Molecular Physics, 2013
Vol. 111, No. 3, 437–454, http://dx.doi.org/10.1080/00268976.2012.728257

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Multiple-state ensemble structure determination from eNOE spectroscopy

Beat Vögeli[a], Peter Güntert[bc] and Roland Riek[ad]*

[a]*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland;* [b]*Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, and Frankfurt Institute for Advanced Studies, J.W. Goethe-Universität, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany;* [c]*Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan;* [d]*The Salk Institute, 10010N. Torrey Pines Road, La Jolla, CA, 92037, USA*

(*Received 10 July 2012; final version received 31 August 2012*)

The recent investigation of an ensemble-based structure calculation using mainly a large collection of exact nuclear Overhauser enhancements (eNOEs) revealed the presence of concerted motion within the protein GB3 (B. Vögeli, S. Kazemi, P. Güntert and R. Riek, Nat. Struct. Mol. Biol., in press). Here, we discuss the method used in this study in detail. Important steps include the NMR pulse sequence, the determination of the eNOEs corrected for spin diffusion, the conversion of eNOE rates into distances, the distance-restraint classification, the use of bundling restraints to generate a compact representation of the structure and the selection of the appropriate ensemble to represent the structure. It is further demonstrated that eNOEs can be obtained between most proton types in a macromolecule. These eNOEs are then used to calculate an ensemble-based structure using CYANA that is capable to reveal long-range concerted motion in the protein. The structure ensembles are cross-validated with jackknife tests applied to the eNOEs, RDCs, scalar couplings, cross-correlated relaxation rates, and with a high-resolution structure obtained independently from X-ray diffraction and refined with RDCs.

Keywords: NMR; biological macromolecules; proteins; dynamics; correlated dynamics; NOE; nuclear Overhauser effect; structure calculation; structure ensemble

1. Introduction

Standard structure determinations by nuclear magnetic resonance (NMR) spectroscopy make use of a large number of experimentally readily accessible Nuclear Overhauser Enhancement (NOE) rates – typically up to 20 per residue in small proteins [1–3]. Since the NOE rate is proportional to the inverse 6th power of the distance between the two dipolar interacting spins [4], the strength of the NOEs lies in the supply of a large amount of through-space distance restraints. In general, these rates are employed in a semi-quantitative manner at most because it is recognized that the measurement of NOEs is flawed by mobility, spin diffusion, low signal-to-noise ratio, and technical limitations [5–18]. We have recently demonstrated that it is possible to obtain accurate amide proton/amide proton NOEs both in deuterated [19] and protonated protein samples [20]. To this purpose we measured NOE buildups as a function of the NOESY mixing time and converted the NOEs into precise distances for detailed structural studies of proteins. For example, distances up to 5 Å obtained from a perdeuterated ubiquitin sample have an experimental random error of only ≈0.07 Å. It is smaller than the pairwise rms deviations from distances extracted from high-resolution NMR or X-ray structures (0.24 Å) or than the pairwise rmsd between those (0.15 Å; pdb codes: 1D3Z and 1UBQ). In the following, we refer to these 'exact' NOEs as 'eNOEs'. Since the eNOE is a time- and ensemble-averaged observable, it contains both structural and dynamical information [21]. The collection of potentially thousands of eNOEs throughout a bio-macromolecule may serve as an excellent probe towards a more complete representation of both its structure and dynamics. For this purpose the eNOE offers advantages over other NMR probes. It is straightforward to measure eNOEs on a standard NMR sample because they do not require additional treatment of the sample, such as the addition of alignment media in the case of residual dipolar couplings (RDCs) [22], or paramagnetic labels for paramagnetic relaxation enhancement (PRE) [23] and pseudocontact shift measurements [24]. Assignments can be performed with the same well-established methods as for conventional NOEs [25]. In addition, the NOE is among the few observables that are measurable even for high molecular weight systems

*Corresponding author. Email: roland.riek@phys.chem.ethz.ch

such as large proteins, protein complexes, or membrane proteins substituted in membrane-mimicking environments [26].

The following work must be regarded as a detailed and extensive elaboration of a previously published paper (B. Vögeli, S. Kazemi, P. Güntert and R. Riek, Nat. Struct. Mol. Biol., DOI: 10.1038/nsmb.2355) on the use of eNOEs obtained for all proton types in a macromolecule for the multiple-state structure determination for the spatial elucidation of motion in proteins. In a first part, we assess the quality of the eNOEs. We demonstrate the close coincidence between distances extracted from eNOEs and those calculated from a high-resolution structure of the third immunoglobulin binding domain of protein G (GB3) [27–29]. In the second part, we establish a new CYANA protocol [30] which calculates multiple-state ensembles of structures in which the conformational restraints are required to be fulfilled on *average* over all members of the ensemble rather than simultaneously for each individual conformer [31–45]. The target function is substantially lowered upon using two and three states rather than one, and reaches a plateau for higher states. Cross-validation reveals that the effect is due to self-consistency of the data. We also show that a realistic spatial sampling is generated.

2. Experimental section

2.1. *Sample expression and purification*

GB3 was expressed and purified as described previously [45]. The uniformly ^{13}C,^{15}N-labeled NMR sample contained 350 µl of 4 mM protein solution in 97% H_2O, 3% D_2O, 50 mM potassium phosphate buffer, pH 6.5 and 0.5 mg/ml sodium azide.

2.2. *NMR spectroscopy*

All experiments were performed on a Bruker 700 MHz spectrometer equipped with a triple resonance cryoprobe at 298 K.

A series of 3D [^{15}N,^{13}C]-resolved [^{1}H,^{1}H]-NOESY spectra was recorded for the measurement of NOE buildups (pulse sequence see Figure 1). After indirect proton chemical shift evolution and [^{1}H,^{1}H]-NOE mixing (mixing time τ_{mix}), simultaneous [^{15}N,^{1}H]-HSQC and [^{13}C,^{1}H]-HSQC elements were employed [46,47]. A spectrum with $\tau_{mix} = 100$ ms was used for the resonance assignment. Diagonal peak decay and cross-peak buildup were analyzed from the spectra measured with mixing times $\tau_{mix} = 20$, 30, 40, 50, and 60 ms. The spectra were recorded with $200(t_1) \times 40(t_2) \times 1024(t_3)$ complex points, maximal evolution times $t_{1max,1H} = 22.0$ ms, $t_{2max,15N} = 14.4$ ms, $t_{2max,13C} = 7.6$ ms, and $t_{3max,1H} = 102.4$ ms, spectral widths $SW_{1,1H} = 13.0$ ppm, $SW_{2,15N} = 39.7$ ppm, $SW_{2,13C} = 30.0$ ppm, $SW_{3,1H} = 14.3$ ppm, an interscan delay of 0.6 s, and 4 scans per increment resulting in a measurement time of 1 day per spectrum. The time-domain data were multiplied with a squared cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to $1024 \times 128 \times 512$ complex points.

All spectra were processed and analyzed using the software packages NMRPipe [48] and NMRDraw [49]. Peak heights were determined by parabolic interpolation.

The diastereotopic assignment of the methyl groups in Leu and Val residues was achieved with a fractional 10%-^{13}C labeled sample following the method proposed in reference [50]. Stereospecific assignment of the methylene groups was based on a 'floating assignment' method searching for the lowest violations. The floating approach correctly assigned all of the seven stereo-specific assignments of methylene groups determined by scalar couplings [70] that have been reproduced by others (Figure 7, [69]). Furthermore, the specific assignment of the other methylene groups are in line with the high-resolution x-ray structure. Details on the method will be published elsewhere.

R_1 and $R_1\rho$ relaxation rates of the backbone ^{15}N nuclear spins were measured using standard pulse sequences at 298 K [3]. R_2 was calculated from R_1 and $R_1\rho$, and an effective global correlation time τ_c of 4.15 ns was obtained from the ratio R_2/R_1 under assumption of isotropic overall tumbling [51] and was used to scale the anisotropic diffusion tensor [52].

$^{3}J_{HN\alpha}$, $^{3}J_{HNC'}$, and $^{3}J_{HNC\beta}$ scalar couplings and ^{15}N-$^{1}H^{N}$ residual dipolar couplings (RDCs) were taken from references [53] and [45]. $^{13}C_\alpha$-$^{1}H_\alpha$ RDCs were extracted from a previously recorded 3D HNCA[HA] E.COSY experiment [54] initially used for extracting $^{1}H^{N}$-$^{1}H_\alpha$ RDCs [45].

In this study, the exact positions of the protons are of particular importance for cross validation of the NOE data. In X-ray structures, protons can only be added at idealized positions. We have previously used large sets of RDCs to determine exact proton positions in the backbone of a high-accuracy X-ray structure [28,29]. In cross-validation checks, we obtained extremely high agreement with $^{3}J_{HNH\alpha}$ scalar couplings and intraresidual and sequential $D_{HNH\alpha}$ RDCs (rmsd values between predicted and experimental values of 0.32 Hz for $^{3}J_{HNH\alpha}$ [53], 1.15 Hz for intraresidual and 1.66 Hz for sequential $D_{HNH\alpha}$ [45]) as reported in

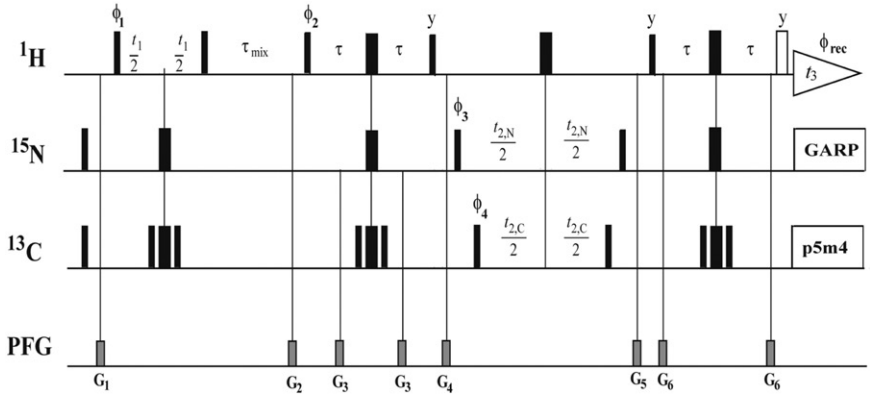

Figure 1. Pulse sequence of the 3D [^{15}N,^{13}C]-resolved [^{1}H,^{1}H]-NOESY experiments for the measurement of NOE buildups. The radio-frequency pulses on ^{1}H, ^{15}N and ^{13}C are applied at 4.7, 116 and 47.7 ppm, respectively. Narrow and wide black bars indicate non-selective 90° and 180° pulses. The white bar indicates a trim pulse of 1 ms duration [75]. Vertical lines connect centered pulses. ^{15}N-decoupling is achieved with GARP [76] and ^{13}C-decoupling with a p5m4 supercycle [77,78] consisting of CHIRP pulses [79]. The delay $\tau = 1.7$ ms is optimized for ^{1}H–^{13}C transfers. Alternatively, τ can be set to 2.7 ms and the ^{13}C 180° pulses during the transfers are shifted by 1.0 ms to optimize both the ^{1}H–^{15}N and ^{1}H–^{13}C transfers. Unless indicated otherwise, all radio-frequency pulses are applied with phase x. The phase cycle is: $\varphi_1 = \{x, x, x, x, -x, -x, -x, -x\}$; $\varphi_2 = \{x, x, -x, -x\}$; $\varphi_3 = \{x, -x\}$; $\varphi_4 = \{x, -x\}$; $\varphi_{rec} = \{-x, x, x, -x, x, -x, -x, x\}$. Pulsed field gradients indicated on the line marked PFG are applied along the z-axis with duration/strength of G$_1$, 1000 µs/47.71 G/cm; G$_2$, 1000 µs/47.71 G/cm; G$_3$, 1000 µs/30 G/cm; G$_4$, 1000 µs/30 G/cm; G$_5$, 1000 µs/39.6 G/cm; G$_6$, 1000 µs/39.6 G/cm. Quadrature detection in the ^{1}H(t_1) and simultaneous ^{15}N/^{13}C(t_2) dimensions [46,47] is achieved by the States-TPPI method [80] applied to the phases φ_1, φ_3, φ_4 and φ_{rec}.

reference [55]. The bond lengths were scaled to 1.02 and 1.09 Å, respectively.

2.3. *Buildup fitting*

The NOESY autorelaxation rate of spin X, ρ_X, is obtained from fits to the normalized diagonal-peak decay [19]:

$$T_{XX}^{NOE}(\tau_{mix}) = e^{-\rho_X \tau_{mix}} \quad (1)$$

Then, the cross-relaxation rate between spins K and L, σ_{KL}, is obtained by inserting ρ_K and ρ_L into the equation describing the normalized cross-peak buildup [19]:

$$T_{KL}^{NOE}(\tau_{mix}) = -\frac{\sigma_{KL}}{(\lambda_+ - \lambda_-)}\left[e^{-\lambda_- \tau_{mix}} - e^{-\lambda_+ \tau_{mix}}\right] \quad (2)$$

with

$$\lambda_\pm = \frac{(\rho_K + \rho_L)}{2} \pm \sqrt{\left(\frac{\rho_K - \rho_L}{2}\right)^2 + \sigma_{KL}^2} \quad (3)$$

The left-hand sides of Equations (1) and (2) were obtained from the spectral peak heights I.

$$T_{KK}^{NOE}(\tau_{mix}) = \frac{I_{KK}(\tau_{mix})}{I_{KK}(0)};\ T_{LL}^{NOE}(\tau_{mix}) = \frac{I_{LL}(\tau_{mix})}{I_{LL}(0)} \quad (4)$$

$$T_{KL}^{NOE}(\tau_{mix}) = \frac{I_{KL}(\tau_{mix})}{I_{KK}(0)} \quad (5)$$

Because the magnetization of the cross-peaks and diagonal peaks relax differently during the HSQC element, Equation (5) is not strictly valid. If both cross-peaks can be evaluated, the following equation correctly provides the transfer function [19]:

$$T_{KL}^{NOE}(\tau_{mix}) = T_{LK}^{NOE}(\tau_{mix}) = \sqrt{\frac{I_{KL}(\tau_{mix}) I_{LK} I(\tau_{mix})}{I_{KK}(0) I_{LL}(0)}} \quad (6)$$

For simplicity, Equation (5) was used to obtain the effective cross-relaxation rates σ_{KL}^{eff} and σ_{LK}^{eff}, and the true cross-relaxation rate was approximated as:

$$\sigma_{LK} = \sigma_{KL} = \sqrt{\sigma_{LK}^{eff} \sigma_{KL}^{eff}} \quad (7)$$

A quality factor is introduced to judge the agreement of the cross-peak buildup and the two-spin description provided in Equation (2) after taking into account the contribution from spin diffusion. Two main contributions to deviations must be considered: First, the validity of the assumptions made in Equation (2) (this depends, for example, on the extent of spin diffusion or possible peak overlap); second, the signal-to-noise ratio. Thus, the quality factor is defined as follows:

$$\chi = \frac{1}{\max|I_{exp}|}\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[I_{exp}(\tau_{mix}(i)) - I_{fit}(\tau_{mix}(i))\right]^2} \quad (8)$$

where N is the number of mixing times, and I_{exp} and I_{fit} are the experimental and back-predicted peak intensities, respectively (note, other definitions may be reasonable as well). This quality factor was in close agreement with benchmarking by human inspection of the fits. Generally, 'good fits' are obtained for $\chi < 0.08$ and 'bad fits' for $\chi > 0.15$. Only cross-peaks which showed no overlap were evaluated.

The autorelaxation rates ρ and cross-relaxation rates σ obtained for GB3 are presented in Tables S1 and S2 in the Supporting Information. The distribution of experimentally determined distances is shown in Figure 2 (including H^N-H^N eNOEs from previous measurements [19,20] and 61 aromatic NOEs, which were translated qualitatively into distances by the conventional approach because of the presence of aromatic ring flips).

The cross-relaxation rate was corrected for a calculated spin-diffusion contribution using a previously determined structure (i.e. X-ray or NMR structure) or a structure determined by the conventional structure calculation approach (see below) [19,20]. A detailed discussion of the procedure is presented in reference [56]. In short, buildup curves were simulated assuming both the two-spin system KL and all three-spin systems KLM for which M is located within spheres of 6 Å radius centered either at K or L. Here, the simulations are based on the cross-relaxation rates calculated from the distances in the RDC-refined X-ray structure (pdb code 2OED) [27] assuming isotropic molecular tumbling. Subsequently, the diffusion-affected intensities were simulated by adding the individual intensity correction of each KLM system to the intensities obtained for the KL system. Fitting Equation (2) to the predicted intensities at the time points τ_m yields apparent cross-relaxation rates. The percentage difference between the the apparent rate and the rate expected for the two-spin approximation was used for correcting the experimental rates.

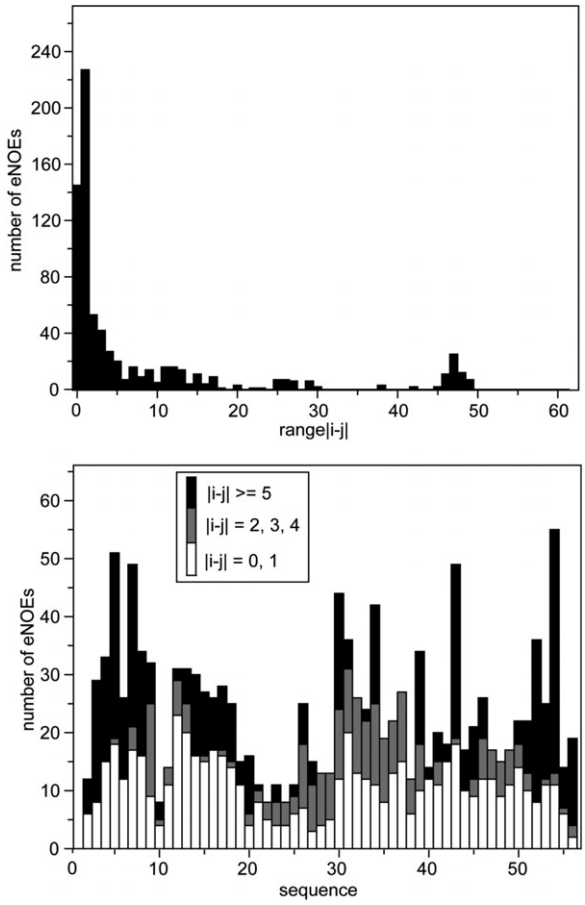

Figure 2. Sequence-resolved eNOE-derived distance restraints for GB3. Top panel: Number of eNOEs versus range $|i-j|$, where i and j are the residue numbers of the involved spins. Bottom panel: Number of eNOEs versus residue number. The counts are classified as short-range (white), medium-range (gray), and long-range (black).

2.4. *Distance calculation*

The cross-relaxation rates were converted into effective distances, which absorb all motional effects, following Equation (9) [4]:

$$r_{KL}^{eff} = \sqrt[6]{\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma^4 h^2}{40\pi^2} \frac{\tau_c}{\sigma_{KL}}} \qquad (9)$$

All the constants have the usual meaning. The molecular tumbling is assumed to be isotropic with a correlation time τ_c. This simplifies the conversion dramatically and introduces an error of maximally 3% in the case of GB3 but much less for most distances [20]. Equation (9) yields the effective distance without the need to determine a scaling factor or 'calibration constant' as in the traditional approach. For the structure calculation discussed below, r_{KL}^{eff} was then translated into an upper and lower distance restraint. If both magnetization pathways were analyzed following Equation (7), the upper and lower distance bound were both set equal to r_{KL}^{eff}. If the NOE rate could be determined from one pathway only, the lower and upper distance bounds were set to $0.85\, r_{KL}^{eff}$ and $1.15\, r_{KL}^{eff}$, respectively.

Equation (9) is exact for motion slower than the tumbling time of the molecule. For fast motion, a double sum over two inverse cubes of r must be used and the expression is further modulated by angular expressions. However, it has been shown that the

angular and distal effects counteract in principle [57,58]. Furthermore, simulations have revealed that the impact of fast motion can be approximately neglected for ^{1}H-^{1}H spin pairs in the protein backbone as long as the local H-X order parameters are larger than 0.5 [59], which is usually fulfilled in globular proteins except for highly flexible tails and loops. Similar conclusions have been drawn also for NOEs involved in side chain based on molecular dynamics studies, but they also reveal that a few NOEs may violate the assumption considerably [57,58]. The spins involved in these critical NOEs are usually located at the far end of very long side chains such as Lys and Arg. In this study, we did not obtain good quality buildup fits for these NOEs and thus did not generate corresponding exact distance restraints.

2.4.1. *NOEs involving methyl groups*

NOEs that involve methyl groups needed an additional adjustment. The three magnetically equivalent protons in a methyl group were treated as one pseudo atom located at the center of mass. Therefore, the experimental distance underestimates the true distance. On the other hand, fast methyl rotation adds an angular dependency to the NOE which lets the extracted distance appear larger. To account for these effects [2,60,61], the lower distance bound was set to $0.8\, r_{KL}^{eff} - 1$ Å and the upper distance bound to $1.2 r_{KL}^{eff}$, respectively.

2.4.2. *NOEs involving aromatics*

NOEs that involve pairs of equivalent protons in aromatic rings (i.e. $^1H^{\delta}$ and $^1H^{\varepsilon}$ in Phe and Tyr) were translated into upper distance bounds of 8 Å following the conventional approach. Distances obtained from eNOEs involving nonequivalent protons (such as $1H^{\zeta}$ protons) in aromatics were treated as eNOEs. However, because of offset effects of the ^{13}C pulses, the concomitant loss in magnetization was corrected for accordingly if only one pathway (i.e. one cross peak) could be evaluated. It is noteworthy that the full potential of NOEs involving a group of slowly exchanging equivalent spins could be exploited with direct structural refinement over the distances between all individual atoms rather than those between the pseudoatoms as this approach puts narrower restraints on these groups [62].

2.5. *Structure calculation*

First, a conventional structure calculation was performed. In the absence of a previously determined structure (such as an X-ray structure), the conventional structure may be used to correct the eNOEs for spin-diffusion effects (see above). Second, an ensemble-based structure calculation was performed that results in a realistic representation of the experimental restraints.

2.5.1. *Conventional structure calculation*

For the conventional structure calculation, 1953 upper distance limits calculated from the cross-peak intensities of the NOESY with a mixing time of $\tau_{mix} = 100$ ms were obtained. Of these, 1041 constituted meaningful restraints, the others being redundant or trivially fulfilled by the covalent structure [63]. Additional experimental restraints were $^3J_{HN\alpha}$, $^3J_{HNC'}$, and $^3J_{HNC\beta}$ scalar couplings, ^{15}N-$^1H^N$ and $^{13}C_{\alpha}$-$^1H_{\alpha}$ RDCs, and angular restraints from $^{13}C_{\alpha}$ chemical shifts. Using these restraints a standard structure calculation was performed with the software package CYANA [30] starting with 100 randomized conformers. Simulated annealing with 50,000 torsion angle dynamics steps was applied [64], and the 20 conformers with the lowest final target function values were analyzed. The 20 conformers with the lowest target function values reflecting the smallest experimental restraint violations and van-der-Waals violations were used to represent the calculated structure. The small r.m.s. deviations from the mean structure of 0.95 Å for the backbone atoms and 1.38 Å for all heavy atoms, respectively, indicate good convergence of the structure calculation (vide infra).

2.5.2. *Multiple-state ensemble structure calculation*

For the multiple-state ensemble calculation, 811 effective distances were supplemented by previously published amide-amide proton distances (averaged over all data sets, see Table S5 in the Supporting Information) [20] yielding a total of 823 effective distances (i.e. 324 from both pathways/multiple H^N-H^N data sets, and 499 from one pathway or pairs involving methyl groups). In addition, 61 NOEs involving one or two aromatic pseudo-atoms were used in a conventional manner by providing upper limits of 8 Å. As for the conventional structure calculation protocol 100 conformers were calculated and the 20 conformers with the lowest target function were then used to represent the calculated structure. Ensembles encompassing simultaneously 1 to 9 states of the entire protein were calculated simultaneously, using the same number of initial conformers and the same simulated annealing schedule as for the conventional structure calculation. Steric repulsion between atoms of different states was

excluded, and the eNOE distance restraints were applied to the $1/r^6$ averages of the corresponding distances in the individual states. Similarly, the ^3J coupling restraints and the RDC restraints were applied to the arithmetic mean of the corresponding quantities in the individual states. $^{13}C_\alpha$ chemical shift derived angular restraints were used in addition. Bundling restraints were applied in order to keep the individual structural states together in space as far as permitted by the experimental restraints [40]. To this end weak upper distance bounds of 1.2 Å were applied to all distances between the same nitrogen and carbon atoms in different states. The weight of these bundling restraints was 100 times lower than for NOE upper distance bounds, except for the backbone atoms N, C^α, C', and C^β, for which a 10 times lower weight than for NOEs was used. The 20 three-state conformers with the lowest target function were deposited in the protein data bank (PDB code 2lum, BMRB ID 18 531).

For the single-state structure the r.m.s. deviations to a RDC-refined X-ray structure were 0.57 Å and 1.17 Å for the backbone or all heavy atoms, respectively [27–29] (vide infra).

3. Results

3.1. *Validation of experimental cross-relaxation rates and distances*

Overall, 1092 buildup could be fitted (typically satisfying the criterion $\chi < 0.15$ with χ given in Equation (8)). 562 of them constitute pairs of both transfer pathways of a specific spin system resulting in $562/2 = 281$ exact cross-relaxation rates. The remaining 530 do not have the corresponding counterpart and resulted in 530 less exact relaxation rates.

The top left panel of Figure S1 in the Supporting Information shows a correlation plot between H^N-H^N cross-relaxation rates obtained in this study and rates determined in previous studies with different mixing times, pulse sequences, and isotope labeling [19,20]. The correlation is excellent with all data sets ($r > 0.99$). For the structure calculation, H^N-H^N distances were calculated from each set and an average overall set was used which has a residue-averaged random error of only 0.06 Å (see table S4 in the Supporting Information) moving 31 NOEs obtained from a single cross peak to the group with 0% error tolerance and yielding an additional 12 eNOEs (in total, 324 from both pathways/multiple H^N-H^N data sets, and 499 from one pathway or pairs involving methyl groups). Figure 3 shows correlation plots between cross-relaxation rates obtained from the eNOE buildups and predicted from the RDC-refined 1.1 Å X-ray structure (PDB code 2OED) [27] with optimized proton positions [28,29]. Figure S2 in the Supporting Information shows plots of small subsets of these cross-relaxation rates. Rates for which both transfer pathways could be evaluated were obtained from equation 7 (red dots in Figure 3). They are in better agreement with the RDC-refined 1.1 Å X-ray structure than those obtained from one pathway only (blue dots). A heuristic order parameter, defined as $S^2_{KL} = \sigma^{eff}_{KL}/\sigma^{theo}_{KL}$, is 0.75 on average as estimated from the slope of a linear regression. A S^2_{KL} value smaller than 1 indicates a dominant influence from angular fluctuations which are faster than the molecular tumbling [57]. In the backbone it is 0.84 and decreases for spins located further out in the side chains (see Table 1 and Tables S3 and S4 in the Supporting Information). Pearson's correlation coefficient is overall 0.93. The backbone yields the highest value (0.97) and the value decreases for the side chains. This observed trend may be attributed to two effects: First, the mobility generally increases with larger separation of the involved protons from the main chain [65,66]. Second, the X-ray structure is RDC-optimized in the backbone but not in the side chains.

Next, the cross-relaxation rates were converted into effective distances under the assumption of isotropic molecular tumbling following Equation (9). The determined effective distances absorb all motional effects. Figure 4 and Figure S3 and S4 in the Supporting Information show correlation plots between the experimental effective and predicted distances. Statistics are listed in Table 1 and Table S3 in the Supporting Information. The slope of the linear regression including all distances is 0.97, whereas it is 0.99 for backbone distances only. The slope drops as the protons are located further out in the side chains which is attributed to motion, as stated already above. Pearson's correlation coefficient for all distances for which both pathways could be evaluated is 0.92, while for the backbone only it is 0.98. Inclusion of distances calculated from only one cross peak lowers it to 0.84 for all distances and 0.96 for the backbone. It can be concluded that the correlation between the experimental and true distances is very good. Taking into account that the reference structure was determined under different conditions and necessarily introduces an additional error from the shortcomings of a single conformer representation, the experimental eNOE data are of extraordinarily high precision. Previously we estimated the experimental error to be on the order of 2% for H^N-H^N NOEs measured on a perdeuterated sample [19] and similar for a protonated sample [20]. Here, we extended the NOEs measurements to the side chains. Since the quality factors χ for NOEs for which

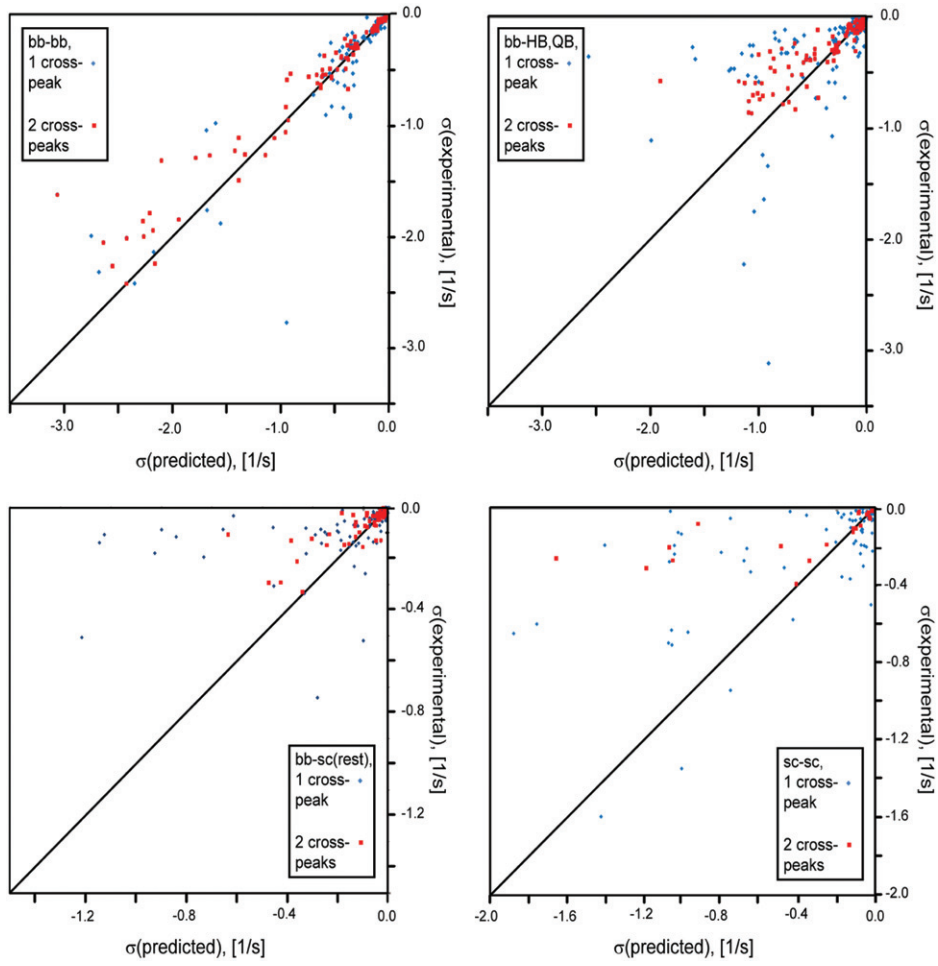

Figure 3. Experimental cross-relaxation rates of GB3 compared with those predicted from the RDC-optimized X-ray structure. The panels show the rates grouped into spin pairs consisting of two spins located in the backbone (top left), one in the backbone and one being a side-chain HB/QB (top right), one in the backbone and one being another side-chain proton (bottom left), and both in the side chain (bottom right) as indicated in the black boxes. Rates for which one or both transfer pathways could be evaluated are shown in blue diamonds or red squares, respectively. Predictions are based on the crystal structure (PDB code 2OED) with RDC-optimized proton positions [27–29]. The diffusion tensor was taken from reference 52 and calculated with formulae 12, 13 and 18 in reference 20. Corrections to the experimental rates due to the presence of spin diffusion were calculated from the 2OED structure with the MATLAB program DOMINO as described in references 19 and 20. Black lines indicate slope 1.

Table 1. Slopes s and Pearson's correlation coefficients r of experimental versus predicted cross-relaxation rates σ_{AB} and internuclear distances r_{AB} grouped into backbone and side-chain spins.

atom A	atom B	s (σ_{AB})[a]	r (σ_{AB})[a]	s (r_{AB})[a]	r (r_{AB})[a]	s(r_{AB})[b]	r (r_{AB})[b]	# pairs	# pairs[b]
all	all	0.75	0.93	0.97	0.92	0.96	0.84	295	832
bb	bb	0.84	0.97	0.99	0.98	0.99	0.96	112	252
bb	HB,HB2, HB3,QB	0.63	0.84	0.98	0.92	0.98	0.86	107	269
bb	sc, rest	0.50	0.71	0.96	0.77	0.95	0.74	57	189
sc	sc	0.24	0.05	0.91	0.82	0.92	0.73	19	122

Notes: [a]Cross-relaxation rates obtained from both cross peaks.
[b]Cross-relaxation rates obtained from one cross peak if the second one could not be evaluated, otherwise from both cross peaks.

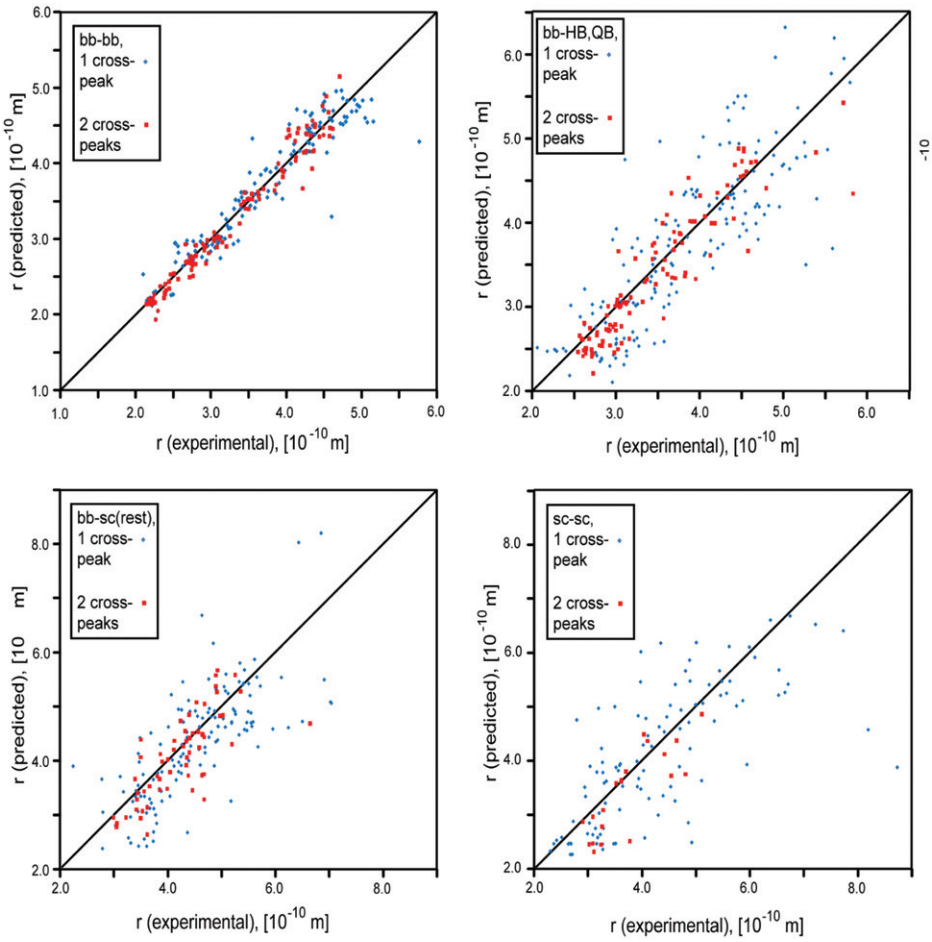

Figure 4. Predicted versus experimental distances in GB3. Experimental distances were calculated under the assumption of isotropic molecular tumbling with a correlation time $\tau_c = 4.15$ ns at 298 K following Equation (9). Predictions are based on the 2OED structure with RDC-optimized backbone proton positions [27–29]. The panels show the distances between (top left) two spins located in the backbone (bb); (top right) one in the backbone and one being a side-chain HB/QB; (bottom left) one in the backbone and one being another side-chain proton (sc, rest); and (bottom right) both in the side chain as indicated in the black boxes. Black lines indicate slope 1.

a side-chain proton was involved are comparable to those for backbone NOEs, the precision and accuracy of these distances are expected to be similar to those in the backbone. However, the correlation of the distances between spins in the side chains is not as good as those in the backbone. This apparent discrepancy is mainly attributed to the more pronounced side-chain motion which is evidently insufficiently represented by the single-state X-ray reference structure.

In order to take full advantage of the high precision of these effective distances, an ensemble-based structure calculation rather than a single-state structure calculation should be performed (vide infra). Distances obtained from both NOE pathways can then be incorporated in the structure calculation with no or very little error tolerance. Distances calculated from only one cross peak are less accurate and correlate less well with the refined X-ray structure than the distances determined via equation 7. A larger error tolerance is/must be anticipated for these less accurate eNOEs. The relative error tolerance can be estimated from the individual contributions to the two-pathway NOEs (σ_{KL}^{eff} and σ_{LK}^{eff} in Equation (7)). This comparison yields an error of 3% if only one peak is evaluated. However, the best peak quality is typically encountered for spin pairs for which both peaks can be evaluated. Therefore, we allowed for 15% to safely avoid overfitting. By inspection of these individual groups of eNOEs in Figure 4 with particular focus on the backbone NOEs and by the discussion given above, the following error tolerances were chosen: ±0%, ±15%, and ±20% for distances calculated from both

cross peaks, one cross peaks, and those involving two methyl groups, respectively. In addition, for distances involving a methyl group the lower limits were reduced by 1 Å per involved methyl group. The large tolerance is motivated by the fact that due to fast rotation of the methyl group a pseudo atom needs to be introduced and that NOE distances are generally underestimated [2,60,61] (see Figure S4 in the Supporting Information). For NOEs involving at least one aromatic proton H^{δ} or H^{ε} of Phe or Tyr an upper limit of 8 Å was used.

3.2. *Ensemble-based structure calculation protocol*

Ensemble-based structure calculations were performed with the software package CYANA assuming that an ensemble of X structural states is required to fulfill simultaneously the experimental distance, RDC and scalar coupling restraints. In parallel, van-der-Waals interactions and angular restraints derived from ^{13}C chemical shifts are also required to be fulfilled by each individual structural state. The squared differences between each eNOE-derived effective distance from Equation (9) and the corresponding distance obtained by r^{-6}-averaging over the states is minimized. RDC and scalar coupling constants are treated similarly by averaging them linearly over the states. Ensemble structure calculations were performed with the number of states ranging from 1 to 9 (Figures 5 and 6).

3.2.1. *Experimental restraints*

A total of 823 eNOEs from the ^{15}N,^{13}C-resolved NOESY experiments were collected, including previously published amide-amide proton distances (averaged over all data sets yielding a residue-averaged random error of 0.06 Å, see Table S4 in the Supporting Information) [20]. These include 324 distances determined either from both magnetization pathways (i.e. both cross-peaks) or from multiple H^N-H^N data sets, and 499 distances determined from either one magnetization pathway only (i.e. only one cross-peak could be evaluated) or between two methyl groups. Addition of 61 conventional NOE upper distance bounds involving equivalent aromatic ring protons yielded a total of 884 distance restraints. The input data was supplemented by 54 angular restraints derived from ^{13}C shifts, 147 restraints from three-bond scalar couplings $^3J_{HNH\alpha}$, $^3J_{HNC\beta}$ and $^3J_{HNC'}$ [53], and 90 residual dipolar couplings D_{HNN} [45] and $D_{H\alpha C\alpha}$ (unpublished). Note that all of these observables are sensitive to motion on the millisecond timescale (the 'slow' NMR timescale).

3.2.2. *Bundling restraints*

In addition to the experimental restraints, bundling restraints [40] had to be introduced in the ensemble-based structure calculation. Due to the $1/r^6$ dependency long distances contribute minimally to the NOE. This may result in an unphysically loose packing of the ensemble. Thus, the structural divergence among the states has to be limited. The rationale behind the bundling restraints is to generate the most uniform ensemble that is in agreement with the experimental data [40]. To achieve this 'bundling' upper distance bounds were applied to all distances between the same atoms in different states. The weight of these bundling restraints is chosen so low that they can easily be overridden by the experimental restraints. The selection and calibration of the bundling restraints within the software package CYANA have three components that can be adjusted. These are (i) the atoms for which a bundling restraint is generated, (ii) the atomic r.m.s. deviation enforced by the bundling restraint, and (iii) the weight of the bundling restraint in comparison to the experimental restraints. Here, all carbon and nitrogen atoms were selected to have a bundling restraint. The r.m.s.d. is chosen to be 1.2 Å such that it is ensured that the different χ_1 rotamer states of the side chains are within the r.m.s.d. without violating the bundling restraint. The symmetry energy term is a harmonic well with a flat bottom of length 1.2 Å.

To get an appropriate weight of the bundling restraints relative to the experimental restraints, several structure calculations were performed in which the weight of NOE distance restraints w was varied from 0 to 1. Within each of these calculations the weight for the side chain-atoms from C_γ outward were reduced tenfold relative to the backbone and C_β atoms. A comparison between these various structure calculations shows that the r.m.s.d. of the ensemble increases strongly with an increasing number of structural states if no bundling restraints are used (i.e. $w = 0$) (Figure 5). As expected, the violations of restraints represented by the target function are decreasing with the number of states (Figure 5). For the structure calculation with a weak weighting factor ($w = 0.001$) the r.m.s.d. is maximal for the two-state ensemble and asymptotically approaches $\approx$1 Å for more states. With a stronger weighting ($w > 0.01$) the r.m.s.d. is uniformly $\approx$0.5 Å for all bundles with more than one conformer and the experimental restraints are well satisfied as judged by the target function. Importantly, this confirms that the bundling restraints are overridden by any experimental input.

The influence of the weighting factor of the bundling restraints can also be gauged by a

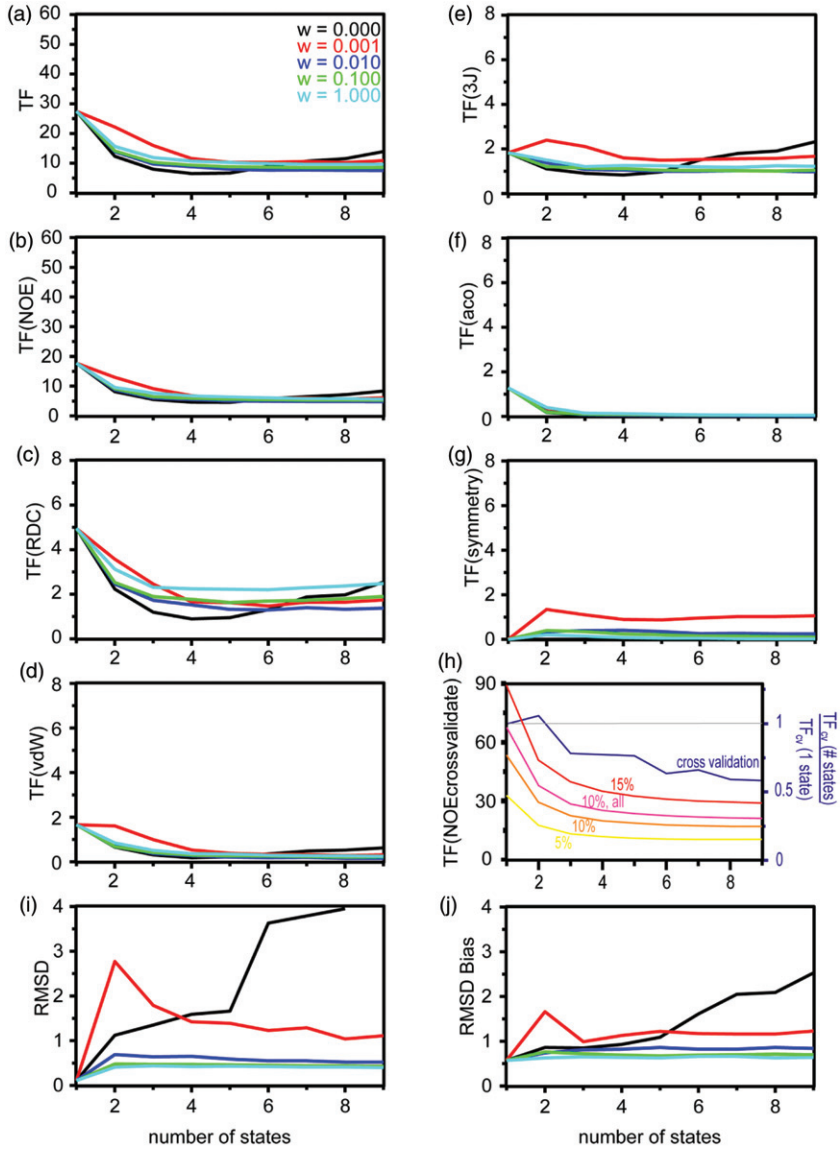

Figure 5. Target functions values (TF) obtained from structure calculations versus the number of simultaneously optimized conformers (a–h) and their r.m.s.d. values (i and j). The overall TF is shown in (a), and the contributions from eNOEs, RDCs, van-der-Waals terms, experimental data, *J* couplings, C_α angular restraints, and the bundling restraint terms are shown in (b) to (g). Cross-validation TF for three-state ensembles calculated with a bundling restraint weight 0.1 are shown in (h). Target functions (TF) obtained from a jackknife procedure are shown in blue, and upon random alteration of the distances obtained from both cross peaks by 5%, 10%, and 15% in yellow, orange, and red, respectively, and upon random alteration of all distances by 10% in pink. R.m.s. deviations within a bundle are plotted in (i) and to the RDC-refined X-ray structure ('Bias r.m.s.d.') in (j). The numeric values are taken from Table S6 in the Supporting Information.

comparison of the H^N–N order parameters calculated from the ensemble and those obtained from H^N–N RDCs [28,29]. The ensemble order parameters are stable for *w* between 1 and 0.01 (Figure 6). However, significantly lower values are obtained for $w = 0.001$. This finding is again in agreement with the ensemble r.m.s.d. values. The analysis of these data indicates that $w = 0.1$ is an appropriate weighting factor for the bundling restraints.

In summary, the use of the bundling restraints in the ensemble-based structure calculation results in a (most) convergent representation of a structural ensemble which is consistent with the experimental data.

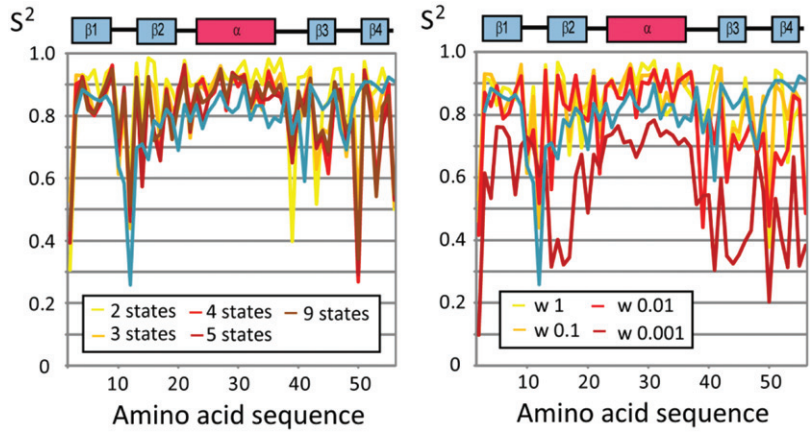

Figure 6. Backbone H^N-N order parameters versus the amino sequence of GB3. Left panel: Order parameters obtained for ensembles calculated with 2, 3, 4, 5, and 9 states and a bundling restraint weight of 0.1. Right panel: Order parameters obtained for three state ensembles with bundling restraint weights of 1, 0.1, 0.01, and 0.001. For comparison, order parameters obtained from RDCs are drawn in blue.

3.2.3. *Selection of the number of states*

The essence of the ensemble-based structure calculation is that the structure of the protein is represented by an ensemble of states rather than a single state. However, the number of states necessary to describe the experimental data is not known a priori. Therefore, an array of structure calculations with the number of states varying from 1 to 9 was performed (and w from 0 to 1). Among these ensembles the one with the minimal number of states that satisfies the experimental data well is regarded as the appropriate representative. This ensemble with X states is obtained if the three following criteria are fulfilled: (i) the target function drops significantly from state $X-1$ to X, (ii) by an increase of the number of states to $X+1$ the target function does not drop significantly anymore, and (iii) by an increase of the number of states to $X+1$ a jackknife error estimation does not produce a lower-than-random target function for omitted eNOEs. Following these criteria, it appears that the ensemble with $X=3$ states is an appropriate representation of the structure (Figure 5). In particular, the target function already levels out for more than three states.

3.2.4. *Multiple state bundle representation of the structure ensemble*

NMR structures are usually presented by a bundle of conformers that describe the precision of the structure determination. The conformers of the bundle are selected by their target function value. Usually, the 10–20 conformers with the lowest target function value from a set of 100 calculated conformers are selected. Here, we employed a similar procedure by calculating 100 three-state ensembles starting from random conformers. The 20 three-state ensembles with the lowest target function value were then selected to represent the ensemble structure of GB3 (note: in some of the figures a smaller number of ensembles are chosen for clarity). Comparison of the r.m.s. deviations of the states within an ensemble and over all 20 ensembles quantifies to which extent the sampling enforced by the experimental restraints coincides with the sampling due to lack of more experimental data. This is analogous to the r.m.s.d. or the precision of a conventional NMR structure bundle. Here, the r.m.s.d. within a single ensemble is 0–15% smaller than the overall r.m.s. deviation.

To generate (sub)bundles representing each of the three states, the r.m.s. deviations were calculated for each structure representing an individual state in all 20 three-state conformers and compared (20×3 individual-state structures). The selection criteria for the r.m.s.d. are adjustable and may depend on the scope. For example, if it is of interest to study the presence of concerted motion of the $\beta1$-strand, the mutual r.m.s. deviations of the C_α atoms located in $\beta1$ (i.e. residues 3–8) are calculated for the 20×3 individual-state structures. For a visual inspection, the (sub)bundles may be colored accordingly. In this way, concerted motion within the β-sheet is easily identified (for more details, see B. Vögeli, S. Kazemi, P. Güntert and R. Riek, Nat. Struct. Mol. Biol., in press). It is obvious that more sophisticated algorithms and approaches may be applicable to extract details from the structural ensemble.

3.2.5. *Cross-validation*

To validate the three-state ensemble representation of GB3, a comparison with independently obtained data and a check of the self-consistency of the input data was performed. The checks were undertaken in two ways (Figure 5h, Table S7 in the Supporting Information). First, the eNOE-derived distances obtained from both cross peaks were arbitrarily changed according to normal distributions with standard deviations of 5, 10, and 15%, and all eNOE-derived distances by 10%, respectively. The target function values of the corresponding structure calculations are considerably larger in all cases, for example by a factor of two in the case of 10% changes. In a second check, the eNOE data is cross-validated by a jackknife error estimation [67]. 10 structure calculations were run from which 10% of the input distance restraints were randomly omitted (each NOE is omitted in exactly one calculation). Then, an overall target function was calculated by summing over the target functions calculated for the omitted data (last column in Table S6 in the Supporting Information). This cross-validation target function tends to decrease with increasing number of states for all bundling restraint weighting factors except for 0. In Figure 5h, a curve is plotted for $w = 0.1$. The target function for the three-state ensemble is 40% lower than the one for the single-state structure. This outcome indicates that the three-state ensemble obtained with the reduced experimental data set is already close in structure to the one calculated with the entire data set. Therefore, the data set is at least in part over-determined.

As an independent check for the three-state ensemble of GB3, the angular spatial sampling of the H^N-N vectors can be quantified by order parameters. Good agreement between those calculated from the ensemble and order parameters derived from RDCs measured under six alignment conditions [28,29] was obtained (Figure 6). It appears that the minimal number of states previously determined to represent the data well is also the minimal number that yields satisfying order parameters.

Another independent cross-validation may be derived from a comparison between the obtained side-chain rotamer states and those in the 1.1 Å X-ray structure with PDB code 1IGD [68] (Figure 6). All the rotamer states in the three-state NMR structure have their counterpart in the X-ray structure with the exception of residues 11 and 47. Interestingly, an anisotropic refinement of the same X-ray data (pdb code 2IGD) is in agreement for residue 11. This reevaluation also resulted in a change of the rotamer states of residues 19 and 24, and finds additional rotamer states for residues 7, 15, 21 and 35. With the exception of residue 7, all these states are also sampled by the ensemble.

More insight into the rotamer states may be obtained from $^3J_{H\alpha H\beta 2,3}$, $^3J_{C'C\gamma}$ and $^3J_{NC\gamma}$ scalar couplings and residual dipolar couplings since they are collected under liquid-state NMR conditions [69–71]. The scalar couplings in reference 69 are not consistent with residues 8 and 52 in the ensemble or in the X-ray structure. On the other hand, the couplings are consistent with residue 47 in the ensemble, but not in the X-ray structure. Residues 8 and 35 appear to undergo rotamer averaging which is supported by RDC data [69]. While this is not the case in the ensemble for residue 8, residue 35 jumps between the two states as suggested by X-ray data. A second set of $^3J_{H\alpha H\beta 2,3}$ couplings [70] is neither consistent with the former scalar couplings nor the ensemble for residues 40, 45, and 46, while it is in contradiction to the former set consistent with the ensemble for residue 52.

The rotamer populations of $\chi 1$ angles of Val, Ile and Thr have also been determined from residual dipolar couplings (RDCs) [71]. H-C RDCs have been used to calculate populations where each rotamer state is allowed to have a systematic phase shift but not an oscillatory motion. Subsequently, these have been used to parameterize Karplus curves for $^3J_{C'\text{-}C\gamma}$ and $^3J_{N\text{-}C\gamma}$ scalar couplings. It must be noted that these calculations are hampered by the combined impacts of rotamer jumps and small-scale oscillatory motions on the RDCs and Karplus curves which can be only defined at a much lower accuracy than for atoms located in the backbone. In Figure 8, we classified the data by populations of $>80\%$, between 50 and 80%, and between 20 and 50% and compare these rotamer states with the calculated ensemble (see Table 2 in the Supporting Information of reference 71). The comparison between the rotamer states of the structural ensemble and the rotamer states predicted by the RDC approach shows high agreement. In detail, for residues 3, 6, 7, 17, 18, 33, 39, 44 and 52 the single predicted rotamer states are in agreement with the structural ensemble. For the rotamer state of residue 11, several rotamer states are predicted by the RDC approach, however with unphysical angles, while the structural ensemble shows some freedom around the $+60^\circ$ rotamer. For the rotamers of residue 21 the RDC approach predicts two rotamer states that are observed by the structural ensemble. For residue 25 the structural ensemble predicts the rotamer states with -60° and 180° angles, which is in line with the RDC data (the 180° rotamer state is however only predicted to be 9% and hence not indicated in Figure 8). Similarly, all the three rotamer states are observed for residue 42 in

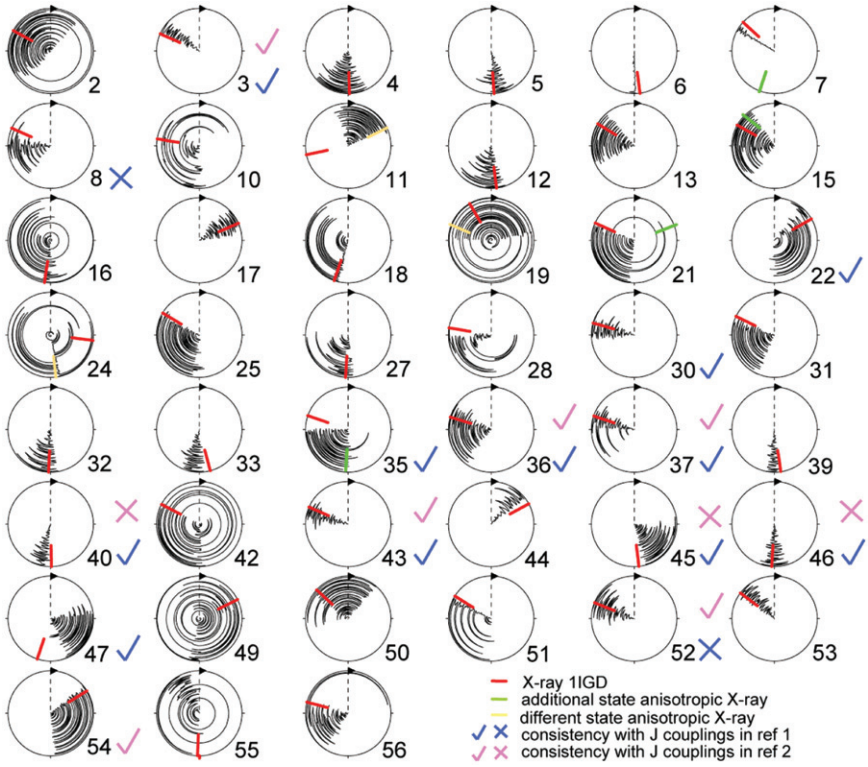

Figure 7. Circle diagrams of the χ^1 angles obtained from 20 three-state ensembles. The corresponding angles from the X-ray structure 1IGD are indicated in red. If the anisotropically refined X-ray structure 2IGD shows a second state in addition to the former it is indicated in green, and if it exhibits a single different state in yellow. Checks and crosses, respectively, indicate consistency with and contradiction to $^3J_{H\alpha H\beta 2,3}$ couplings presented in references 1 [69] and 2 [70].

the structural ensemble and also in the RDC measurements, but again two rotamer states are predicted to be populated only around 10%. For residues 49, 51, and 53 there are discrepancies between the two methods, while for residue 54 again the same two rotamer states are proposed by both methods. In summary, the rotamer states appear to be well represented by the structural ensemble. In addition, we measured $^3J_{C'\text{-}C\gamma}$ and $^3J_{N\text{-}C\gamma}$ scalar couplings of aromatic residues [72] and estimated the most populated rotamer states as listed in Table S8 in the Supporting Information assuming only staggered rotamers (idealized positions) [73]. We analyzed the data qualitatively. Much care has to be taken, as there is a relatively large experimental error and the parametrization of the Karplus curve is not trivial. Two residues, 33 and 45, occupy the 'trans' conformation, and four residues, 3, 30, 43 and 52, occupy a gauche conformation ($-60°$). All most populated states are in accordance with the ensemble. Residue 52 appears to have a small bias towards 'trans', which is not observed in the ensemble. On the other hand, residue 45, which has a bias towards gauche ($+60°$) does not yield a detectable peak at all in the corresponding spectrum for $^3J_{C'\text{-}C\gamma}$ identification which points to a close-to-zero population of gauche ($-60°$), and the $^3J_{N\text{-}C\gamma}$ coupling is somewhat smaller for residue 45 than for the other 'trans' rotamer residue. This may point to some residual population of gauche ($+60°$) in line with the structural ensemble, but the uncertainty does not allow for a clear-cut conclusion.

Finally, rotamer states can also be analyzed by cross-correlated relaxation (CCR) rates. Dipole/dipole CCR rates between H^N-N and C_β-C_γ are not consistent with the X-ray data for residues 33, 35, 40, 43 and 52 [74]. While part of the inconsistency is caused by the φ backbone angle as in residue 40, it is striking that residues 35 and 52 are again two of the controversial cases. Conclusively, there is an overall good agreement between the sampling in the ensemble and previously analyzed X-ray or NMR data. Some inconsistencies may be explained by different sample conditions such as crystalline/liquid state or different buffer conditions.

3.2.6. *Possible reduction of the experimental data*

Since the cross-validation procedures indicated that the experimental data set is in part over-determined,

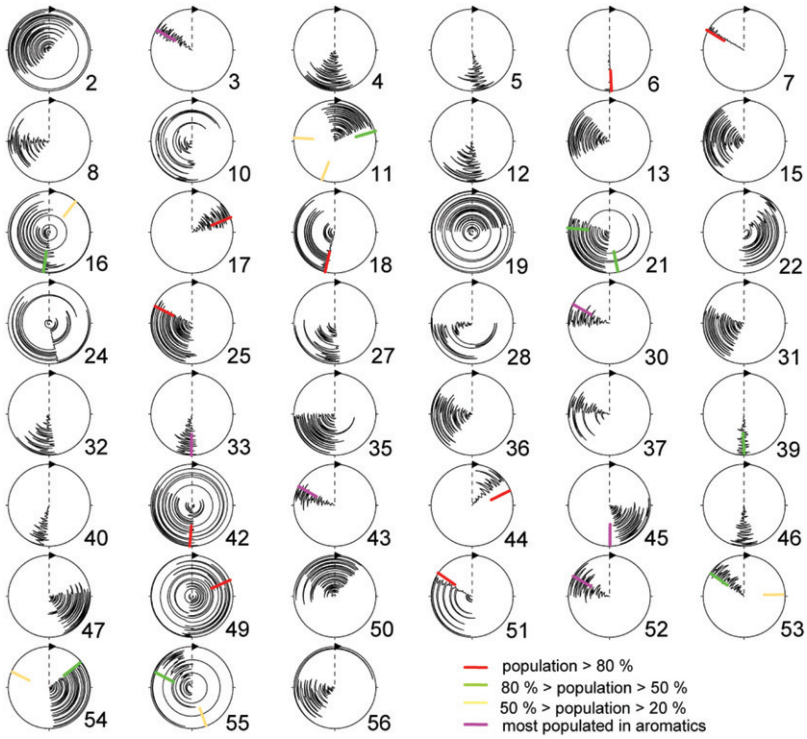

Figure 8. Circle diagrams of the χ^1 angles obtained from 20 three-state ensembles. The corresponding angles in accordance to their population obtained in an RDC study [71] are indicated as well. In addition, for aromatic residues, the most populated rotamer states are indicated as calculated from $^3J_{C'C\gamma}$ and $^3J_{NC\gamma}$ scalar couplings.

Table 2. rms deviations from reference structure 2OED with RDC-optimized proton positions.

Input data	# states	rmsd (bb heavy atoms) (res1-56) (Å)	rmsd (all heavy atoms) (res1-56) (Å)[a]	rmsd from reference (bb heavy atoms) (res3-55) (Å)	rmsd from reference (all heavy atoms) (res3-55) (Å)[a]
all	1	0.11	0.60	0.57	1.17
	3	0.47	0.86	0.72	1.31
no *J* coupling, RDCs	1	0.27	0.67	0.68	1.37
	3	0.49	0.86	0.79	1.39

Note: [a]2OED structure with RDC-optimized proton positions [27–29].

the question arises as to whether it is possible to reduce the experimental input by omitting the *J* coupling and the RDC data. From an experimental point of view, it would be particularly interesting to remove the RDC measurements as they require the preparation of another sample. To gain insight into this question, an ensemble-based structure calculation in absence of the *J* couplings and RDCs was performed (Table 2, Figure 9). This structure ensemble is very similar to its ensemble counterpart calculated with the additional data as demonstrated in Figure 9. However, the backbone r.m.s.d. increases considerably for the single-state. Interestingly, the r.m.s.d. is virtually identical for the three-state ensemble. Extraction of the concerted motion is less straight-forward, although it is still present (data not shown). Therefore, we recommend to measure a small set of *J* couplings and/or RDCs in addition to the eNOEs.

4. Discussion

4.1. *Summary of the present state*

A protocol is presented which enables the NMR structure determination of an ensemble of states by the measurements of many eNOEs as well as a small

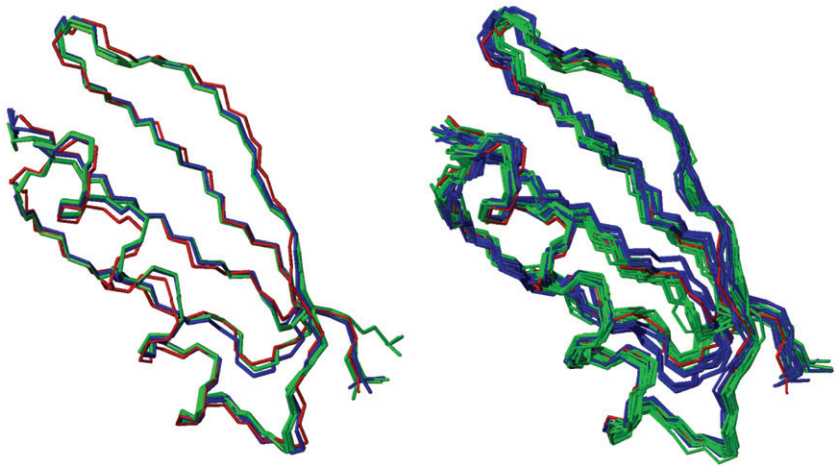

Figure 9. Backbone heavy atom representation of GB3. Structures obtained from eNOEs are compared with the high-accuracy reference structure 2OED with RDC-optimized proton positions shown in red [28–30]. The ensemble is calculated with and without J coupling and RDC data (blue and green). The single-state structures are shown in the left panel and three-state ensembles in the right panel, respectively.

set of J couplings and RDCs. A practical application is illustrated for the model protein GB3. The protocol includes the following important steps: (i) the NMR pulse sequence, (ii) the determination of eNOEs corrected for spin diffusion, (iii) the $1/r^6$ conversion of eNOE rates into distances, (iv) the classification of distances restraints, (v) the use of bundling restraints to generate a representation of the structure with least divergence compatible with the experimental data, (vi) the ensemble-based structure calculation using the software CYANA, and (vii) the selection of the ensemble with the appropriate number of states to represent the structure. We showed that the quality of the eNOE data obtained from GB3 is excellent. The ensemble structures were cross-validated with jackknife tests as well as with high-resolution structures independently obtained from X-ray diffraction, RDCs, scalar couplings and cross-correlated relaxation rates. The present structure ensemble appears to be a good representation of the solution state of GB3.

4.2. *Current limitations of the method*

The current study made use of a highly concentrated, relatively small protein. For a potential application to larger systems, it is important to consider the following current limitations.

(a) A relatively short interscan delay of 0.6 s was chosen to reduce spectrometer time. It is likely that the Boltzmann equilibrium is not fully reestablished during the scans. This leads to reduced starting magnetization with considerable non-uniformity among all types of atoms. However, the eventually non-equilibrated starting magnetization does not influence the determination of the rate constant if the cross-peaks on both sides of the diagonal are available for the analysis. However, it causes an additional systematic error if only one cross-peak can be evaluated [19,20]. In such cases, there are also other non-equilibrated relaxation effects and therefore the upper and lower distance limits from these NOEs have been implemented with a 15% tolerance corresponding to a reduction or increase of the NOE rate of more than a factor of 2. For larger systems, however, as the molecular tumbling time increases with increasing molecule size it is recommended to increase the inter-scan delay accordingly.

(b) In the presented approach isotropic molecular tumbling is assumed, which is a very good approximation for GB3 with a D_{long}/D_{trans} ratio of 1.4. Even for D_{long}/D_{trans} less than 3 the distance error is estimated to be less than 5% for rigid spin pairs [20] and under realistic assumption about dynamics considerably smaller. Thus, the anisotropy becomes only relevant for extreme cases.

(c) The very high protein concentration of 4 mM used may appear to be a challenge for many biological relevant proteins of interest. However, while the extracted intensities of strong peaks are not dominated by the

signal-to-noise, weak peaks are. If samples with lower concentration are used, the weakest peaks will be sacrificed. For example, half the concentration yields half the signal-to-noise, which results in a decrease of the number of distance restraints because of the decrease of the measurable ^{1}H-^{1}H distance range of 10% (e.g. from 5 to 4.5 Å). Alternatively, the number of scans has to be increased.

(d) The spin diffusion does not have an adverse impact for larger molecules. As the cross-relaxation rate constant is approximately proportional to the tumbling time, the relative intensity change is identical for all sizes. However, the optimal maximal mixing time must be adjusted to the rotational tumbling time of the protein as highlighted in refs. [20,56].

(e) The collection of eNOE rates depends critically on the number of diagonal peaks for which the intensity can be determined reliably. While for GB3 (56 residues) sufficient diagonal peaks have been evaluated, it is expected that for considerably larger proteins the number of evaluable diagonal cross- peaks will decrease significantly due to extensive peak overlap. Whether 4-dimensional spectroscopy in combination with sparse data sampling or/and a reduction of the signal overlap with residue- or atom type-selective isotope labelling may alleviate this problem remains to be demonstrated.

(f) The assumption of a small impact of fast motion on the NOE has been demonstrated for H-X order parameters which are larger than 0.5 [59]. Similar conclusions have been drawn from molecular dynamics studies, but they also reveal that a few NOEs may violate the assumption considerably [57,58]. Post estimated order parameters from a 102 picosecond trajectory of molecular dynamic simulation of lysozyme [58]. Overall, the order parameter is 0.91 with 0.94 ± 0.36 and 0.87 ± 0.29 for interior and surface NOEs, respectively. The radial contribution is 1.05 ± 0.44 and 1.05 ± 0.32 in both cases, and the angular contribution is 0.90 ± 0.09 and 0.83 ± 0.13, respectively. While for most of the NOEs these two motional contributions cancel significantly, for somewhat more than 10% of the NOEs the extracted distances had an accuracy of less than 10%. Note, however, that in these calculations the rigid structure was an energy-minimized average simulation structure, while without energy minimization, all the standard deviations are smaller by a factor of two. A detailed analysis of these critical NOEs show that half of them involve side-chain arginine and lysine protons in the protein interior and not surprisingly, the most extreme averaging involves dihedral transitions. In contrast, the largest violations obtained from a 800 ps MD simulation of the cyclic decapeptide antamanide in chloroform are mostly observed for NOEs involving side chain atoms of a phenylalanine. Conclusively, NOEs involving atoms located at the far end of very long highly flexible side chains must be applied with caution. It must be noted, however, that the strongest driving force in the ensemble calculated is the cumulative impact of a dense network of NOEs which should reveal such discussed potentially wrong distances.

In conclusion, the present ensemble-based structure determination of GB3 resulted in a three-state ensemble that describes the experimental input data well. As highlighted in the accompanying paper (B. Vögeli, S. Kazemi, P. Güntert and R. Riek, Nat. Struct. Mol. Biol., in press), the structure ensemble has distinct structural states that reveal conformational-exchange dynamics in a correlated/concerted fashion. Thus, ensemble-based structure determination in conjunction with eNOEs constitutes an interesting approach to elucidate concerted motion of proteins and its role in protein activity and function.

References

[1] K. Wüthrich, *NMR of Proteins and Nucleic Acid* (Wiley, New York, 1986).
[2] D. Neuhaus and M.P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis* (Wiley, New York, 2000).
[3] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, M. Rance and N.J. Skleton, *Protein NMR Spectroscopy. Principles and Practice* (Academic Press, San Diego, 2007).
[4] I. Solomon, Phys. Rev. **99**, 559 (1955).
[5] A. Kalk and H.J.C. Berendsen, J. Magn. Reson. **24**, 343 (1976).
[6] A. Kumar, G. Wagner, R.R. Ernst and K. Wüthrich, J. Am. Chem. Soc. **103**, 3654 (1981).
[7] J.W. Keepers and T.L. James, J. Magn. Reson. **57**, 404 (1984).
[8] E.T. Olejniczak, F.M. Poulsen and C.M. Dobson, J. Magn. Reson. **59**, 518 (1984).
[9] G.M. Clore and A.M. Gronenborn, J. Magn. Reson. **61**, 158 (1985).
[10] W. Massefski and P.H. Bolton, J. Magn. Reson. **65**, 526 (1985).

[11] E.T. Olejniczak, R.T. Gampe and S.W. Fesik, J. Magn. Reson. **67**, 28 (1986).
[12] J.F. Lefevre, A.N. Lane and O. Jardetzky, Biochemistry **26**, 5076 (1987).
[13] R. Boelens, T.M.G. Koning and R. Kaptein, J. Mol. Struct. **173**, 299 (1988).
[14] B.T. Farmer, S. Macura and L.R. Brown, J. Magn. Reson. **80**, 1 (1988).
[15] G.M. Clore and A.M. Gronenborn, J. Magn. Reson. **84**, 398 (1989).
[16] R. Boelens, T.M.G. Koning, G.A. Van der Marel, J.H. Van Boom and R. Kaptein, J. Magn. Reson. **82**, 290 (1989).
[17] A.K. Suri and R.M. Levy, J. Magn. Reson. **B 106**, 24 (1995).
[18] S.J.F. Vincent, C. Zwahlen, P.H. Bolton, T.M. Logan and G. Bodenhausen, J. Am. Chem. Soc. **118**, 3531 (1996).
[19] B. Vögeli, T.F. Segawa, D. Leitz, A. Sobol, A. Choutko, D. Trzesniak, W. van Gunsteren and R. Riek, J. Am. Chem. Soc. **131**, 17 215 (2009).
[20] B. Vögeli, M. Friedmann, D. Leitz, A. Sobol and R. Riek, J. Magn. Reson. **204**, 290 (2010).
[21] E.T. Olejniczak, C.M. Dobson, M. Karplus and R.M. Levy, J. Am. Chem. Soc. **106**, 1923 (1984).
[22] N. Tjandra and A. Bax, Science **278**, 1111 (1997).
[23] J. Iwahara and G.M. Clore, Nature **440**, 1227 (2006).
[24] G. Otting, J. Biomol. NMR **42**, 1 (2008).
[25] P. Güntert, Eur. Biophys. J. **38**, 129 (2009).
[26] R. Horst, G. Wider, J. Fiaux, E.B. Bertelson, A.L. Horwich and K. Wüthrich, Proc. Natl. Acad. Sci. USA **103**, 15 445 (2006).
[27] T.S. Ulmer, B.E. Ramirez, F. Delaglio and A. Bax, J. Am. Chem. Soc. **125**, 9179 (2003).
[28] L. Yao, B. Vögeli, D.A. Torchia and A. Bax, J. Phys. Chem. **B 112**, 6045 (2008).
[29] L. Yao, B. Vögeli, J. Ying and A. Bax, J. Am. Chem. Soc. **130**, 16 518 (2008).
[30] P. Güntert, Prog. Nucl. Magn. Reson. Spectrosc. **43**, 105 (2003).
[31] R. Brüschweiler, M. Blackledge and R.R. Ernst, J. Biomol. NMR **1**, 3 (1991).
[32] A.M.J.J. Bonvin, J.A.C. Rullmann, R.M.J.N. Lamerichs, R. Boelens and R. Kaptain, Proteins Struct. Func. Genet. **15**, 385 (1993).
[33] J. Kemmink, C.P.M. van Mierlo, R.M. Scheek and T.E. Creighton, J. Mol. Biol. **230**, 312 (1993).
[34] A.M.J.J. Bonvin, R. Boelens and R. Kaptain, J. Biomol. NMR **4**, 143 (1994).
[35] D.F. Mierke, R.M. Scheek and H. Kessler, Biopolymers **34**, 559 (1994).
[36] J. Fennen, A.E. Torda and W.F. van Gunsteren, J. Biomol. NMR **6**, 163 (1995).
[37] A.M.J.J. Bonvin and A.T. Brünger, J. Mol. Biol. **250**, 80 (1995).
[38] A.M.J.J. Bonvin and A.T. Brünger, J. Biomol. NMR **7**, 72 (1996).
[39] P. Bernado and M.J. Blackledge, Am. Chem. Soc. **126**, 4907 (2004).
[40] G.M. Clore and C.D. Schwieters, J. Am. Chem. Soc. **126**, 2923 (2004).
[41] G.M. Clore and C.D. Schwieters, Biochemistry **43**, 10 678 (2004).
[42] K. Lindorff-Larsen, R.B. Best, M.A. DePristo, C.M. Dobson and M. Vendruscolo, Nature **433**, 128 (2005).
[43] G. Bouvignies, P. Bernado, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler and M. Blackledge, Proc. Natl. Acad. Sci. USA **102**, 13 885 (2005).
[44] G.M. Clore and C.D. Schwieters, J. Mol. Biol. **355**, 879 (2006).
[45] B. Vögeli, L. Yao and A. Bax, J. Biomol. NMR **41**, 17 (2008).
[46] B.T. Farmer, J. Magn. Reson. **93**, 635 (1991).
[47] S.M. Pascal, D.R. Muhandiram, T. Yamazaki, J.D. Forman-Kay and L.E. Kay, J. Magn. Reson. Ser. **B 103**, 197 (1994).
[48] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer and A. Bax, J. Biomol. NMR **6**, 277 (1995).
[49] B.A. Johnson and R.A. Blevins, J. Biomol. NMR **4**, 603 (1994).
[50] D. Neri, G. Otting and K. Wüthrich, Tetrahedron **46**, 3287 (1990).
[51] L.E. Kay, D.A. Torchia and A. Bax, Biochemistry **28**, 8972 (1989).
[52] J.B. Hall and D. Fushman, J. Biomol. NMR **27**, 261 (2003).
[53] B. Vögeli, J.F. Ying, A. Grishaev and A. Bax, J. Am. Chem. Soc. **129**, 9377 (2007).
[54] A.C. Wang and A. Bax, J. Am. Chem. Soc. **118**, 2483 (1996).
[55] B. Vögeli and L. Yao, J. Am. Chem. Soc. **131**, 3668 (2009).
[56] J. Orts, B. Vögeli and R. Riek, J. Chem. Theory Comput. (2012), doi: 10.1021/ct3002249.
[57] R. Brüschweiler, B. Roux, M. Blackledge, C. Griesinger, M. Karplus and R.R. Ernst, J. Am. Chem. Soc. **114**, 2289 (1992).
[58] C.B. Post, J. Mo. Biol. **224**, 1087 (1992).
[59] D. Leitz, B. Vögeli, J. Greenwald and R. Riek, J. Phys. Chem. **B 115**, 7648 (2011).
[60] K. Wüthrich, M. Billeter and W. Braun, J. Mol. Biol. **169**, 949 (1983).
[61] T.M.G. Koning, R. Boelens and R. Kaptein, J. Magn. Reson. **90**, 111 (1990).
[62] C.M. Fletcher, D.N.M. Jones, R. Diamond and D. Neuhaus, J. Biomol. NMR **8**, 292 (1996).
[63] P. Güntert, W. Braun and K. Wüthrich, J. Mol. Biol. **217**, 517 (1991).
[64] P. Güntert, C. Mumenthaler and K. Wüthrich, J. Mol. Biol. **273**, 283 (1997).
[65] K. Houben and R. Boelens, J. Biomol. NMR **29**, 151 (2004).
[66] C. Fares, N.A. Lakomek, K.F.A. Walter, B.T.C. Frank, J. Meiler, S. Becker and C. Griesinger, J. Biomol. NMR **45**, 23 (2009).

[67] A.T. Brunger, G.M. Clore, A.M. Gronenborn, R. Saffrich and M. Nilges, Science **261**, 328 (1993).
[68] J.P. Derrick and D.B. Wigley, J. Mol. Biol. **243**, 906 (1994).
[69] E. Miclet, J. Boisbouvier and A. Bax, J. Biomol. NMR **31**, 201 (2005).
[70] L.Y. Lian, J.P. Derrick, M.J. Sutcliffe, J.C. Yang and G.C.K. Roberts, J. Mol. Biol. **228**, 1219 (1992).
[71] J.J. Chou, D.A. Case and A. Bax, J. Am. Chem. Soc. **125**, 8959 (2003).
[72] J.S. Hu, S. Grzesiek and A. Bax, J. Am. Chem. Soc. **119**, 1803 (1997).
[73] C. Perez, F. Löhr, H. Ruterjans and J.M. Schmidt, J. Am. Chem. Soc. **119**, 7081 (2001).
[74] B. Vögeli and R. Riek, J. Biomol. NMR **46**, 135 (2010).
[75] A. Bax and D.G. Davies, J. Magn. Reson. **65**, 355 (1985).
[76] A.J. Shaka, P.B. Barker and R. Freeman, J. Magn. Reson. **194**, 547 (1985).
[77] R. Tycko, A. Pines and J. Guckenheimer, J. Chem. Phys. **83**, 2775 (1985).
[78] T. Fujiwara and K. Nagayama, J. Magn. Reson. **77**, 53 (1988).
[79] J.M. Böhlen and G. Bodenhausen, J. Magn. Reson. Ser. **A 102**, 293 (1993).
[80] D. Marion, M. Ikura, R. Tschudin and A. Bax, J. Magn. Reson. **85**, 393 (1989).