, CA), and four out of the six RDCs measured per bi-peptide ($^1D_{HN–N}$, $^1D_{CA–HA}$, $^1D_{C'–CA}$, $^2D_{C'–HN}$) were then used as input restraints in the Rosetta-NMR software (Rohl and Baker 2002; Raman et al. 2010). In order to avoid computational problems related to conformational averaging of measured RDCs in highly flexible loop regions, we have used a HET-SOFAST (Schanda et al. 2006b) experiment to identify residues

Fig. 5 RDC data used for calculations on IIB-MTL. $^1D_{HN–N}$ and $^2D_{HN–C'}$ were obtained from the RCSB (Legler et al. 2004). The other two RDC types were simulated using the calculated alignment tensor and adding Gaussian noise (1 Hz SD for $^1D_{CA–HA}$ and 0.5 Hz SD for $^1D_{CA–C'}$). In the 90 and 80% data sets, RDCs from the following residues were excluded: [87, 88, 89, 90, 42, 43, 44, 45, 46, 47] (90%) or [12, 13, 14, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 87, 88, 89, 90, 42, 43, 44, 45, 46, 47] (80%). In the extra noise data sets, the RDC data were simulated adding Gaussian noise with two (or four times) the estimated deviation for each type of RDC: i.e. 2 Hz (respectively 4 Hz) SD for $^1D_{CA–HA}$ and $^1D_{HN–N}$, and 1 Hz (respectively 2 Hz) SD for $^1D_{CA–C'}$ $^2D_{HN–C'}$. Below each data set, the resulting ten lowest score structure bundle obtained is shown ▶

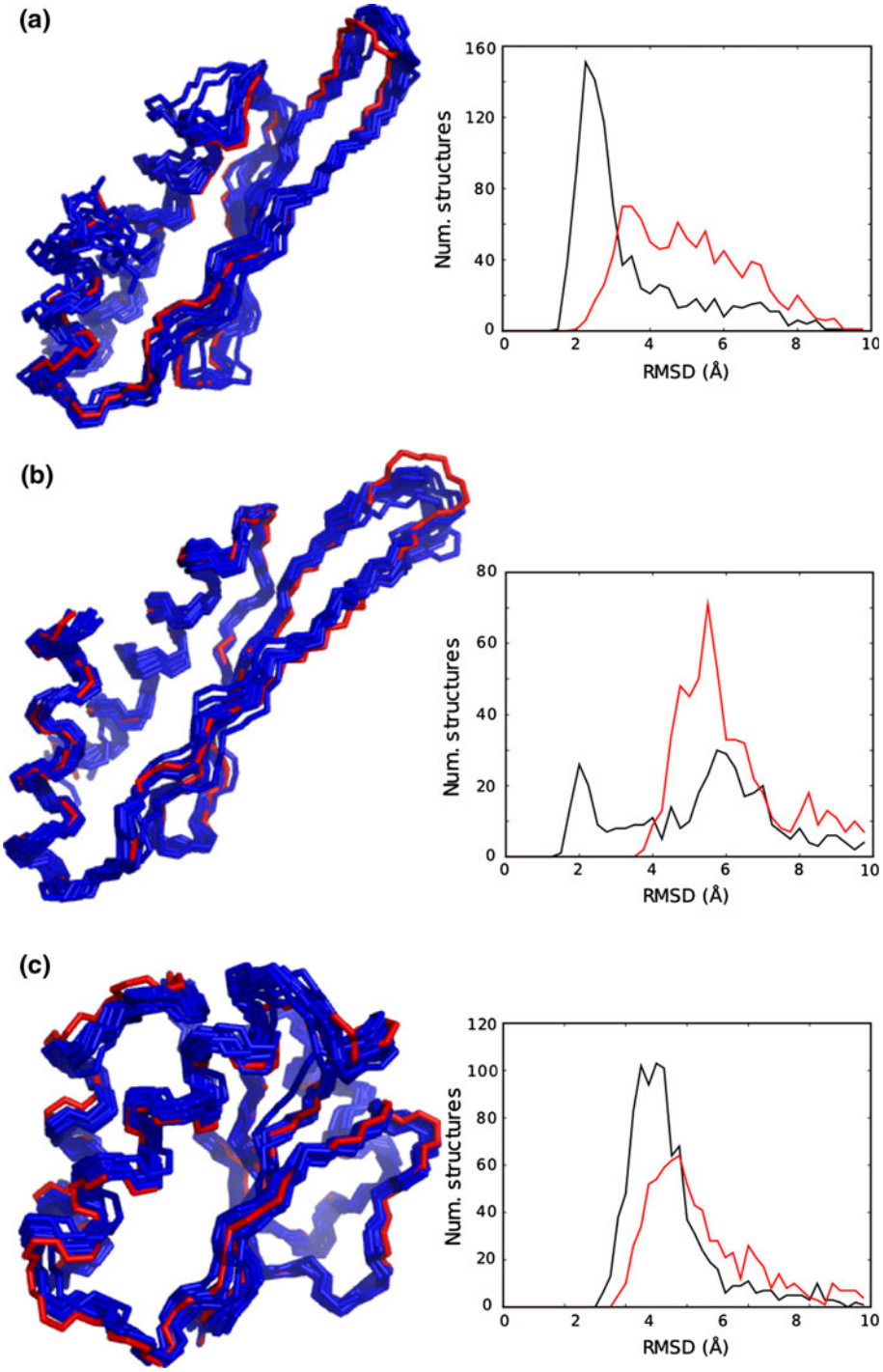

Fig. 4 Protein folds obtained using the present protocol. **a** HYL1-dsRBD1 **b** HYL1-dsRBD2 and **c** IIB-MTL. The ten lowest score structures (*blue*) are superimposed on the deposited crystal structure (*red*). The *graph* below shows the distribution of the RMSD of the calculated structures at the fragment insertion step either with (*black*) or without (*red*) RDC restraints, with respect to a reference structure

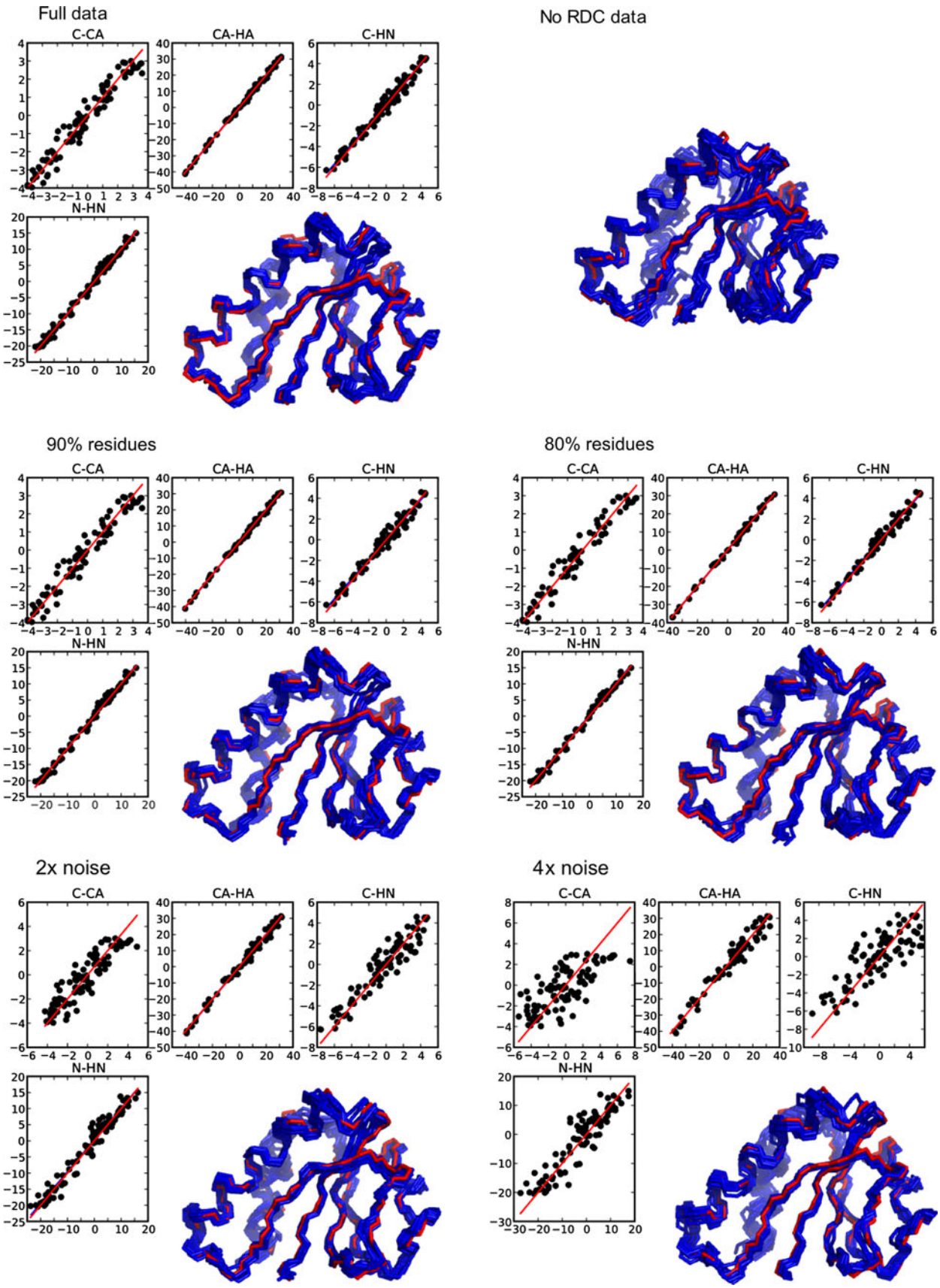
Full data
C-CA
CA-HA
C-HN
N-HN
No RDC data
90% residues
C-CA
CA-HA
C-HN
N-HN
80% residues
C-CA
CA-HA
C-HN
N-HN
2x noise
C-CA
CA-HA
C-HN
N-HN
4x noise
C-CA
CA-HA
C-HN
N-HN

Table 1 Precision and accuracy of Rosetta-NMR calculations

Protein	Precision	Accuracy
Ubiquitin	0.47 ± 0.06	0.67 ± 0.05
HYL1-1	1.63 ± 0.34	1.71 ± 0.41
HYL1-2	1.35 ± 0.36	1.23 ± 0.19
IIB-MTL	0.80 ± 0.11	1.18 ± 0.04
90% data	0.89 ± 0.15	1.19 ± 0.08
80% data	1.07 ± 0.23	1.28 ± 0.08
2× noise	0.87 ± 0.17	1.20 ± 0.11
4× noise	1.21 ± 0.35	1.27 ± 0.22
No data	1.58 ± 0.34	1.47 ± 0.12

RMSD values are given in Å for the 10 lowest score structures at the relaxation step. Precision refers to the average pairwise RMSD in the ensemble, whereas accuracy refers to the RMSD between the average and the deposited (reference) structure. For IIB-MTL, the calculation has been repeated for different input data sets as described in the text

located in protein regions with high conformational flexibility (figure S2). In these experiments, the parameter λ_{NOE} provides a qualitative measure of the average proton density and structural rigidity at each amide site, with large λ_{NOE} values indicative of increased local dynamics. The RDC data for residues (bi-peptides) above a threshold value of $\lambda_{NOE} > 0.7$ were not used for the fold calculations. The high intrinsic sensitivity of the HET-SOFAST experiment allows data recording in only a few minutes. The protocol used for structure calculation is schematically outlined in Fig. 3. Fragments were selected using the CS-Rosetta MFR protocol (Shen et al. 2008), and 1,000 centroid representations of protein structures were assembled with the standard Rosetta ab initio approach (Rohl et al. 2004). The 100 best structures, ranked according to their Rosetta score, were then selected and subject to the Rosetta full atom relaxation protocol. In this final part of the calculation a high-resolution all-atom model is generated, in which side chain conformations are added to the backbone model. The model is then further refined using a Monte Carlo minimization with an all-atom energy function (Rohl et al. 2004). Finally the 10 best structures were selected based on the Rosetta full atom score. The set of four measured RDC constraints was included in both the fragment assembly and relaxation protocols. The $^{3}D_{CA-HN}$ and $^{4}D_{HA-HN}$ RDCs that were not included in the fold calculation protocol were then used as independent data for cross-validation of the Rosetta results (figures S3, S6 and S7).

In the following we report on the results obtained for three different small proteins (<10 kDa) for which a set of six types of RDCs were measured as described above: ubiquitin as a test case with a well-known structure, and two double-stranded RNA binding domains (dsRBDs) of HYL1, a protein implicated in microRNA processing in *Arabidopsis thaliana*, for which no structure was available at the beginning of this study. While this work was in progress, crystal structures of both dsRBDs of HYL1 were published (Yang et al. 2010). For ubiquitin, the backbone pairwise RMSD of the ten lowest energy structure bundle was 0.47 ± 0.06 Å, and the RMSD of this final ensemble with respect to the NMR- refined solution structure (1D3Z) was 0.67 ± 0.05 Å. For the dsRBDs domains of HYL1, the result of the fold calculation shows the canonical $\alpha\beta\beta\beta\alpha$ topology characteristic of dsRBDs (Fig. 4a, b). The lowest energy structures superimpose with an average backbone pair-wise RMSD of 1.63 ± 0.34 Å for HYL1-1 and 1.35 ± 0.36 Å for HYL1-2. The deviations from the X-ray structures are 1.71 ± 0.41 Å and 1.23 ± 0.19 Å for HYL1-1 and HYL1-2, respectively. These results indicate that our fold calculation protocol yields both precise and accurate results. The use of RDC restraints improves the convergence of the calculation towards the correct fold (Figs. 4 and S4, see above), and thus dramatically reduces computational power and time requirements. In our approach, accurate structures are obtained using Rosetta directly at the fast low-resolution exploration phase from about 1,000 repetitions of the calculation protocol. In the absence of such RDC-derived geometric restraints, a significantly larger number of structures, typically between 5,000 and 10,000, needs to be generated in order to sufficiently explore the conformational space. In addition, to obtain accurate protein models a full atom description of the protein (including side chains) is required, and minimization of the resulting structures is computationally time consuming (Shen et al. 2010).

Finally, we have investigated the dependence of the fold calculation protocol proposed here on the quality and completeness of the RDC data available for the computation. For this analysis, we have chosen the cytoplasmic b domain of the mannitol transporter of the Escherichia coli phospho-transferase system (Legler et al. 2004) (10.5 kDa, IIB-MTL) as a model system using a set of experimental and simulated RDC data. Experimental $^{1}D_{HN-N}$, $^{2}D_{C'-HN}$ values were taken from the RCSB database and used together with the deposited structure to calculate the alignment tensor by singular value decomposition. The missing RDC values were then obtained from the structure and the calculated alignment tensor using software PALES (Zweckstetter and Bax 2000). In order to simulate the experimental conditions, random Gaussian noise was added to the calculated RDC values (Fig. 5). We did not attempt to take into account possible errors in the alignment tensor estimation. To test the robustness of the fold calculations we added 2× or 4× extra noise to the input data, or alternatively we removed RDC values from 10% or 20% of the residues in the protein. In addition we performed calculations using only one or two (instead of four)

types of RDCs. In all cases, we analyzed both the distribution of the structures obtained at the low-resolution exploration phase (figures S4 and S5), as well as the precision and accuracy of the ten final structures selected (Table 1). Reducing the amount of accurate RDC data to ~80% of all potentially measurable RDCs does not have a major influence on the results. However, a decrease in the number of different types of RDCs per bi-peptides lead to substantially worse structural ensembles, both in terms of accuracy and precision. This result highlights the importance of sensitive NMR pulse schemes for the accurate measurement of coupling constants in the context of fast structure calculation protocols based on RDCs and chemical shifts as the sole experimental restraints.

Conclusion

In summary, we have presented a protocol for time-efficient protein fold determination using a comprehensive set of chemical shift and RDC data as input to the Rosetta-NMR calculation protocol. A series of sensitivity- and resolution-optimized NMR experiments is proposed that allows accurate quantification of six RDC values per residue in a short overall experimental time. Such optimized experiments not only save spectrometer time, but they may be crucial in cases where the protein sample is only stable for a short time under the conditions required for NMR data acquisition. Furthermore, fast and accurate measurement of a set of backbone RDCs is of crucial importance for the detailed characterization of weak molecular complexes by NMR requiring the availability of such RDC data for a wide range of molecular ratios of the interactions partners (Ortega-Roldan et al. 2009). Using four RDC values per peptide plane as input data to the Rosetta program improves the convergence thus allowing fold calculation of a small protein to be performed overnight on a desktop computer. NMR constraint-driven fold determination may also enable the study of timely unstable proteins that may escape investigation by classical NMR approaches. Protein folds obtained using computational tools such as Rosetta combined with a sparse set of experimental data can provide first insights into protein function at the early stage of an NMR study.

Acknowledgments The authors thank Martin Blackledge for stimulating discussions, Nicolas Bologna for providing clones of HYL1 domains, Isabel Ayala for help with sample preparation and the Grenoble Partnership for Structural Biology for access to the high field NMR and isotope labeling platforms. This work was supported by the French research agency (ANR JCJC05-0077 and ANR JCJC06-0034), HFSP (RG-0057/2006), the Agencia Nacional para la Promoción de la Ciencia y la Tecnología (PICT-2007-720), and the European commission (I3, Contract No. 026145).

References

Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog NMR Spectrosc 46:23–61

Bodenhausen G, Ernst R (1982) Direct determination of rate constants of slow dynamic processes by two-dimensional "accordion" spectroscopy in nuclear magnetic resonance. J Am Chem Soc 104:1304–1309

Brutscher B (2001) Accurate measurement of small spin–spin couplings in partially aligned molecules using a novel J-mismatch compensated spin-state-selection filter. J Magn Reson 151:332–338

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

Coggins BE, Venters RA, Zhou P (2010) Radial sampling for fast NMR: concepts and practices over three decades. Prog NMR Spectrosc 57:381–419

Cornilescu G, Marquardt JL, 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

Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D (2009) Simultaneous prediction of protein folding and docking at high resolution. Proc Natl Acad Sci USA 106:18978–18983

Felli IC, Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. Chemphyschem 10:1356–1368

Fowler CA, Tian F, Al-Hashimi HM, Prestegard JH (2000) Rapid determination of protein folds using residual dipolar couplings. J Mol Biol 304:447–460

Freeman R, Kupce E (2003) New methods for fast multidimensional NMR. J Biomol NMR 27:101–113

Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM (1999) Solution structure of the 40, 000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. Nat Struct Biol 6:166–173

Geen H, Freeman R (1991) Band-selective radiofrequency pulses. J Magn Reson 93:93–141

Griesinger C, Soerensen OW, Ernst RR (1985) Two-dimensional correlation of connected NMR transitions. J Am Chem Soc 107:6394–6396

Hus JC, Marion D, Blackledge M (2001) Determination of protein backbone structure using only residual dipolar couplings. J Am Chem Soc 123:1541–1542

Jain NU, Wyckoff TJO, Raetz CRH, Prestegard JH (2004) Rapid analysis of large protein–protein complexes using NMR-derived orientational constraints: the 95 kDa complex of LpxA with acyl carrier protein. J Mol Biol 343:1379–1389

Jung Y, Zweckstetter M (2004) Backbone assignment of proteins with known structure using residual dipolar couplings. J Biomol NMR 30:25–35

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Kozminski W (2010) Random sampling in multidimensional NMR spectroscopy. Prog. NMR Spectrosc 57:420–434

Koenig BW, Kontaxis G, Mitchell DC, Louis JM, Litman BJ, Bax A (2002) Structure and orientation of a G protein fragment in the receptor bound state from residual dipolar couplings. J Mol Biol 322:441–461

Kontaxis G, Delaglio F, Bax A (2005) Molecular fragment replacement approach to protein structure determination by chemical shift and dipolar homology database mining. Methods Enzymol 394:42–78

Korukottu J, Bayrhuber M, Montaville P, Vijayan V, Jung Y, Becker S, Zweckstetter M (2007) Fast high-resolution protein structure determination by using unassigned NMR data. Angew Chem Int Ed Engl 46:1176–1179

Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE (2010) A transient and low-populated protein-folding intermediate at atomic resolution. Science 329:1312–1316

Kupce E, Freeman R (1993) Polychromatic selective pulses. J Magn Reson Series A 102:122–126

Legler PM, Cai M, Peterkofsky A, Clore GM (2004) Three-dimensional solution structure of the cytoplasmic B domain of the mannitol transporter IImannitol of the Escherichia coli phosphotransferase system. J Biol Chem 279:39115–39121

Lescop E, Brutscher B (2009) Highly automated protein backbone resonance assignment within a few hours: the "BATCH" strategy and software package. J Biomol NMR 44:43–57

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187:163–169

Lescop E, Kern T, Brutscher B (2010) Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H–15 N correlation experiments. J Magn Reson 203:190–198

Montalvao RW, Cavalli A, Salvatella X, Blundell TL, Vendruscolo M (2008) Structure determination of protein–protein complexes using NMR chemical shifts: case of an endonuclease colicin-immunity protein complex. J Am Chem Soc 130:15990–15996

Ortega-Roldan JL, Jensen MR, Brutscher B, Azuaga AI, Blackledge M, van Nuland NAJ (2009) Accurate characterization of weak macromolecular interactions by titration of NMR residual dipolar couplings: application to the CD2AP SH3-C:ubiquitin complex. Nucleic Acids Res 37:e70

Ottiger M, Bax A (1998) Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. J Biomol NMR 12:361–372

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104:3519–3540

Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D (2010) NMR structure determination for larger proteins using backbone-only data. Science 327:1014–1018

Rohl CA, Baker D (2002) De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. J Am Chem Soc 124:2723–2729

Rohl CA, Strauss CEM, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383:66–93

Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc 122:7793–7797

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127: 8014–8015

Schanda P, Van Melckebeke H, Brutscher B (2006a) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043

Schanda P, Forge V, Brutscher B (2006b) HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. Magn Reson Chem 44(Spec No):S177–S184

Sgourakis NG, Lange OF, Dimaio F, André I, Fitzkee NC, Rossi P, Montelione GT, Bax A, Baker D (2011) Determination of the structures of symmetric protein oligomers from NMR Chemical shifts and residual dipolar couplings. J Am Chem Soc 133:6288–6298

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690

Shen Y, Bryan PN, He Y, Orban J, Baker D, Bax A (2010) De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. Protein Sci USA 19: 349–356

Smith MA, Hu H, Shaka AJ (2001) Improved Broadband Inversion Performance for NMR in Liquids. J Magn Reson 151:269–283

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114

Valafar H, Prestegard JH (2003) Rapid classification of a protein fold family using a statistical analysis of dipolar couplings. Bioinformatics 19:1549–1555

Wang X, Tash B, Flanagan JM, Tian F (2011) RDC derived protein backbone resonance assignment using fragment assembly. J Biomol NMR 49:85–98

Yang SW, Chen H, Yang J, Machida S, Chua N, Yuan YA (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. Structure 18: 594–605

Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 122:3791–3792