Journal of Structural Biology 191 (2015) 306–317

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Complementarity and congruence between exact NOEs and traditional NMR probes for spatial decoding of protein dynamics

Beat Vögeli [a,*], Simon Olsson [a,b], Roland Riek [a], Peter Güntert [a,c,d]

[a] Laboratory of Physical Chemistry, Vladimir-Prelog-Weg 2, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland
[b] Institute for Research in Biomedicine, Via Vincenzo Vela 6, CH-6500 Bellinzona, Switzerland
[c] 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
[d] Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form 16 July 2015
Accepted 20 July 2015
Available online 20 July 2015

Keywords:
Exact NOE
eNOE
Protein ensemble
Structure determination
GB3

ABSTRACT

The study of the spatial sampling of biomolecules is essential to understanding the structure–dynamics–function relationship. We have established a protocol for the determination of multiple-state ensembles based on exact measurements of the nuclear Overhauser effect (eNOE). The protocol is practical since it does not require any additional data, while all other NMR data sets must be supplemented by NOE restraints. The question arises as to how much structural and dynamics information is shared between the eNOEs and other NMR probes. We compile one of the largest and most diverse NMR data sets of a protein to date consisting of eNOEs, RDCs and J couplings for GB3. We show that the eNOEs improve the back-prediction of RDCs and J couplings, either upon use of more than one state, or in comparison to conventional NOEs. Our findings indicate that the eNOE data is self-consistent, consistent with other data, and that the structural representation with multiple states is warranted.

1. Introduction

The determination of the average conformation of a biomolecule has emerged as a cornerstone towards the understanding of the structure–function relationship. However, it is increasingly recognized that the structure and function is often bridged by the dynamics of the molecule. Therefore, elucidation of the spatial sampling of the molecule is also essential to the study of the function. Common probes of dynamics are carbon and nitrogen relaxation rates, residual dipolar couplings (RDCs) and scalar (*J*) couplings Torchia, 2015; Anthis and Clore, 2015; Shapiro, 2013; Ban et al., 2013; Sekhar and Kay, 2013. However, all of those probes must be supplemented by nuclear Overhauser enhancement (NOE) data in *de novo* structure calculation. We have established a protocol for the determination of protein ensembles represented by multiple states based on the exact measurement of the NOE (eNOE) Vögeli, 2014; Vögeli et al., 2014, 2012. Such eNOEs are complementary in nature to the commonly used data. Provided that the NOE can be quantified accurately (Vögeli et al., 2009, 2010), it is a very sensitive reporter of temporarily visited short proton–proton distances because of its proportionality to the inverse sixth power of the distance (Solomon, 1955; Bell and Saunders, 1970). This property is particularly useful when a dense network of exact NOEs can be collected (Fesik et al., 1986; Kessler et al., 1988; Brüschweiler et al., 1991; Constantine et al., 1995; Bonvin and Brünger, 1995).

We have previously collected 884 eNOEs for the 56-residue protein GB3 (Vögeli et al., 2013, 2012). When combined with a small number of backbone *J* couplings and RDCs, three simultaneous states with fixed equal weights were required to back-predict the input data satisfactorily (Vögeli et al., 2013, 2012). While the NOE is primarily sensitive to the distance fluctuations between protons (Brüschweiler et al., 1992), RDCs depend on the orientation of the vector connecting a proton and ^{13}C or ^{15}N, or two heavy atoms (Tjandra and Bax, 1997), and relevant *J* couplings are mostly determined by the dihedral angles with the rotation axis given by two heavy atoms (Karplus, 1963). Here, we investigate how much structural and dynamics information is shared between the eNOEs and the other NMR probes. To address this question, we compare eNOEs to RDCs and *J* couplings, all of which are sensitive to motions faster than milliseconds.

We compile an extensive NMR data set consisting of eNOEs, RDCs and *J* couplings for GB3, which forms one of the largest and

* Corresponding author at: Laboratory of Physical Chemistry, HCI F217, Vladimir-Prelog-Weg, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland.
E-mail address: beat.voegeli@phys.chem.ethz.ch (B. Vögeli).

http://dx.doi.org/10.1016/j.jsb.2015.07.008
1047-8477/

most diverse data set to date. This data set forms an excellent basis for many cross-validation tests. By means of various leave-out checks with an updated protocol for CYANA (Güntert, 2003, 2009), we analyze the self-consistency and redundancy of the input data using ensembles with various numbers of states. Because NOEs are indispensable in an NMR structure calculation, we assess the information content of the RDCs and *J* couplings by analyzing the effect on ensembles when either eNOE or conventional NOE input data sets are supplemented by RDCs and *J* couplings. Finally, we assess the stability of the determination of long-range correlations by cross-validation tests.

2. Results and discussion

We compiled an NMR data set consisting of 984 eNOEs, 1477 RDCs and 225 *J* couplings for GB3. To that purpose, the previously published set consisting of 884 eNOEs (Vögeli et al., 2013, 2012) was supplemented by eNOEs involving methylenes with a degenerate proton pair and Val with a degenerate pair of methyl groups (no Leu counterparts were obtained here). The RDCs were collected from the literature, in total originating from 8 different alignment conditions. 1335 RDCs are obtained from spin pairs located in the backbone, 129 in side chains, and 13 bridging the backbone and side chains. All *J* couplings, most of which are taken from the literature, extend over three covalent bonds, of which 147 are from the backbone, and 78 between a backbone and a side-chain spin. All data have small experimental errors, which are carefully estimated. These errors are then used for the widths of a flat bottom, which is inserted into harmonic wells contributing to the CYANA target function in the structure calculation. A detailed account on the measurements of the data, along with details on estimations of the alignment tensors, Karplus curve coefficients and experimental errors is provided in the Section 4. The complete restraints data set is provided in the accompanying data article (Vögeli et al., 2015).

We calculated structural ensembles for one to nine states from the complete data set using the CYANA protocol described in Section 4. The overall target function is shown in red in Fig. 1. The average of the target function value over the 20 lowest-energy conformers of ca. 117 Å^2 obtained for one state decreases to ca. one third of its value for two states. Additional decrease is observed when going to three and four states, where it reaches a plateau value of ca. 30 Å^2. For a comparison, the target function of the previously calculated ensemble (pdb code for three-state ensemble 2LUM) is shown in green (Vögeli et al., 2012). The input data consisted of the 884 eNOEs mentioned above, 90 RDCs and 147 backbone *J* couplings. The target function value is ca. 27 Å^2 for one state and levels out slightly below 10 Å^2. The reason for the large increase upon the use of the complete data set is the dominant contribution from the RDCs (ca. 80 Å^2). The relative weights of the eNOEs, RDCs and *J* couplings is somewhat arbitrary, but changes of the weight of the RDC contributions by an order of magnitude produced virtually identical ensembles.

In the next step, we repeated the calculations by systematically omitting subsets of the complete input data. For the left-out data, we back-calculated cross-validation target function values that quantify the disagreement between the left-out data and the ensembles. Finally, we repeated all leave-out tests with 1953 previously determined conventional NOEs (Vögeli et al., 2013, 2012) instead of the eNOEs.

2.1. Cross-validation of input data

2.1.1. eNOEs

First, we assess the self-consistency of the eNOE data set. The bundles were recalculated ten times with the complete RDC and *J* coupling data set, but 10% of the eNOE data was removed from the input (Brünger et al., 1993). Each eNOE was left out exactly once such that a complete target function could subsequently be reconstructed. The target function decreases continuously and significantly from one to five states, where it reaches a plateau at half of the value at the single state (Fig. 2). This indicates a partial information redundancy of the eNOE data.

2.1.2. Residual dipolar couplings

To assess the extent of shared information between the eNOEs and the RDCs, target function values of selected RDC data sets were calculated from ensembles obtained from either only the eNOEs or only the conventional NOEs, under exclusion of the RDCs and the *J* couplings.

For the RDC target functions, new alignment tensors were determined for each number of states separately by carrying out singular value decomposition (SVD) on the averaged spherical harmonics. It is important to note that this procedure underestimates the true decrease of the true cross-validation target function with increasing number of states. A very conservative estimate suggests

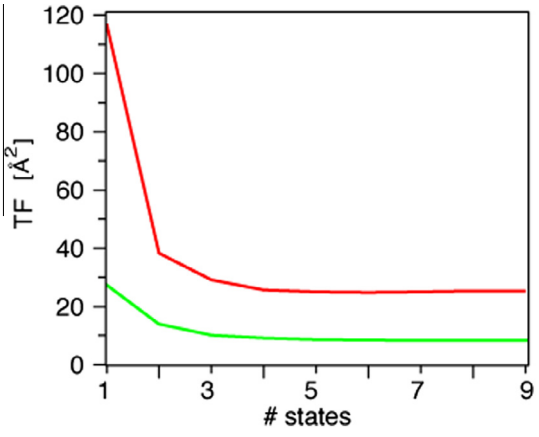

Fig. 1. Target functions obtained from multi-state ensemble determination of GB3. The target function (TF) obtained with the complete set of eNOEs, RDCs and *J* couplings used in this study is shown in red. The target function of the previously published ensembles (pdb code of three-state ensemble: 2LUM) calculated with a slightly smaller eNOE set, a small RDC set and only the backbone *J* couplings is plotted in green (Vögeli et al., 2012).

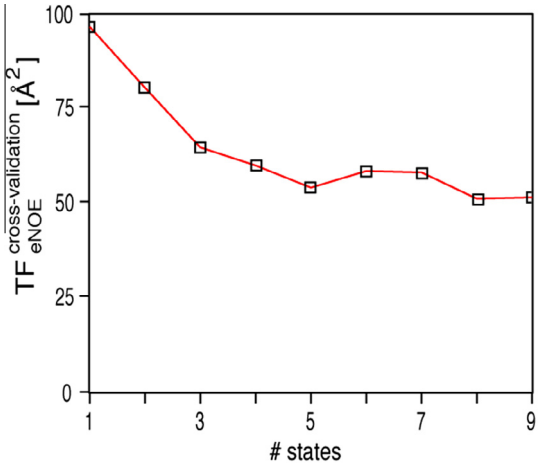

Fig. 2. Cross-validation of the eNOE data. The eNOE target function (TF) of left-out eNOEs is plotted versus the number of states. Ten ensembles per state number were calculated, where 10% of the eNOEs were removed from the input data. Each eNOE was left out exactly once.

an additional drop by 20–30 Å^2 of the target function value (see Section 4 for details). In the following, we interpret the cross-validation results as the lower limit for self-consistency.

Fig. 3 shows back-calculated target functions of RDC sets from ensembles determined from only the conventional NOEs (dashed lines) and only the eNOEs (solid lines). Target functions encompassing all RDCs are shown in red, and those from RDCs obtained from the backbone and side chains in blue and green, respectively. The significant overall decrease upon use of eNOEs instead of conventional NOEs is dominated by a 50% decrease of the contribution from the backbone RDCs. A plausible explanation is the need of very accurate geometry to improve the RDC back-prediction, which is better achieved in the backbone due to the higher density of eNOE restraints. The eNOEs also produce the lowest target function values for side-chain RDCs (three states). These results are impressive as conventional NOEs and eNOEs provide primarily translational information expectedly complementary to the angular information of the RDCs. Interestingly, the RDC cross-validation in the backbone does not show improvement when going from one to multiple states, while there is substantial improvement in the side chains. This is expected as the states separation in the backbone is mostly of translational nature, while rotamer hops in the side chains are readily picked up by angle-dependent observables such as RDCs.

2.1.3. J couplings

Next, we assess the extent of shared information between the eNOEs and the *J* couplings. The target functions of the *J* coupling data set were again calculated from ensembles obtained from either only the eNOEs or only the 1953 conventional NOEs.

A similar effect is observed for target functions of back-calculated *J* couplings as the one for the RDCs (Fig. 4). The use of eNOEs instead of conventional NOEs for bundle calculation reduces the target function to less than 50%, and a slight decrease is observed when going from one to two, and to three states. Again, both findings are remarkable, since the *J* couplings are dominated by dihedral angles involving heavy atoms.

2.1.4. Residual dipolar couplings with variable internuclear distances

An additional validation is obtained from back-predicted intraresidual $^3D_{HN,H\alpha}$ and sequential $^4D_{HN,H\alpha}$ RDCs (Fig. 5). Because CYANA cannot accommodate RDCs with variable

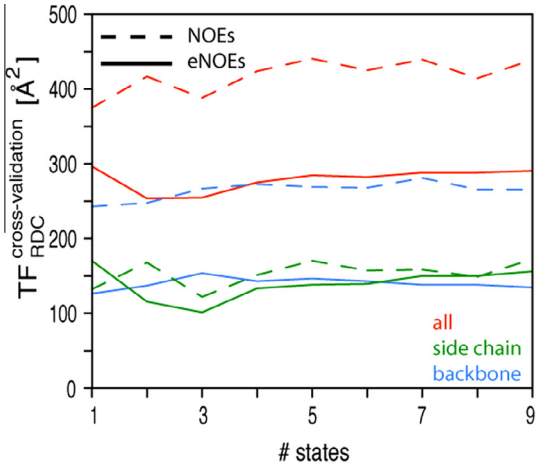

Fig. 3. Cross-validation of the eNOE data with RDCs. Back-calculated target functions (TF) of RDCs from ensembles determined from only conventional NOEs (dashed lines) and only eNOEs (solid lines) are plotted versus the number of states. TFs encompassing all RDCs are shown in red, and those from RDCs obtained from the backbone and side chains in blue and green, respectively.

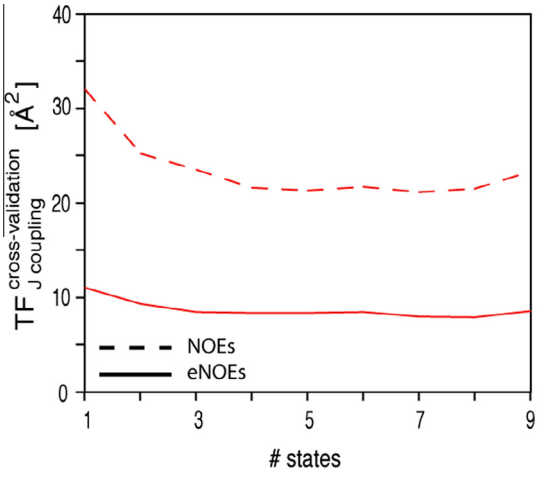

Fig. 4. Cross-validation of the eNOE data with *J* couplings. The back-calculated target functions (TF) of *J* couplings from ensembles determined from only conventional NOEs (dashed lines) and only eNOEs (solid lines) are plotted versus the number of states.

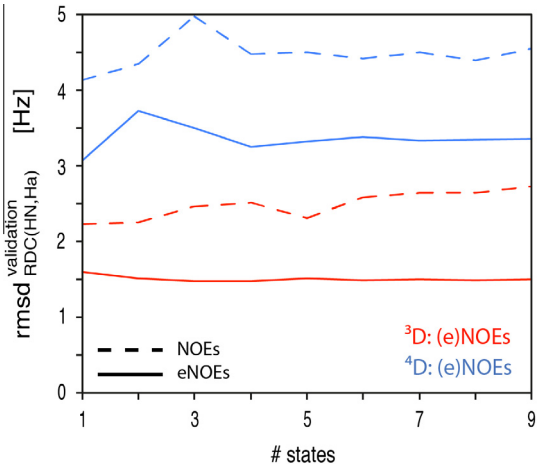

Fig. 5. Validation of the eNOE data with additional RDCs with variable internuclear distances. The root-mean-square deviations (rmsd) between experimental $^3D_{HN,H\alpha}$ (red) and $^4D_{HN,H\alpha}$ RDCs (blue) and those back-calculated from ensembles determined from only conventional NOEs (dashed lines) and only eNOEs (solid lines) are plotted versus the number of states.

internuclear distances as restraints in a structure calculation and no experimental errors are available, we calculate the root-mean-square deviations (rmsd) between the experimental and back-predicted RDCs rather than the target functions. The rmsd values obtained from the ensembles calculated from conventional NOEs are typically ca. 30% larger than when calculated from the eNOEs. In contrary to the RDCs and *J* couplings analyzed above, the prediction of these RDCs from two or three states does not substantially improve the agreement with the experimental values.

Because these RDCs were not used in the structure calculation, they may be used for a validation of structures computed using the complete input data set (all other RDCs, conventional or eNOEs, and *J* couplings, see Supporting material, Fig. S1). For the $^4D_{HN,H\alpha}$ RDCs, which depend on the ψ torsion angles, the eNOE ensembles yield lower rmsd values than those from conventional NOEs (ca. 15% less). On the other hand, the $^3D_{HN,H\alpha}$ RDCs yield similar rmsd values for conventional NOEs and eNOEs, presumably because the employed *J* couplings, which define the φ torsion angles, restrict the relative positioning of H^N_i and H^α_i. However, a similar difference is obtained for $^3D_{HN,H\alpha}$ as for the $^4D_{HN,H\alpha}$ RDCs when *J*

couplings, but no RDCs are used in the ensemble calculation. Apparently, the RDCs are required as a supplement to the *J* couplings to replace the information loss when going from eNOEs to conventional NOEs.

2.2. Structural correlations

For simplicity, the correlations inherent to two-state bundles are analyzed first. Two-state bundles calculated from various combinations of input data are shown in Fig. 6. The grouping into two states is particularly obvious for the segment of residues 45 and 46 of β strand 3, in which we manually assigned from each of the 20 two-state ensembles one state to the red and the other to the blue group, respectively. In all ensembles calculated from the complete eNOE data set, a strong correlation is present with the two states in the loop comprising residues 49–51 and with the adjacent β strand 4. To a lesser extent, the correlations also propagate into β strands 1 and 2. The fact that all data sets containing all eNOEs produce such a pattern suggests that the correlations are not coded by a few specific restraints only, but are contained in multiple ways in the complete input data set. Similar states and correlations are obtained from reduced eNOE data sets, either created by deleting 10% of the complete eNOE set (second, third and fourth rows in Fig. 6), or by taking the input set used for the previously published three-state ensemble 2LUM (but used for a two-state calculation here, top row right) (Vögeli et al., 2013, 2012). Slight structural differences must be caused by unique information that the missing restraints carry. A completely different ensemble is obtained if conventional NOEs are the sole input into structure calculation. The states are not as compact and no segment displays two distinct states that can be grouped with respect to all ensembles. Grouping of the states fails to produce any correlations. When the NOE set is supplemented by the RDC and *J* coupling data sets, a similar ensemble is regained as those obtained from eNOEs. This suggests that similar information is inherent to the combined NOE/*J* coupling/RDC data set as in the eNOE set alone. It must be emphasized, however, that the experimental effort to collect the eNOE restraints is much smaller.

Interestingly, it appears that the correlations between the states in residues 45/46 and the adjacent loop are switched upon use of the RDC data sets as compared to the use of eNOE (and *J* coupling) data sets only (see top row in Fig. 6). None of the φ and ψ angles of residues 45–47 is significantly different in the ensembles with and without the switch. Instead, the effect can only be identified by inspection of the translational behavior. The question arises whether this difference is caused by the limitations of a simple two-state model or whether the information contents of the eNOE and the RDC data sets are contradicting in that particular respect. Removing either all RDCs obtained from residues that were mutated to modify the Pf1-induced alignment conditions (see Section 4) or from all side-chain residues does not alter the correlations. Note that the crossing-over of the states is located to residue 45, whereas only the construct giving rise to alignment tensor 2 has a nearby mutation (residue 47). Therefore, we removed all backbone RDCs induced by alignment medium 2. Again, we did not observe any changes. Neither did we observe changes upon removing the backbone RDCs involving residues 45, 46, 45/46 or 46/47. However, the crossing-over effect disappears when all backbone RDCs are removed (a large rms deviation between individual states is obtained due to poor tensor definition) or when those involving the backbone of residues 45/46/47 are removed. As the RDC data sets are self-consistent (Yao et al., 2008a,b; Yao and Bax, 2007), it is unlikely that the negatively charged segment Asp46/Asp47 of GB3 is involved in electrostatic interaction with the negatively charged Pf1 which alters the dynamics very slightly.

Grouping of the states carried out at residues 45 and 46 for various three-state ensembles is shown in Fig. 7 (first and third row). The findings are very similar to those from the two-state ensembles. In particular, the correlations that are obtained for two-state structure determination are conserved. This indicates that the correlations are not simply an artifact enforced by the simplicity of the two-state assumption. Similar correlations are observed in the four-state ensembles. In general, increasing the number of states fills the gaps between the distinct clusters of states of the ensembles with a lower number of states.

An alternative grouping of states, this time focusing on residue 9, is shown in the second and fourth row of Fig. 7. This representation illustrates mostly correlations in the proximity of residue 9. Interestingly, there is long-range correlation to a small extent with the grouping at residues 45 and 46 in all ensembles based on eNOEs and the one based on conventional NOEs/RDCs/*J* couplings.

In a complementary approach, we analyzed the pairwise linear correlations of the fluctuations of the C^{α} atom positions in the two-state ensembles using heat maps of correlation matrices generated with the program THESEUS (Theobald and Wuttke, 2006) (see Figs. 8 and S2 in the Supporting material). Similar patterns are obtained if eNOE data is used in the structure calculation. For the ensembles determined using complementary RDC and *J* coupling data in addition to the eNOE data, the same correlation pattern is enhanced and a few additional distinct correlation patterns emerge. The majority of the correlations is lost when conventional NOEs are the only input data. Upon addition of the RDC and *J* coupling data, the pattern resembles the average of the two eNOEs ensembles shown in Fig. 8. Unlike the correlations observed here, the correlations revealed by manual state assignment are non-linear (see next section). Thus, the additional linear correlations may indicate partial information redundancy between the eNOE and complementary data sets. In the way the input data is integrated into the ensembles determination protocol, potential 'double counting' of information from different data types will occur. This hypothesis is consistent with the cross-validation tests shown above.

2.3. Principal component analysis and state separability

To more quantitatively analyze how well separated the two states of the ensembles determined using the different input data sets are, we perform principal component analysis (PCA) Vögeli et al., 2012; Bahar et al., 2010 of each of the ensembles independently. The PCA projects the ensembles onto an orthonormal basis which is sorted according to the variance of the backbone C^{α} coordinate positions. We find a significant concentration of the C^{α} coordinate variance along a small number of principal components (PCs) as quantified by the eigenvalue spectrum entropy (see Section 4 and Fig. S3 in the Supporting information).

Next, we use the first ten PCs to define a separation probability which measures how often the two states in each ensemble are assigned to different states. The higher the separation probability, the better separated the two states are. An example for state separation of the ensemble obtained from the full data set along the two first PCs is shown in Fig. S3 in the Supporting material. We find that the full data set and the tenfold cross-validation ensembles (using 90% of the eNOEs together with the RDCs and *J* couplings) generally separate the two states quite well with a separation probability of more than 0.7. This is appreciably larger than the random true-positive separation probability of 0.51 ± 0.1 (Fig. S4 in the Supplemental information). On the other hand, the ensembles generated from the conventional NOEs alone, from eNOEs and *J* couplings, or from the previously used data set (2LUM) score lower, but still with non-random separation probabilities. Thus, we see a similar picture as that obtained using the manual assignment

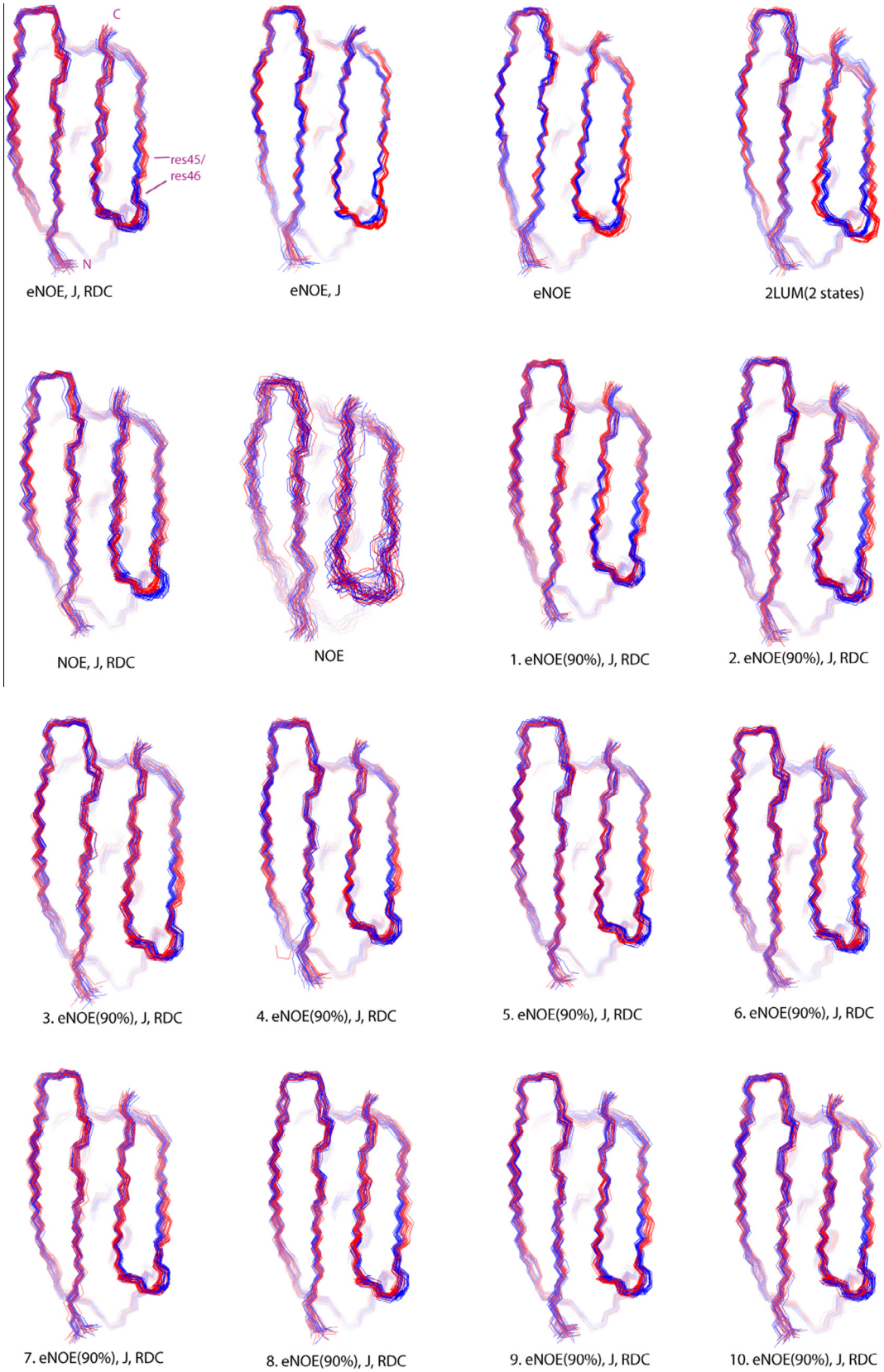

Fig. 6. Two-state ensembles of GB3 calculated from various sets of eNOEs, conventional NOEs, RDCs and *J* couplings. The two states grouped at residues 45/46 are shown in red and blue. Ensembles based on the complete eNOE data set and on the previously published eNOE data set used for the three-state ensemble 2LUM (Vögeli et al., 2013, 2012) are shown in the top row. Ensembles calculated from conventional NOEs are shown in the second row on the left. eNOE (90%) indicates randomly selected 90% of the complete eNOE data set, whereas in sets 1–10 each eNOE is omitted exactly once. The resulting ensembles are shown in rows two (right), three and four.

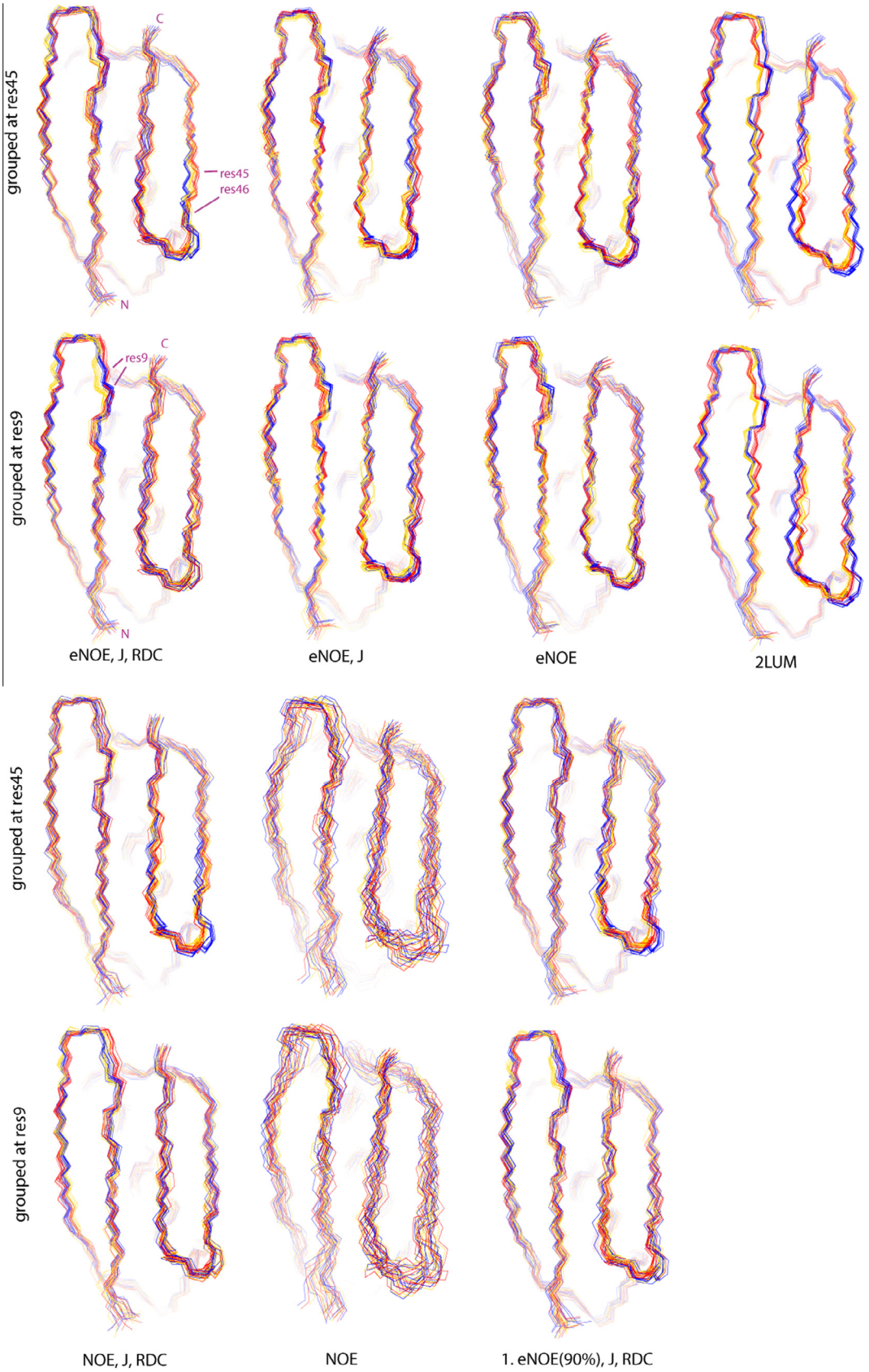

Fig. 7. Three-state ensembles of GB3 calculated from various sets of eNOEs, conventional NOEs, RDCs and *J* couplings. The three states are shown in gold, red and blue. Ensembles based on the complete eNOE data set and the previously published three-state ensemble 2LUM are shown in the top two rows, either grouped at residues 45/46 (first row) or 9 (second row). Ensembles calculated from conventional NOEs and one example of an ensemble with 10% randomly deleted eNOEs are shown on the bottom two rows, either grouped at residues 45/46 (third row) or 9 (fourth row).

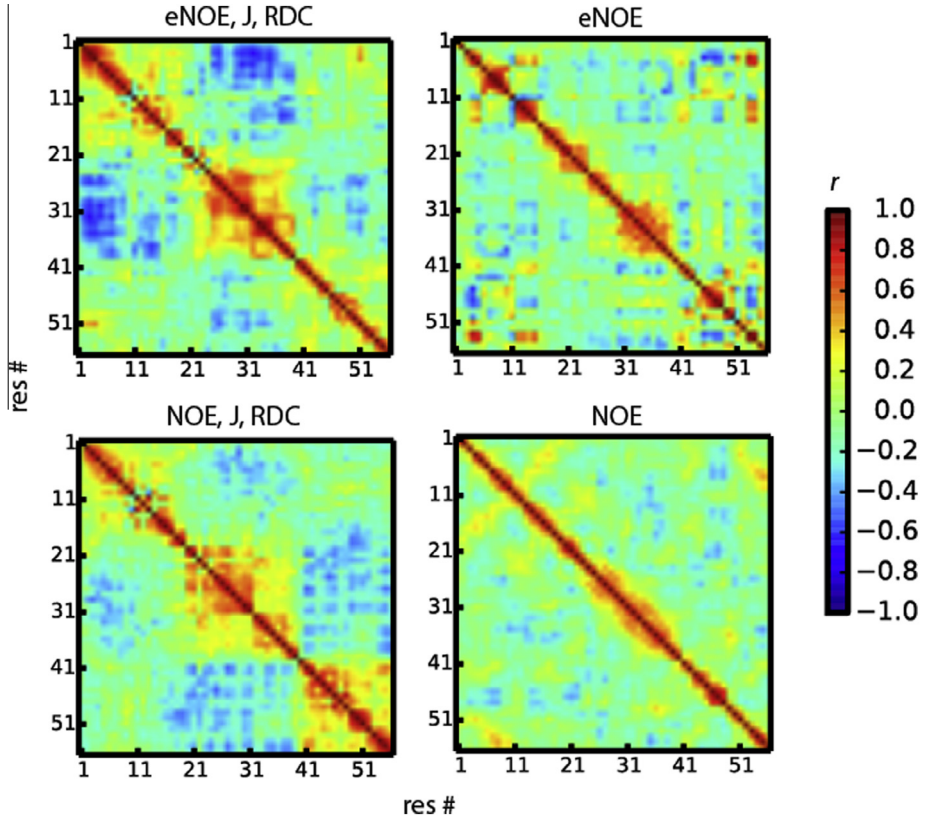

Fig. 8. Heat maps of pairwise correlations of C^{α} positional fluctuations of GB3 ensembles. The diverging color gradient ranges from perfect negative ($r = -1$, blue) to perfect positive ($r = 1$, red) correlation. The bundles are calculated from various sets of eNOEs (top) or conventional NOEs (bottom) alone (right) or supplemented with RDCs and *J* couplings (left). The plots were generated in THESEUS (Theobald and Wuttke, 2006).

in local segments. In general, when eNOE data is used the separation of the states is clearer as compared to conventional NOEs. This can be largely overcome by inclusion of *J* coupling and RDC data.

The automated state separation method considers only linear structural correlations, which is fundamentally different from the manual state separation but similar to the correlation matrix heat maps discussed above. To illustrate this, the state assignments for the full data set ensemble using the first three PCs are shown in structural representation in Fig. S6 in the Supplemental information. As expected, we observe that the PC-based separations do not coincide with the manual grouping of either the loop comprising residues 9–12 or the segment with residues 45 and 46. Instead, the modes reveal more subtle features distributed over the entire protein.

3. Conclusion

eNOEs define tighter distance limits than conventional NOEs (Chi et al., 2015). We investigated how much structural and dynamics information is shared between the eNOEs and other NMR observables that are also sensitive to sub-milliseconds motions. We compiled a large NMR data set consisting of eNOEs, RDCs and *J* couplings for GB3. We demonstrated that the eNOEs are able to substantially improve the back-prediction of RDCs and *J* couplings upon use of more than one state in the structure determination. The back-prediction is also superior to an analogous approach using conventional NOEs. These findings are remarkable, as eNOEs are primarily sensitive to the distance fluctuations between protons, and RDCs and *J* couplings are mostly determined by angles involving heavy atoms. eNOEs are well suited to define distances and their fluctuations between segments far separated in the amino-acid sequence. RDCs, on the other hand, define only the orientations of such segments well. In general, small length changes of a vector connecting two protons (say, tenths of an Ångstrom) have a large impact on the NOE, while a change in the vector orientation of a few degrees does barely alter the NOE. The situation may be exactly opposite for an RDC or a *J* coupling.

Our findings suggest that the eNOEs are self-consistent and in favorable cases may be sufficiently accurate to supersede the acquisition of complementary angular information from RDC or *J* coupling data. This is supported by extensive cross-validation tests, which are shown to yield ensemble models insensitive to both the number of chosen states and different training sets used. This notion is further supported by quantitative and qualitative analysis of linear and non-linear correlations of the protein-backbone conformations of ensembles determined with different data sets.

We have previously observed the interplay between angular and radial information either through the use of prior information (Olsson et al., 2011) or through the validation of structural ensembles with complementary data (Fenwick et al., 2011; Olsson et al., 2014, 2015). This suggests that it may be generally viable to avoid acquisition of multiple complementary observables, without compromising the quality of the information obtained. As demonstrated in this study, the eNOE-based ensembles provide geometrical details only attainable with conventional NOEs when complemented by RDC and *J* coupling data. Finally, we point out that an eNOE data set can be used as the sole input for structure calculation. This stands in contrast to RDCs and *J* couplings, which must be supplemented by conventional NOEs or prior knowledge.

4. Materials and methods

4.1. Input data

4.1.1. eNOEs

All cross-relaxation rate constants presented in references (Vögeli et al., 2013, 2012) were taken. Upper and lower limits of the distance restraints were set with an allowed distribution range of 0% and 15% for bidirectional and unidirectional eNOEs, respectively. However, NOEs involving methyl groups were processed differently. In this study, the CYANA protocol was executed with an individual treatment of each methyl proton by r^{-6} summation of the corresponding distances. This approach has been shown to be more efficient than the use of pseudo atoms (Fletcher et al., 1996). Therefore, previous input distance restraints were scaled by a factor of $3^{-1/6} = 0.83268$ per methyl group such that the corresponding cross-relaxation rate constant is a sum over all individual contributions. This would be strictly true if the methyl motion was slow (slower than nanoseconds). Since there is fast rotation present as well, we added an additional tolerance of ±8.5%.

eNOEs that involve either methylene groups with chemically equivalent protons or chemically equivalent methyl groups in Val and Leu were added to the previous data set. The apparent cross-relaxation rate constants were fitted to the same formulae as used for single atoms or methyl groups, corrected for spin diffusion (Orts et al., 2012) and normalized to the equivalent of a superposition of contributions from all pairs of single atoms. Note that this is an approximation because the spins do not undergo fast exchange. Instead, the spectral peaks are superpositions of the individual buildup/decay curves. Therefore, all upper limits and lower limits were given an additional tolerance of 5% in addition to the 0%/15% for bidirectional/unidirectional eNOEs.

Conventional NOEs were generated similarly as described in Chi et al. (2015).

4.1.2. Residual dipolar couplings

$^1D_{HN,N}$ (43 Hz range) and $^1D_{C\alpha,H\alpha}$ (92 Hz range) RDCs of wild-type GB3 under alignment induced by Pf1 (tensor 1) were taken from reference Vögeli et al. (2008). For $^1D_{HN,N}$, the errors were uniformly set to 0.5 Hz because half the pairwise rms deviation from values obtained from a new sample was 0.57 Hz, and singular value decomposition (SVD) with a 160-member multi-state ensemble (Clore and Schwieters, 2006), 2OED (Ulmer et al., 2003) and RDC-proton optimized 2OED (Yao et al., 2008a,b; Derrick and Wigley, 1994) yields fitting errors of 0.54 Hz, 0.76 Hz and 1.29 Hz, respectively. Analogously, the errors for $^1D_{C\alpha,H\alpha}$ were uniformly set to 1.5 Hz because SVD on the same structures yields fitting errors of 1.67 Hz, 2.02 Hz and 2.51 Hz, respectively, and if the alignment tensor is taken from the H^N–N RDCs, only slightly larger errors are observed with 2.07 Hz, 2.40 Hz and 3.52 Hz. An initial tensor estimate was obtained from SVD of the H^N–N RDCs on the RDC-refined X-ray structure (Yao et al., 2008a,b; Derrick and Wigley, 1994) and omission of the highly mobile residues 12, 40 and 41. This tensor was then used for both $^1D_{HN,N}$ and $^1D_{C\alpha,H\alpha}$. The relative scaling is −2.0327 (corresponding to the bond lengths of 1.02 Å and 1.09 Å).

For the deuterated mutants K19AD47K (tensor 2), K19ED40 N (tensor 3), K19EK4A-C-His6 (tensor 4), K19EK4A-N-His6 (tensor 5) and K19AT11K (tensor 6) two slightly different experiments were run to obtain H^N–N and C^α–C′ RDCs under alignment via Pf1 (Yao et al., 2008b). For these data sets, the averages are used and the errors are half the individual pairwise differences. For the (54, 54, 52, 54, 54) $^1D_{HN,N}$, the overall rms deviations for the sets corresponding to tensors (2, 3, 4, 5, 6) are (0.14, 0.17, 0.16, 0.11, 0.07) Hz for ranges of (28, 30, 35, 27, 30) Hz. If only one value was available, the error was set to twice the overall rms deviation. For the (54, 54, 50, 54, 54) $^1D_{C\alpha,C'}$, the errors for the sets corresponding to tensors (2, 3, 4, 5, 6) are (0.041, 0.078, 0.152, 0.073, 0.047) Hz for a range of (6.2, 6.2, 7.4, 6.7, 7.3) Hz. If only one value was available, the error was set to twice the overall rms deviation. The relative scaling is −0.1866 (corresponding to the bond lengths of 1.02 Å and 1.525 Å). For a seventh alignment condition under Pf1 (tensor 7, mutant K19EK4A), only one data set of $^1D_{HN,N}$ (37 Hz range) and $^1D_{C\alpha,C'}$ (8.1 Hz range) is available. The error was set uniformly to twice the largest error of the other five data sets (0.34 Hz and 0.30 Hz).

For H^N–C′, one set of RDCs is available for each mutant yielding tensors 2–6 (Yao et al., 2008b). The relative scaling of the coupling is −0.3123 (assuming an interatomic distance of 2.035 Å corresponding to lengths of 1.020 Å and 1.329 Å for the H^N–N and C′–N bonds and a projection angle of 119.5° (Vögeli, 2011). Rms deviations from back-calculated values obtained from SVD are typically 0.4 Hz (using a directly fitted tensor or a tensor obtained from $^1D_{HN,N}$). Therefore, the errors were uniformly set to 0.2 Hz.

For each of the protonated mutants K19AD47K (tensor 2), K19ED40N (tensor 3), K19EK4A-C-His6 (tensor 4), K19EK4A-N-His6 (tensor 5), K19AT11K (tensor 6) and K19EK4A (tensor 7), one $^1D_{C\alpha,H\alpha}$ data set is available (Yao et al., 2008a). To account for Pf1 concentration differences, the values were rescaled by the slopes between the $^1D_{HN,N}$ values obtained from the protonated samples (not used in this study) and the deuterated samples (see above). The alignment tensors are assumed to be the same as those determined from the $^1D_{HN,N}$ sets of the deuterated samples. Rms deviations from back-calculated values obtained by SVD are typically 1.5 Hz (or ca. 2.0–3.5 Hz using the tensors obtained from $^1D_{HN,N}$). Therefore, the errors are uniformly set to 1.0 Hz.

For the structure calculations, tensors 2–7 were determined from SVD with all measured RDCs in the backbone.

$^2D_{C\beta,H\alpha}$, $^1D_{C\beta,H\beta2}$, $^1D_{C\beta,H\beta3}$ and $^1D_{H\beta2,H\beta3}$ values were obtained from alignment with Pf1 phage (Miclet et al., 2005). A 3D HBCBCA type experiment provides four independent values for $^2D_{C\beta,H\alpha}$, which allows for an estimation of individual errors. $^1D_{C\beta,H\beta2}$, $^1D_{C\beta,H\beta3}$ and $^1D_{H\beta2,H\beta3}$ values are obtained from linear combinations of the effectively measured ($^1D_{C\beta,H\beta2} - {}^1D_{H\beta2,H\beta3}$), ($^1D_{C\beta,H\beta3} - {}^1D_{H\beta2,H\beta3}$) and ($^1D_{C\beta,H\beta2} + {}^1D_{C\beta,H\beta3}$) values. Here, the errors of these values were propagated into individual errors of the couplings of interest. Sample conditions were similar to those of the previously mentioned measurement of $^1D_{HN,N}$ with Pf1 yielding tensor 1 (sample conditions for the isotropic coupling measurements were 50 mM sodium phosphate buffer at pH 5.6 and 298 K, and for the anisotropic coupling measurement similar, except for pH 6.5). $^1D_{C\alpha,H\alpha}$ couplings obtained from the same experiment were compared (Pearson's correlation coefficient 0.99) and rescaled to the $^1D_{C\alpha,H\alpha}$ couplings mentioned above in order to estimate the alignment tensor magnitude relative to tensor 1. Then, tensor 1 was used for the structure calculations (the scaling factor was 0.814 and the errors were not scaled).

$^1D_{C\beta,H\beta}$ (Val, Ile, Thr), $^1D_{C\beta,H3\beta}$ (Ala), $^1D_{C\gamma1/2,H3\gamma1/2}$ (Val, Ile, Thr) and $^1D_{C\delta1/2,H3\delta1/2}$ (Leu, Ile) values were obtained from alignments with Pf1 phage and PEG (Clore and Schwieters, 2006). For Pf1, a set of $^1D_{C\alpha,H\alpha}$ recorded on the same sample was used to scale the couplings such that the previously mentioned tensor 1 for alignment with Pf1 could be used (scaling factor 0.90 to account for Pf1 concentration difference, Pearson's correlation coefficient 0.99) and also for the estimate of the error (rms deviation between those two sets is 2.32 Hz, which gives an error of ca. $2.32/2^{1/2}$ Hz = 1.5 Hz here). The couplings within methyl groups were scaled by −1/3.17 such that they could be used as effective $^1D_{C\alpha,C\beta}$, $^1D_{C\beta,C\gamma1/2}$ and $^1D_{C\gamma(1),C\delta1/2}$ couplings in the structure calculation (Ottiger and Bax, 1999). The rescaled errors would be ca. 0.2 Hz, but were uniformly set to 0.5 Hz. For PEG, no $^1D_{HN,N}$ couplings

were available and the alignment tensor was determined from a set of $^1D_{C\alpha,H\alpha}$ obtained in the same experiment (tensor 8). The measurements were carried out once. The errors were estimated as follows: An SVD from the $^1D_{C\alpha,H\alpha}$ couplings of Pf1 gave an rms deviation of ca. 2.7 Hz (this tensor was not used in the structure calculations). An analogous SVD from the $^1D_{C\alpha,H\alpha}$ couplings of PEG yielded an rms deviation of ca. 1 Hz, while the coupling amplitude is half as large. Thus, the RDC data from PEG appear 1.5 times better and if it is assumed that the absolute measuring errors are similar for PF1 and PEG, it would be safe to use the same errors for PEG as for Pf1. However, it is likely that tensor 8 obtained from $^1D_{C\alpha,H\alpha}$ is not as accurate as tensor 1, which is obtained from $^1D_{HN,N}$. Therefore, a uniform error of 2 Hz was chosen for $^1D_{C\alpha,H\alpha}$ and $^1D_{C\beta,H\beta}$. The errors of the couplings involving methyl groups were uniformly set to 0.5 Hz.

Next, all fitted alignment tensors were corrected for the rescaling due to uniform motion throughout the molecule. Iterative re-determination of the tensors increases them by 4% in the first cycle of a two-state ensemble calculation (CYANA target function value +0.91 Å^2, or −1.08 Å^2 when using effective bond lengths) and converges to an increase of about 5% after the second cycle. This result is in good agreement with a tensor rescaling by 1/0.95 based on SVD of the 160-member ensemble calculated by Clore and Schwieters (2006). This method may be used to determine the absolute tensor scaling, which is still a matter of debate (Sabo et al., 2014). In the following, we use the re-determined tensor after the initial two-state ensemble calculation.

In addition, intraresidual $^3D_{HN,H\alpha}$ and sequential $^4D_{HNH\alpha}$ RDCs taken from reference (Vögeli et al., 2008) were used for validation only. The alignment tensors were fitted separately.

For the RDC cross-validation target functions, alignment tensors were determined for each number of states separately by carrying out singular value decomposition (SVD) on the averaged spherical harmonics. We used all backbone RDCs, except for alignment medium 1, where we only used the H^N–N RDCs. The procedure underestimates the true decrease of the cross-validation target function with increasing number of states. The effect becomes evident when the RDC target function is back-calculated from an ensemble obtained from the complete data set including the RDCs. In this case, the cross-validation target function underestimates the true value by 15 Å^2 for one state, but overestimates it by 14 Å^2 at two states and then converges to an overestimation of 5 Å^2. It is, however, difficult to quantitatively apply corrections to the cross-validation target functions. Therefore, the cross-validation results may be interpreted as the lower limit for self-consistency.

4.1.3. Scalar couplings

$^3J_{HN,H\alpha}$ values are averages over couplings derived from CT-MQ($^1H^N$, $^{13}C^\alpha$) + SQ($^1H^N$)–HNCA spectra and *J*-modulated HMQC spectra (Vögeli et al., 2007; Kuboniwa et al., 1994). Based on the pairwise rms deviation between the values obtained from the two types of measurements, the error of their averaged values equals 0.14 Hz (Vögeli et al., 2007). For each coupling, we use the averages over two sets of each type as CYANA input and the standard deviation as input error (overall 0.15 Hz). If both data sets of one type (*J*-modulated HMQC spectra) are missing the error was set to 0.3 Hz. The averaged measured values were corrected for the residual dipolar couplings between H^N and H^α due to the natural alignment of GB3 in the magnetic field at 600 MHz. The alignment tensor was estimated from sums of H^N–N residual dipolar and scalar couplings at 500 and 800 MHz fields. The H^N–H^α RDCs were back-predicted from a RDC-refined X-ray structure (Yao et al., 2008a,b; Derrick and Wigley, 1994).

$^3J_{HN,C\beta}$ values are averages over couplings derived from a CT-MQ($^1H^N$, $^{13}C^\alpha$)–HNCA and a HNCA[CB] E.COSY experiment (Vögeli et al., 2007). The individual errors were based on the pairwise rms deviation between these two sets of values, with an overall error of their averaged values of 0.07 Hz (Vögeli et al., 2007). If the value of one data set was missing, 0.1 Hz was used.

$^3J_{HN,C'}$ values are averages over couplings derived from a CT-MQ($^1H^N$, $^{13}C^\alpha$)–HNCA and a HNCA[C′] E.COSY experiment (Vögeli et al., 2007). The individual errors were based on the pairwise rms deviation between these two sets of values, with an overall error of their averaged values of 0.1 Hz (Vögeli et al., 2007). If the value of one data set was missing, 0.2 Hz was used.

The Karplus curve coefficients for $^3J_{HN,H\alpha}$, $^3J_{HN,C\beta}$ and $^3J_{HN,C'}$ were determined from fits to the RDC-refined X-ray structure (Yao et al., 2008a,b; Derrick and Wigley, 1994) under the assumption of uniform fluctuations of 10° of the φ angles (Brüschweiler and Case, 1994). The highly dynamic residues 12 and 40 were excluded, and three ubiquitin residues with positive φ angles, namely residues 46, 60 and 64, were included in the fits, using angles from an NMR-refined X-ray structure (Hu and Bax, 1997). The obtained Karplus curve coefficients (*A*–*C*) for $^3J_{HN,H\alpha}$, $^3J_{HN,C\beta}$ and $^3J_{HN,C'}$ then are (8.754, −1.222, 0.111), (3.693, −0.514, 0.043) and (4.516, −1.166, −0.038) Hz, respectively.

It has been shown that the values for $^3J_{HN,H\alpha}$, $^3J_{HN,C\beta}$ and $^3J_{HN,C'}$ of GB3 can be predicted somewhat better when density functional theory (DFT) calculations are performed on the structure of PDB deposition 2OED than from simple parametrization of the experimental data by Karplus curves (Wang et al., 2013). This demonstrates that some discrepancy between the Karplus curves and experimental data is caused by hydrogen bonding, substituent and electrostatic effects rather than fluctuation of the dihedral angles. These errors are very small for $^3J_{HN,H\alpha}$ and $^3J_{HN,C'}$ and are clearly dominated by the experimental errors (which are used for the width of the flat bottom CYANA potential). The situation is somewhat different for $^3J_{HN,C\beta}$. Here, the error due to the substituent effect is ca. 0.08 Hz. Thus, the errors for $^3J_{HN,C\beta}$ were set to the propagated errors obtained from the individual random experimental errors plus a uniform error of 0.08 Hz. This increases the overall error from 0.07 Hz to 0.11 Hz.

$^3J_{H\alpha,H\beta2}$ and $^3J_{H\alpha,H\beta3}$ values are available from two sources (Miclet et al., 2005; Lian et al., 1992). In reference Miclet et al. (2005), a 3D HBCBCA type experiment provides two independent values, allowing for a cross check. Sample conditions were similar to those of the other measurements (for example, pH 5.6, 50 mM sodium phosphate buffer, 298 K). The systematic error is mostly equivalent to the errors in the extracted peak positions caused by $^1H^\alpha$ transverse relaxation during the S^3CT element and is estimated to be 1 Hz (see Fig. 4 in the Supplemental information in reference Miclet et al. (2005)). The systematic and individual random errors were propagated into an overall error. In reference Lian et al. (1992), the values were obtained from a 2D PE.COSY experiment with an estimated error of 2 Hz. A 100 mM sodium phosphate buffer was used at pH 4.2. Due to the smaller errors and buffer conditions more consistent with ours, the data set of reference (Miclet et al., 2005) is used, which is also more complete. All couplings are in agreement with our previously calculated ensemble except for residues 8 and 52. Residue 8 seems to undergo averaging as indicated by the two nearly identical values of $^3J_{H\alpha,H\beta2}$ and $^3J_{H\alpha,H\beta3}$ close to 7 Hz. These values are used here since a potentially wrong stereoassignment would not have an impact. The couplings of residue 52 are also in disagreement with the data set in reference (Lian et al., 1992), the X-ray structure (Derrick and Wigley, 1994) and our eNOE-based stereospecific assignment (Orts et al., 2013), which all suggest a single rotamer state. On the other hand, the $^3J_{C',C\gamma}$ and $^3J_{N,C\gamma}$ couplings for residue 52 in reference (Vögeli et al., 2013) appear to be slightly averaged over at least two rotamer states. Due to these inconsistencies, the $^3J_{H\alpha,H\beta2}$ and $^3J_{H\alpha,H\beta3}$ couplings of residue 52 are not used here. The substituent-effect-corrected Karplus parametrization (*A*–

Table 1
Restraints used for ensemble calculation of GB3.

Type				#		
eNOEs	Upper/lower distance limit	Total		984		
		Bidirectional	Total		355	
			involving pseudo atom[a]			31
			Else			324
		Unidirectional	Total		568	
			Involving pseudo atom[a]			69
			Else			499
	Upper distance limit	Aromatics[b]			61	
Scalar couplings	Total			225		
	Backbone	Total			147	
		$^{3}J_{HN,H\alpha}$				49
		$^{3}J_{HN,C\beta}$				49
		$^{3}J_{HN,C'}$				49
	Backbone-side chain	Total			78	
		$^{3}J_{H\alpha,H\beta2/3}$				24
		$^{3}J_{N,C\gamma(1/2)}$				27
		$^{3}J_{C',C\gamma(1/2)}$				27
RDCs	Total			1477		
	Backbone	$^{1}D_{HN,N}$	Total		372	
			medium 1			50
			medium 2			54
			medium 3			54
			medium 4			52
			medium 5			54
			medium 6			54
			medium 7			54
		$^{1}D_{C\alpha,H\alpha}$	Total		387	
			medium 1			43
			medium 2			50
			medium 3			49
			medium 4			47
			medium 5			50
			medium 6			49
			medium 7			50
			medium 8			49
		$^{1}D_{C\alpha,C'}$	Total		320	
			medium 2			54
			medium 3			54
			medium 4			50
			medium 5			54
			medium 6			54
			medium 7			54
		$^{1}D_{HN,C'}$	Total		256	
			medium 2			52
			medium 3			51
			medium 4			47
			medium 5			54
			medium 6			52
	Backbone-side chain	$^{2}D_{C\beta,H\alpha}$	medium 1		13	13
	Side chain	$^{1}D_{C\beta,H\beta(2/3)}$	Total		54	
			medium 1			38
			medium 8			16
		$^{1}D_{H\beta2,H\beta3}$	medium 1		11	11
		$^{1}D_{Cm,Hm3}$	Total		64	
			medium 1			32
			medium 8			32
Angular restraints[c]	Total			52		
	Phi					26
	Psi					26

[a] Pseudo methylene atom or pseudo atom for both methyl groups in Val and Leu.
[b] NOEs involving aromatic groups, set to <8 Å.
[c] Only used during initial stage of structure calculation.

C) = (7.23, −1.37, 2.40) is used as proposed for Arg, Asx, Glx, His, Leu, Lys, Met, Phe, Pro, Trp and Tyr in reference (Perez et al., 2001).

We determined $^{3}J_{C',C\gamma}$ and $^{3}J_{N,C\gamma}$ couplings for aromatic residues using the pulse sequences proposed in reference (Hu et al., 1997). $^{3}J_{C',C\gamma(1/2)}$ and $^{3}J_{N,C\gamma(1/2)}$ couplings for Val, Ile and Thr residues were taken from reference (Chou et al., 2003). The Karplus coefficients for the aromatic residues were taken from reference (Perez et al., 2001) proposing (A–C) to be (2.31, −0.87, 0.49/1.29, −0.49, 0.34), and for the methyl bearing residues (2.76, −0.67, 0.19/2.01, 0.21, −0.12) for Thr and (3.42, −0.59, 0.17/2.64, 0.26, −0.22) for Val and Ile as proposed in reference (Chou et al., 2003).

4.1.4. Torsion angle restraints from C^{α} chemical shifts

A φ and ψ dihedral angle restraint list was generated from C^{α} chemical shifts with CYANA (Güntert, 2003, 2009). The allowed ranges were chosen conservatively and are either −200° to −80°

for φ and 40–220° for ψ or, respectively, −120° to −20° and −100–0°. An exception is the C-terminus with −120–80° and −100–60°. The list contains some restraints in the loops (initial calculations produced a violation for residue 49). As the restraints are based on statistical considerations, they may include some incorrect restraints. However, omission of all of them in the structure calculation leads to bad convergence. Therefore, residue 49 is omitted, but other loop residues (10, 13, 22, 47, 48 and 56) are kept during the initial calculation steps.

4.2. Ensemble structure determination protocol

Multi-state ensembles were calculated as previously described but using all input data described above as listed in Table 1 (Vögeli et al., 2013, 2012). In short, a total of 984 upper/lower distance limits were used for the multiple-state ensemble calculation. This data set consists of 355 upper/lower limits that were set equal to the eNOE-derived distances (from bidirectional eNOEs or from highly reliable averages over multiple H^N–H^N data sets) and 568 upper/lower limits that were obtained from adding/subtracting 15% from the eNOE-derived distances (from unidirectional eNOEs), of which 31 and 69 involve a pseudo methylene or a pseudo atom for both methyl groups in Val and Leu. In addition, we used 61 NOEs involving aromatic protons where only an upper distance limit of 8 Å was set. Other input parameters included 225 scalar couplings, 1477 RDCs obtained from 8 alignment conditions and 52 angular restraints from C^α chemical shifts during the initial stages of the structure calculation. The typical number of restraints per residue is 60 and each residue having well above 30 (except for residue 1), when counting those between different residues half. A bar plot of the number of restraints per residue is shown in Fig. S7 in the Supporting material. The CYANA target function term for an observable of type X is $w_X \max(|X_{exp} - X_{calc}| - \mathrm{err}(D)_{exp}, 0)^2$ with a weighting factor that has the unit $Å^2/\mathrm{unit}(X)2$. Note that the root-mean-square deviations (rmsd) between experimental and back-calculated values of left-out restraints give very similar statistics for the cross-validations and are therefore not shown for restraints that were used in the structure calculation.

100 conformers were calculated with 50,000 torsion-angle dynamics annealing steps and the 20 conformers with the lowest target function were then used to represent the calculated structure. Ensembles encompassing 1 to 9 states of the entire protein were calculated simultaneously. Steric repulsion between atoms of different states was excluded, and the eNOE distance restraints were applied to the r^{-6} averages of the corresponding distances in the individual states. The ^3J and the RDC restraints were applied to the arithmetic mean of the quantities in the individual states. Bundling restraints were applied in order to keep the individual structural states together in space as far as permitted by the experimental restraints (Vögeli et al., 2013; Clore and Schwieters, 2004). 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 the NOEs was used. The effective bonds length of H^N–N and H^α–C^α were set to 1.02 and 1.09 Å, respectively. Analogous calculations with 1.041 and 1.117 Å result in nearly identical target functions and ensembles.

4.2.1. Relative weights for annealing

As mentioned above, torsion angle restraints from C^α shifts improve convergence during the CYANA structure calculation (Güntert, 2003, 2009), but are not desired to restrain the final structure due to their statistical rather than well-defined theoretical relationship. Therefore, the weight of the angular restraint contribution to the target function is increasingly reduced to zero during the calculation (Güntert and Bucher, 2015). On the other hand, scalar couplings and RDCs are known to slow down convergence if they are fully active at the initial stage of the calculation due to local extrema. Therefore, their contribution weight to the target function is ramped up starting from zero during calculation. This procedure renders the distance restraints computed from the eNOEs, which do not have individual local extrema, the main factors at the initial stage. The weights of the backbone and side-chain bundling restraints were set one and two orders of magnitude smaller than the one of distance restraints from eNOEs. Interestingly, increasing the weights of the bundling restraints resulted in similar target functions and statistics. Upon reducing the weights by one order of magnitude, the target function was ca. 25 $Å^2$ larger and the rms deviation from the mean more than twice as large. This trend points to a compromised convergence, which could only partially be rescued by doubling the annealing steps or the number of calculated structures.

4.2.2. Principal component analysis and state separation

Principal component analysis (PCA) Vögeli et al., 2012; Bahar et al., 2010 was performed on each of the 20 two-state ensembles calculated with the different input data independently on vectors of the Cartesian coordinates of all C^α atoms following an optimal superposition. Eigenvalue spectrum entropies were obtained by normalizing all strictly positive eigenvalues to 1, and computing the Shannon information. The random reference was computed as the Shannon entropy of the corresponding uniform distribution. All the entropies are normalized by the number of non-zero eigenvalues.

We then assigned all the states in each of the bundles to one of two classes (red or blue) according to the ten first principal components (PCs) independently. These PCs account for more than 80% of the total structural variance of the C^α atoms in all cases. For each of the PCs we then compute a 'separation probability' defined as the number of times the two states were assigned to the same class for each bundle, divided by the total number of ensembles. The values shown in Fig. S5 in the Supporting material are averages obtained from the first ten PCs. The random separation probabilities were computed by performing assignment to 10,000 standard normal samples, and the error bar corresponds to the standard deviation around the mean separation probability.

Acknowledgments

B.V. was supported by the Swiss National Science Foundation with Grant 140214 and the by ETH Research Grant ETH-04 13-1. S.O. is funded by an Independent Postdoc grant from The Danish Council for Independent Research|Natural Sciences (ID: DFF-4002-00151). P.G. gratefully acknowledges financial support by the Lichtenberg program of the Volkswagen Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2015.07.008.

References

Anthis, N.J., Clore, G.M., 2015. Q. Rev. Biophys. 48, 35–116.
Bahar, I., Lezon, T.R., Bakan, A., Shrivastava, I.H., 2010. Chem. Rev. 110, 1463–1497.
Ban, D., Sabo, T.M., Griesinger, C., Lee, D., 2013. Molecules 18, 11904–11937.
Bell, R.A., Saunders, J.K., 1970. Can. J. Chem. 48, 1114–1122.
Bonvin, A.M.M.J., Brünger, A.T., 1995. J. Mol. Biol. 250, 80–93.
Brünger, A., Clore, G.M., Gronenborn, A.M., Saffrich, R., Nilges, M., 1993. Science 261, 328–331.
Brüschweiler, R., Case, D.A., 1994. J. Am. Chem. Soc. 116, 11199–11200.
Brüschweiler, R., Blackledge, M., Ernst, R.R., 1991. J. Biomol. NMR 1, 3–11.

Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M., Ernst, R.R., 1992. J. Am. Chem. Soc. 114, 2289–2302.
Chi, C., Strotz, D., Riek, R., Vögeli, B., 2015. J. Biomol. NMR 62, 63–69.
Chou, J.J., Case, D.A., Bax, A., 2003. J. Am. Chem. Soc. 125, 8959–8966.
Clore, G.M., Schwieters, C., 2004. J. Am. Chem. Soc. 126, 2923–2938.
Clore, G.M., Schwieters, C., 2006. J. Mol. Biol. 355, 879–886.
Constantine, K.L., Mueller, L., Andersen, N.H., Tong, H., Wandler, C.F., Friederichs, M.S., Bruccoleri, R.E., 1995. J. Am. Chem. Soc. 117, 10841–10854.
Derrick, J.P., Wigley, D.B., 1994. J. Mol. Biol. 243, 906–918.
Fenwick, R.B., Esteban-Martín, S., Richter, B., Lee, D., Walter, K.F.A., Milovanovic, D., Becker, S., Lakomek, N.A., Griesinger, C., Salvatella, X., 2011. J. Am. Chem. Soc. 133, 10336–10339.
Fesik, S.W., O'Donnell, T.J., Camp Jr., R.T., Olejniczak, E.T., 1986. J. Am. Chem. Soc. 108, 3165–3170.
Fletcher, C.M., Jones, D.N.M., Diamond, R., Neuhaus, D., 1996. J. Biomol. NMR 8, 292–310.
Güntert, P., 2003. Prog. Nucl. Magn. Reson. Spectrosc. 43, 105–125.
Güntert, P., 2009. Eur. Biophys. J. 38, 129–143.
Güntert, P., Bucher, L., 2015. J. Biomol. NMR. http://dx.doi.org/10.1007/s10858-015-9924-9.
Hu, J.S., Bax, A., 1997. J. Am. Chem. Soc. 119, 6360–6368.
Hu, J.S., Grzesiek, S., Bax, A., 1997. J. Am. Chem. Soc. 119, 1803–1804.
Karplus, M., 1963. J. Am. Chem. Soc. 85, 2870–2871.
Kessler, H., Griesinger, C., Lautz, J., Müller, A., van Gunsteren, W.F., Berendsen, H.J.C., 1988. J. Am. Chem. Soc. 110, 3393–3396.
Kuboniwa, H., Grzesiek, S., Delaglio, F., Bax, A., 1994. J. Biomol. NMR 4, 871–878.
Lian, L.Y., Derrick, J.P., Sutcliffe, M.J., Yang, J.C., Roberts, G.C.K., 1992. J. Mol. Biol. 228, 1219–1234.
Miclet, E., Boisbouvier, J., Bax, A., 2005. J. Biomol. NMR 31, 201–216.
Olsson, S., Boomsma, W., Frellsen, J., Bottaro, S., Harder, T., Ferkinghoff-Borg, J., Hamelryck, T., 2011. J. Magn. Reson. 213, 182–186.
Olsson, S., Vögeli, B., Cavalli, A., Boomsma, W., Ferkinghoff-Borg, J., Lindorff-Larsen, K., Hamelryck, T., 2014. J. Chem. Theory Comput. 10, 3484–3491.
Olsson, S., Ekonomiuk, D., Sgrignani, J., Cavalli, A., 2015. J. Am. Chem. Soc. http://dx.doi.org/10.1021/jacs.5b01289.
Orts, J., Vögeli, B., Riek, R., 2012. J. Chem. Theory Comput. 8, 3483–3492.
Orts, J., Vögeli, B., Riek, R., Güntert, P., 2013. J. Biomol. NMR 57, 211–218.
Ottiger, M., Bax, A., 1999. J. Am. Chem. Soc. 121, 4690–4695.
Perez, C., Löhr, F., Ruterjans, H., Schmidt, J.M., 2001. J. Am. Chem. Soc. 123, 7081–7093.
Sabo, T.M., Smith, C.A., Ban, D., Mazur, A., Lee, D., Griesinger, C., 2014. J. Biomol. NMR 58, 287–301.
Sekhar, A., Kay, L.E., 2013. Proc. Natl. Acad. Sci. USA 110, 12867–12874.
Shapiro, Y.E., 2013. Prog. Biophys. Mol. Biol. 112, 58–117.
Solomon, I., 1955. Phys. Rev. 99, 559–565.
Theobald, D.L., Wuttke, D.S., 2006. Bioinformatics 22, 2171–2172.
Tjandra, N., Bax, A., 1997. Science 278, 1111–1114.
Torchia, D., 2015. Prog. Nucl. Magn. Reson. Spectrosc. 84–85, 14–32.
Ulmer, T.S., Ramirez, B.E., Delaglio, F., Bax, A., 2003. J. Am. Chem. Soc. 125, 9179–9191.
Vögeli, B., 2011. J. Biomol. NMR 50, 315–329.
Vögeli, B., 2014. Prog. Nucl. Magn. Reson. Spectrosc. 78, 1–46.
Vögeli, B., Ying, J., Grishaev, A., Bax, A., 2007. J. Am. Chem. Soc. 129, 9377–9385.
Vögeli, B., Yao, L., Bax, A., 2008. J. Biomol. NMR 41, 17–28.
Vögeli, B., Segawa, T., Leitz, D., Sobol, A., Choutko, A., Trzesniak, D., Van Gunsteren, W., Riek, R., 2009. J. Am. Chem. Soc. 131, 17215–17225.
Vögeli, B., Friedmann, M., Leitz, D., Sobol, A., Riek, R., 2010. J. Magn. Reson. 204, 290–302.
Vögeli, B., Kazemi, S., Güntert, P., Riek, R., 2012. Nat. Struct. Mol. Biol. 19, 1053–1057.
Vögeli, B., Orts, J., Strotz, D., Güntert, P., Riek, R., 2012. Chimia 66, 787–790.
Vögeli, B., Güntert, P., Riek, R., 2013. Mol. Phys. 111, 437–454.
Vögeli, B., Orts, J., Strotz, D., Chi, C., Minges, M., Wälti, M.A., Güntert, P., Riek, R., 2014. J. Magn. Reson. 241, 53–59.
Vögeli, B., Olsson, S., Riek, P., Güntert, P., 2015. Data in brief, compiled data set of exact NOE distance limits, residual dipolar couplings and scalar couplings from GB3 (submitted).
Wang, B., He, X., Merz, K.M., 2013. J. Chem. Theory Comput. 9, 4653–4659.
Yao, L., Bax, A., 2007. J. Am. Chem. Soc. 129, 11326–11327.
Yao, L., Vögeli, B., Torchia, D.A., Bax, A., 2008a. J. Phys. Chem. B 112, 6045–6056.
Yao, L.S., Vögeli, B., Ying, J.F., Bax, A., 2008b. J. Am. Chem. Soc. 130, 16518–16520.