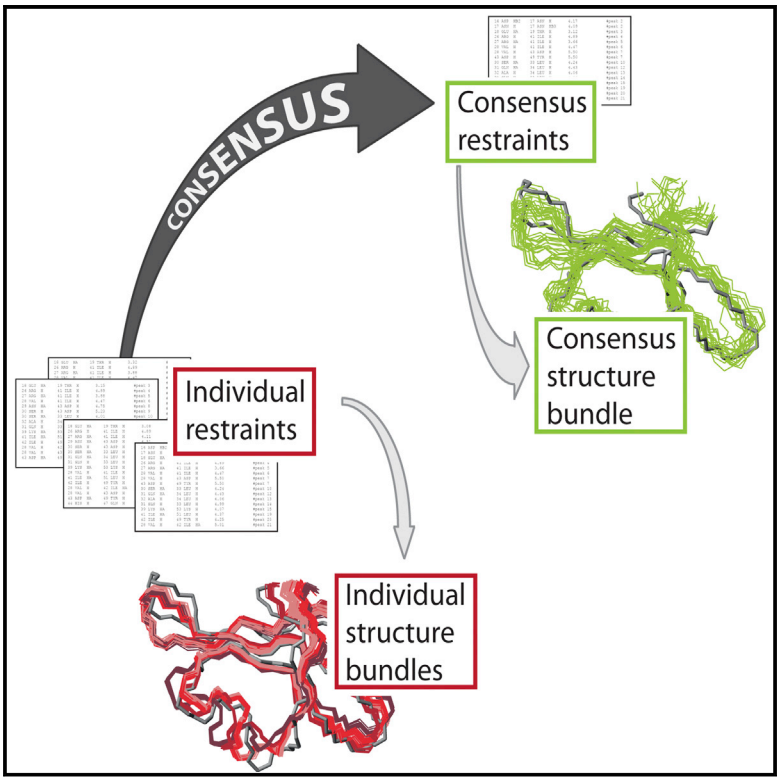
Resource

Structure

Increased Reliability of Nuclear Magnetic Resonance Protein Structures by Consensus Structure Bundles

Graphical Abstract



Highlights

- Protocol for NMR structure determination with automated NOESY assignment
- Consensus structure bundle whose radius is a good measure of structural accuracy
- Accuracy is not overestimated by unreasonably high bundle precision
- The protocol was applied to 4,066 data sets of 18 different proteins

Authors

Lena Buchner, Peter Güntert

Correspondence

guentert@em.uni-frankfurt.de

In Brief

Overestimation of structural accuracy by unrealistically tight structure bundles is a common issue with available methods for NMR structure calculation. Buchner and Güntert develop an approach that minimizes this problem by using consensus restraints from multiple automated NOESY assignment/structure calculation runs to recalculate a consensus structure bundle.

Buchner & Güntert, 2015, Structure *23*, 425–434
February 3, 2015
http://dx.doi.org/10.1016/j.str.2014.11.014

CellPress

Structure
Resource

Increased Reliability of Nuclear Magnetic Resonance Protein Structures by Consensus Structure Bundles

Lena Buchner[1] and Peter Güntert[1,2,3,*]
[1]Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt, Max-von-Laue-Straße 9, 60438 Frankfurt am Main, Germany
[2]Frankfurt Institute of Advanced Studies, Goethe University Frankfurt, Ruth-Moufang-Straße 1, 60438 Frankfurt am Main, Germany
[3]Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachioji, Tokyo 192-0397, Japan
*Correspondence: guentert@em.uni-frankfurt.de
http://dx.doi.org/10.1016/j.str.2014.11.014

SUMMARY

Nuclear magnetic resonance (NMR) structures are represented by bundles of conformers calculated from different randomized initial structures using identical experimental input data. The spread among these conformers indicates the precision of the atomic coordinates. However, there is as yet no reliable measure of structural accuracy, i.e., how close NMR conformers are to the "true" structure. Instead, the precision of structure bundles is widely (mis)interpreted as a measure of structural quality. Attempts to increase precision often overestimate accuracy by tight bundles of high precision but much lower accuracy. To overcome this problem, we introduce a protocol for NMR structure determination with the software package CYANA, which produces, like the traditional method, bundles of conformers in agreement with a common set of conformational restraints but with a realistic precision that is, throughout a variety of proteins and NMR data sets, a much better estimate of structural accuracy than the precision of conventional structure bundles.

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is, besides X-ray crystallography, the technique most widely used to determine 3D structures of macromolecules. NMR structures are typically represented as bundles of conformers, each conformer being the result of a minimization procedure that optimizes the agreement between the 3D structure and the experimental data. The structural ensemble is characterized by its precision representing the positional uncertainty of the atomic coordinates as well as its accuracy, which is a measure of the closeness to the true structure (Spronk et al., 2004; Zhao and Jardetzky, 1994).

Structural precision is commonly quantified by the root-mean-square deviation (rmsd) radius of the structure bundle, i.e., the average rmsd value between the individual conformers and the mean coordinates of the bundle. Structural accuracy can be quantified by the rmsd bias, i.e., the rmsd between the mean coordinates of the structure bundle and a reference structure (or the mean coordinates of a reference structure bundle) that is assumed to represent the "true structure" (Güntert, 1998). Rmsd values are calculated for the atoms in the structured regions of the protein, which can be identified by visual inspection or algorithms such as CYRANGE (Kirchner and Güntert, 2011).

Experimental data are provided in the form of structural restraints, the most common being distance restraints from nuclear Overhauser effect spectroscopy (NOESY) experiments as well as angular restraints, from, e.g., chemical shift analysis with the program TALOS+ (Shen et al., 2009). The conversion of NOESY peaks into distance restraints requires the assignment to atom pairs and the calibration of peak intensities into upper distance limits. This NOESY assignment is crucial for the outcome of a structure calculation and errors can have severe consequences for the quality of the resulting structure (Jee and Güntert, 2003). It should thus be performed as objectively as possible. Several software tools, therefore, combine automatic NOESY peak assignment and structure calculation in an iterative way (Herrmann et al., 2002; Huang et al., 2006; Rieping et al., 2007).

It has been shown that even a very small number of incorrect distance restraints can lead to a highly precise but completely misfolded protein structure that is not always recognized as such (Nabuurs et al., 2006). This can be attributed to the lack of an independent and reliable measure of NMR structure quality. Instead, some NMR spectroscopists tend to compare the precision of a structure ensemble with the X-ray resolution and use it as a measure of the structure quality. Consequently, there is a widespread ambition to improve the precision, i.e., to minimize the rmsd radius of structure bundles in the belief that it will increase the quality of the structure. This misconception is the cause of a widely observed overestimation of NMR structure accuracy (Spronk et al., 2003, 2004), which limits the reliability of NMR structures without further validation. Currently existing validation tools can reveal errors but do not always guarantee a reliable result. They perform especially well when validating completely misfolded protein structures (Nabuurs et al., 2006). However, deviations in the range of 2–3 Å rmsd bias from the reference structure are not likely to be recognized, although they occur much more frequently than severely erroneous structures (Saccenti and Rosato, 2008). There have also been attempts to combine various validation measures into an estimate

of structural accuracy; e.g., the rmsd from the true structure can be estimated by a linear combination of (suitably normalized) validation parameters (Bagaria et al., 2012), or, similarly, an "equivalent resolution" can be obtained from multiple validation scores (Bagaria et al., 2013; Laskowski et al., 1996). However, while over a large number of different protein structures there is a visible correlation between these accuracy estimates and the true accuracy, the predictive power for a given, single NMR structure determination remains limited.

The precision of a structure bundle directly relates to the amount of meaningful long-range distance restraints, i.e., the information content of a restraint set (Nabuurs et al., 2003). The most severe possible problem of using distance restraints for structure calculation is the bias resulting from erroneous NOESY peak assignments. Potential error sources include the sequence-specific resonance assignments, the identification of true NMR signals, the assignment of NOESY peaks to atom pairs, and the calibration of upper distance limits. An incorrectly assigned nuclear Overhauser effect (NOE) can distort a structure, whereas a too tight but correctly assigned NOE will have a much smaller impact. Although automation increases the reproducibility and reduces the bias originating from subjective user choices, the algorithms are still not perfect, especially in cases of limited data quality, e.g., signal overlap, low signal-to-noise ratios, or sparse data. Iterative combined automated NOESY peak assignment and structure calculation with CYANA can converge toward misfolded protein structures that are not immediately recognized as such because of their high bundle precision (Jee and Güntert, 2003). Similar problems can also arise with other algorithms for automated NOE assignment (Rosato et al., 2012). This occurs predominantly in cases where the protein fold is poorly defined in early cycles of the calculation and subsequent NOESY peak assignments in the following cycles are based on incorrect assumptions about the protein fold. Errors are only rarely reflected in lower precision of the final structure calculation result, but rather in a highly precise but inaccurate protein structure.

Since the outcome of a structure calculation in such cases depends partly on the random initial structures, structure calculations based on the same set of experimental data but different random starting structures converge potentially to different structure bundles. The degree of deviation strongly depends on the data quality. However, some extent of deviation is observed on a regular basis even when using input data of good quality, indicating that precision significantly exceeds accuracy. Despite this fact, in general only one final structure calculation is performed and its results are reported, although a different solution, obtained with a different random number generator seed, would represent the NMR data equally well. Many structure bundles determined by NMR spectroscopy thus have a precision that overestimates accuracy and distortions can remain unrecognized when using simply the bundle precision and the agreement between structure and experimental data as a measure of quality.

Several attempts to solve this problem have been conducted. Spronk et al. (2003) have developed a tool that maximizes the rmsd radius while maintaining the agreement with the experimental data and the geometric quality. This approach improves the sampling of structures within a given set of distance restraints. However, structural distortions due to erroneous distance restraints are not addressed. Since the set of distance restraints remains unchanged, the method does not necessarily improve the accuracy of a protein structure. Inferential structure determination was introduced as a fundamentally different approach to structure determination by NMR spectroscopy (Rieping et al., 2005). The method uses Bayesian inference to derive an objective probability distribution to evaluate the structural ensemble that is generated based on a Monte Carlo simulation. It is independent of empirical parameter estimates and increases the completeness of the sampling of conformational space that is in agreement with the experimental data.

In this article, we introduce a protocol for combined automatic NOESY peak assignment and structure calculation that provides a solution to the problem of overestimated bundle precision. The protocol aims at yielding protein structures for which the bundle precision is a reliable measure of the structural accuracy and where the structure bundle covers well the conformational space that is allowed by the experimental data.

RESULTS AND DISCUSSION

A schematic overview of the algorithm implemented in the CYANA software package (Güntert, 2009; Güntert et al., 1997) is given in Figure 1. The method first performs 20 independent runs of combined automated NOESY assignment and structure calculation using the standard CYANA automatic structure determination procedure with the same input data. Each run starts from a different set of random structures (Figure 1A), comprises seven cycles and a final structure calculation, and yields a final structure bundle as well as the corresponding set of distance restraints. Because the NOESY peaks are assigned independently in each of the 20 runs, the sets of distance restraints from each run in general differ from each other. From each run, the conformer with the lowest CYANA target function value is collected to form a new bundle of 20 conformers, to which we refer in the following as the combined structure bundle (Figure 1B). A further crucial step is to combine the individual sets of distance restraints in order to obtain a consensus set of distance restraints including assignments from all individual runs, which is then used to recalculate the final protein structure bundle, again composed of 20 conformers (Figure 1C), which we refer to as the consensus bundle. This final structure calculation is a simple standard CYANA structure calculation without automatic NOE assignment. In contrast to the combined bundle, all conformers in the final consensus bundle are optimized against a single "consensus" set of distance restraints. In the following, we show that (1) the rmsd radius of the consensus bundle provides a good measure of structural accuracy, and (2) the conformers of the consensus bundle fulfill the consensus set of distance restraints better than the conformers of the initial individual runs fulfill their corresponding sets of distance restraints.

We evaluated the method using NMR data sets of eight different proteins provided by the Northeast Structural Genomics Consortium (NESG) for the CASD-NMR project in 2011–2012 (Rosato et al., 2009, 2012). Two sets of unassigned NOESY peak lists were available for each protein, one containing manually refined NOESY peaks lists and one containing raw

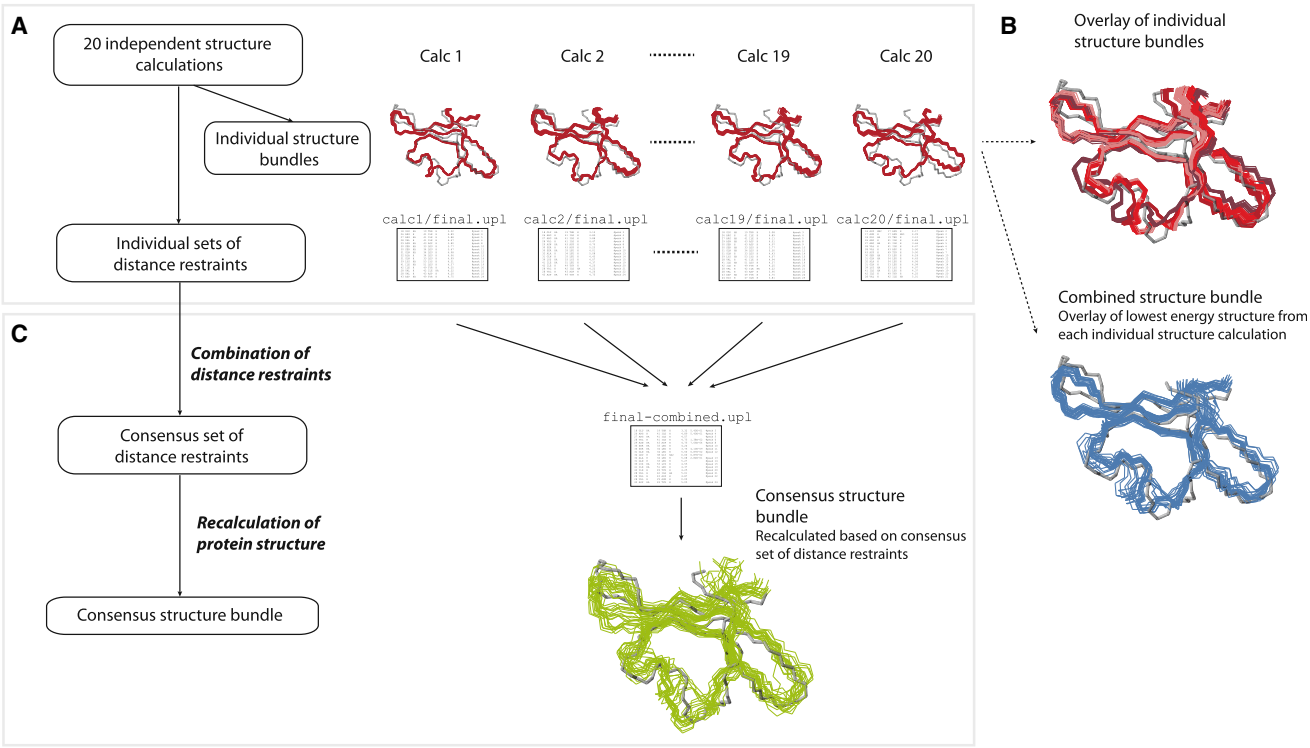


Figure 1. Schematic Overview of the Algorithm Implemented in the CYANA Software Package

(A) Twenty independent runs of the standard CYANA automatic NOESY assignment and structure calculation procedure using the same input data, each starting from a different set of random structures.

(B) Overlay of structure bundles from independent runs (top) and combined structure bundle consisting of the lowest target function structure of each run (bottom).

(C) Combination of all individual sets of distance restraints and recalculation of the protein structure using the consensus set of distance restraints yields the consensus structure bundle.

NOESY peak lists from an early stage of spectral analysis. Structure calculations were performed using the standard combined automatic NOESY peak assignment and structure calculation procedure from the CYANA software package and our method. The CASD-NMR data set was especially well suited for the present study because of the large variability of input data quality and subsequent considerable variety among the different structure calculation results. Despite large differences in structural accuracy, ranging from totally correct to severely erroneous structures, all structure bundles calculated by the standard CYANA approach exhibited a high precision. The test data set thus represents well the aforementioned problem of overestimated bundle accuracy.

In order to evaluate the reliability of a structure calculation result, we calculated the rmsd to the mean structure of the bundle (rmsd radius) representing the precision as well as the rmsd to the reference structure (rmsd bias) as a measure of the accuracy. A structure bundle is considered reliable if precision and accuracy are in good agreement and the reference structure is thus included in the structure bundle. The structure quality can then be estimated solely based on the bundle precision, which is useful in cases where no reference structure is available. We compare the reliability of protein structures determined by the conventional structure calculation procedure with the results from our method.

As an example, Figure 2 shows structures of the protein StT322 (only the ordered parts in the reference structure) calculated from raw peak lists using the standard structure calculation procedure (Figure 2A) as well as an overlay of four selected structure bundles from independent structure calculations using the same procedure (Figure 2B). The reference structure is presented in gray for comparison. Figure 2A clearly shows the completely incorrect global fold of the protein structure when superimposed onto the reference structure. The error is also reflected by the high rmsd bias with respect to the reference structure of 5.78 Å. Nonetheless, the structure bundle is very tight and well defined (precision measured by an rmsd radius of 0.07 Å), illustrating clearly the misconception of precision being related to structural quality. The overlay of four selected structure bundles (Figure 2B), each being the result of the same structure calculation using different random start structures for the minimization procedure, shows large deviations among the structures, indicating clearly that one individual structure calculation result is not fully representing the data set. The average accuracy is 6.75 Å, whereas the average precision is 0.17 Å. The result of our method is presented in Figures 2C and 2D. Figure 2C shows the combined structure bundle and Figure 2D the consensus structure bundle based on the consensus distance restraints. The rmsd to the mean structure increases to 4.73 Å and 4.61 Å, respectively, indicating

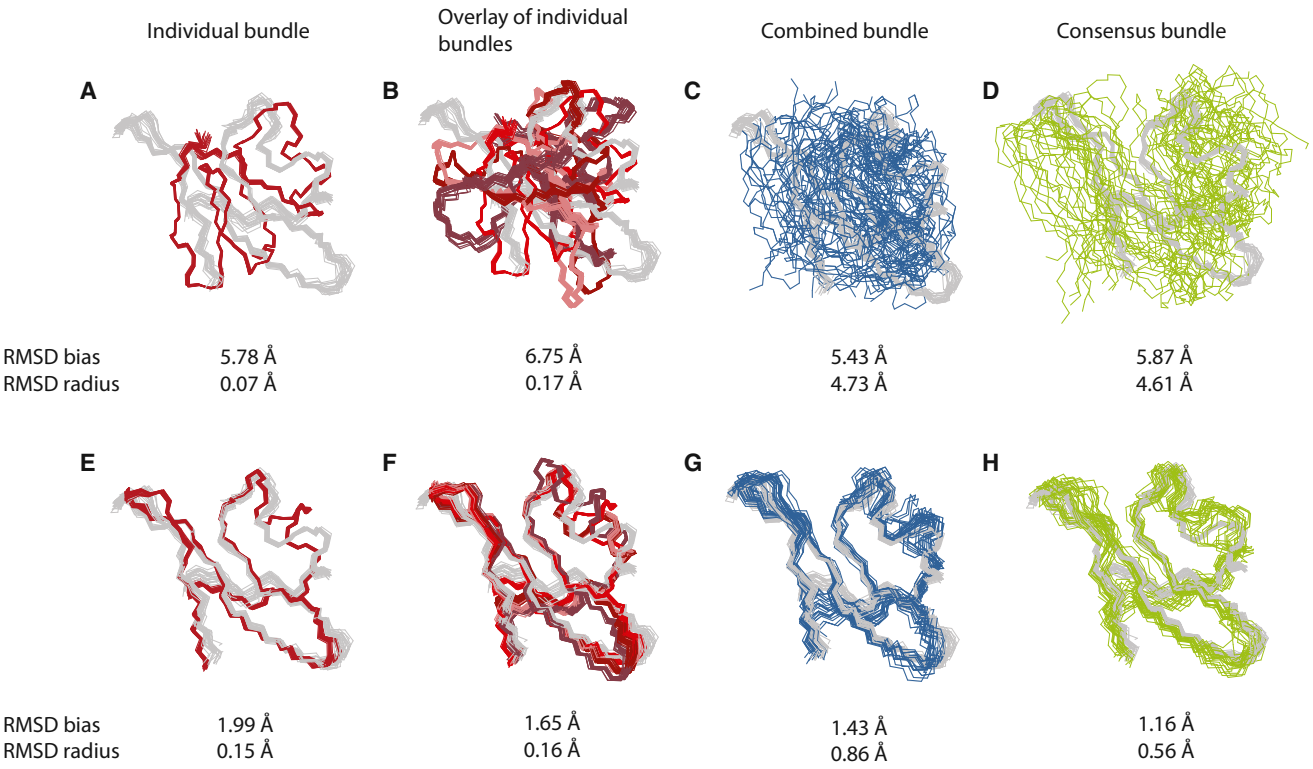


Figure 2. Structures of the Protein StT322 Calculated Using the Classic Automatic Structure Calculation Procedure and Our Method

(A and E) The structure bundle calculated using the classic CYANA automatic NOESY assignment and structure calculation protocol is shown in red.

(B and F) An overlay of four independent structure calculation results based on the same input data but different random starting structures using the classic structure calculation protocol is depicted in red.

(C and G) The combined structure bundle consisting of the lowest target function structure of each of the 20 individual structure bundles is shown in blue.

(D and H) The consensus structure bundle based on the consensus set of distance restraints is presented in green.

In all diagrams the reference structure is shown in gray for comparison. Structures were superimposed for optimal fit of the backbone atoms of the ordered residues 23–63. In (A–D), raw peak lists were used as input data. In (E–H), refined peak lists were used as input data.

a complete lack of any common structural elements among the individual structures. This result furthermore illustrates well that the recalculated consensus structure bundle closely resembles the combined structure bundle. The rmsd bias in both cases remains similar to the average bias of the individual calculations (5.43 Å and 5.87 Å). Both structure bundles represent well the variety of individual structure calculations depicted in Figure 2B and show that the structure calculation actually did not converge to a unique fold. This example illustrates well that our method yields a structure bundle in which the overestimation of accuracy is dramatically reduced and the precision is a faithful measure of the data quality. These results hold for both the combined structure bundle and the consensus structure bundle, but only the latter is calculated from a single set of conformational restraints such that all its conformers fulfill the same restraints.

The first example of Figures 2A–2D is based on experimental input data of very low quality, and the incorrect global fold of a single structure calculation could be identified by the majority of validation tools (Figure S1 available online; Table S1). Figures 2E–2H show the structure calculation results for the same protein using manually refined peak lists. The individual structure calculation result depicted in Figure 2E shows an overall correct global fold. However, the reference structure is again not included in the structure bundle. The accuracy of the individual structure calculation result is 1.99 Å, indicating a correct global fold but a not very accurate local structure. Again the precision of 0.15 Å overestimates the accuracy considerably. In contrast to the aforementioned example, the accuracy is in a range where currently available validation software does not produce reliable results, as can be seen from the very limited correlation between the rmsd bias and the validation parameters of different validation software tools (Figure S1; Table S1). The overlay of several structure bundles (Figure 2F) shows significant deviations among the individual structure bundles. All of these individual results represent equally well the experimental data when evaluating the final CYANA target function (data not shown). The consensus structure bundle from our method is depicted in Figure 2H. The accuracy of 1.16 Å is in the same range as the average accuracy of the 20 individual calculations. However, the rmsd radius increases almost 4-fold from 0.15 Å to 0.56 Å. Visual inspection as well as the evaluation of rmsd values clearly shows the more complete representation of the experimental data by the new structural ensemble where the reference structure is included in the structure bundle. The consensus structure bundle again represents well the combined structure bundle presented in Figure 2G. The example of Figure 2 illustrates well that

Table 1. Structure Calculation Results

		Individual Bundles (Average)				Combined Bundle		Consensus Bundle		
Protein	Data Type	Rmsd Radius Cycle 1 (Å)	Rmsd Radius (Å)	Rmsd Bias (Å)	Target Function (Å^2)	Rmsd Radius (Å)	Rmsd Bias (Å)	Rmsd Radius (Å)	Rmsd Bias (Å)	Target Function (Å^2)
HR2876B	raw	0.70	0.10	1.04	7.54	0.58	0.87	0.40	0.88	1.31
	refined	0.60	0.25	0.62	1.24	0.43	0.53	0.49	0.75	0.52
HR5460A	raw	3.16	0.32	1.82	13.9	0.98	1.54	0.80	1.21	1.94
	refined	1.00	0.38	1.32	9.21	0.80	1.13	0.74	1.15	1.60
HR6430A	raw	0.52	0.19	0.67	5.84	0.47	0.53	0.55	0.92	4.94
	refined	0.48	0.21	0.66	5.37	0.42	0.57	0.45	0.79	4.77
HR6470A	raw	0.48	0.29	0.56	0.32	0.38	0.5	0.39	0.67	0.28
	refined	0.46	0.29	0.46	0.29	0.29	0.42	0.38	0.50	0.28
OR135	raw	0.58	0.19	0.91	0.28	0.50	0.77	0.42	0.90	0.00
	refined	0.48	0.17	0.69	0.48	0.39	0.58	0.32	0.67	0.01
OR36	raw	0.98	0.23	1.17	4.59	0.76	0.9	0.63	0.94	0.06
	refined	0.89	0.42	0.91	0.42	0.64	0.81	0.71	0.98	0.01
StT322	raw	5.52	0.17	6.75	14.94	4.73	5.43	4.29	6.28	1.16
	refined	0.80	0.16	1.65	0.52	0.86	1.43	0.60	1.20	0.09
YR313A	raw	3.44	0.46	2.04	1.40	1.62	1.50	1.07	1.55	0.47
	refined	1.44	0.69	1.69	0.47	1.52	1.33	0.83	1.32	0.35

The rmsd radius of a structure bundle is the average rmsd value between the individual conformers and the mean coordinates of the bundle. It characterizes the precision. For individual bundles, the first column reports the rmsd radius for the structure obtained in the first cycle of automated NOE assignment and structure calculation, and the second column reports the rmsd radius for the final structure bundle. The rmsd bias is the rmsd between the mean coordinates of the structure bundle and a reference structure (or the mean coordinates of a reference structure bundle). rmsd radius and rmsd bias characterize the precision and accuracy of a structure bundle, respectively. rmsd values are calculated for the backbone atoms N, C^{α}, C' in the structured regions of the protein (see Experimental Procedures).

our structure calculation method is beneficial in situations with input data of good and bad quality alike, since in both cases the reliability can be increased significantly.

To evaluate the reliability of all structure calculation results with the different proteins and experimental data sets from CASD-NMR, we calculated the ratio between accuracy (rmsd to the reference structure) and precision (rmsd to the mean structure). Ideally, this accuracy-to-precision ratio should be 1 in order to be able to estimate the structure quality solely based on the bundle precision, with values above 1 indicating that the apparent precision overestimates the accuracy. All rmsd values that were used to calculate the ratios between accuracy and precision are given in Table 1. Figure 3 shows the results for the individual structure calculations using the standard approach (red), the combined structure bundle based on 20 individual structure bundles (blue), and the recalculated structure bundle based on the consensus set of distance restraints (green). The results are presented for eight different proteins and two data sets for each protein.

For the conventional individual structure calculations, the accuracy-to-precision ratio (averaged over the 20 individual runs with each data set) shows a very large variability in the range between 1.6 and 39.7 among the different proteins and data sets. One protein (HR6470A) has comparatively low ratios in the range between 1.6 and 1.9 and thus represents an exception among these test proteins. Rather low values were also observed for the proteins OR36 (2.2), YR313A (2.4), and HR2876B (2.5) when using refined peak lists. All other proteins show higher ratios, indicating a considerable overestimation of accuracy by the bundle precision when using the conventional structure calculation procedure. For most proteins, the ratio decreases when optimizing the experimental input data. However, even when using highly correct input data in the form of manually refined peak lists, the ratios range between 1.6 and 10.3 (average 3.7).

The combined structure bundle contains the lowest energy structure from each individual structure calculation, and thus represents a large part of the conformational space that can be explained by the given input data set. The ratio between accuracy and precision (Figure 3, blue; values in the range between 0.9 and 1.7) decreases considerably in all cases when compared with the individual calculations. This finding clearly shows the beneficial effect of repeating the same structure calculation several times, making the calculation result more reliable and enabling the use of bundle precision as a direct measure of the structural accuracy. One exception is again the protein HR6470A, for which no significant difference is observed in the ratios between the individual structure calculations and the combined structure bundle. This is the only example whereby a single structure calculation already resulted in an accurate and reliable structure bundle, which could, however, be unrecognized if the reference structure is unknown.

The essential part of our method is the combination of the individual restraint data sets to obtain a single consensus set of distance restraints representing the entire conformation space allowed by the input peak lists. The ratios for the consensus structure bundles based on the combined set of distance restraints are shown in Figure 3 (green). In general, the values

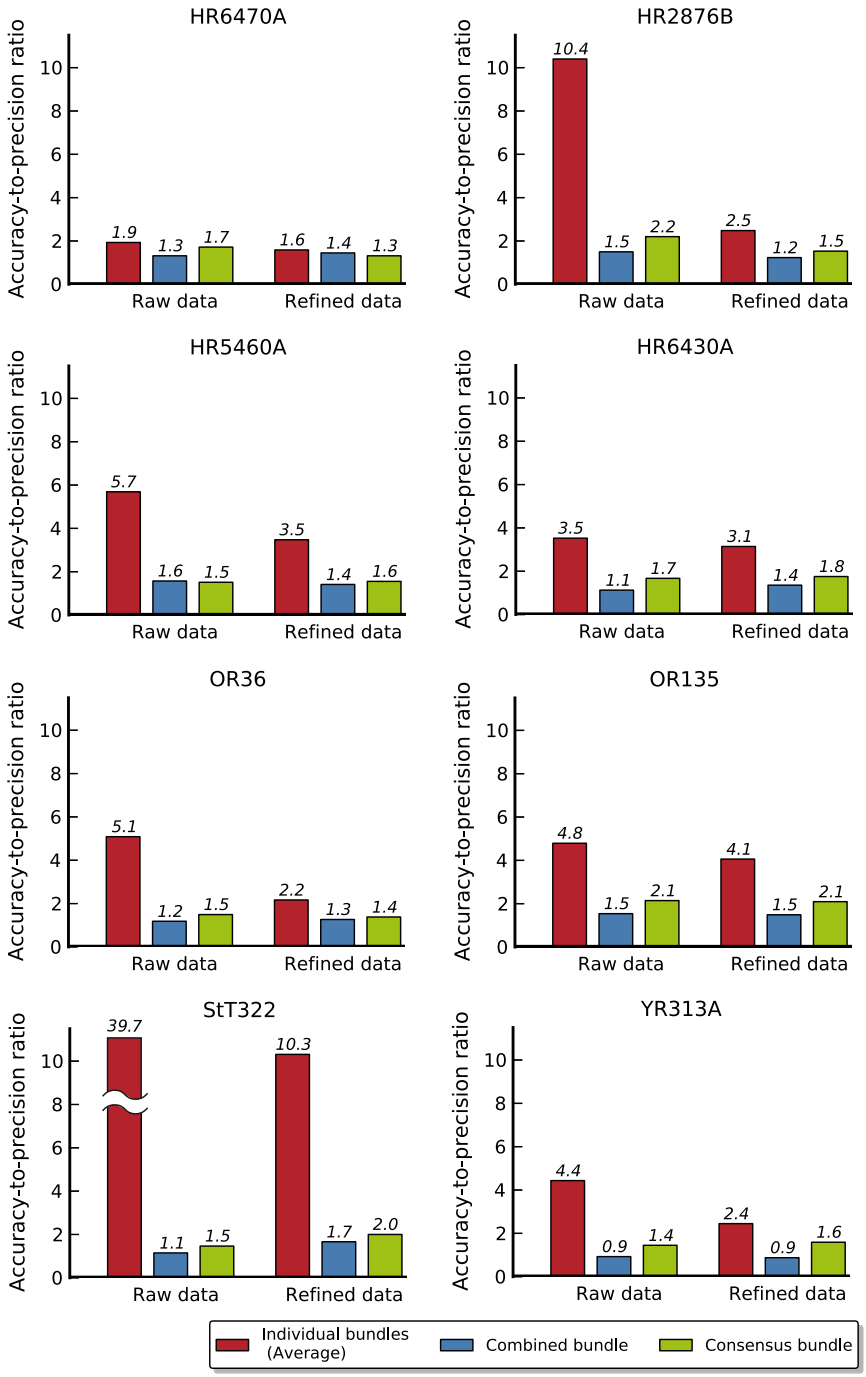


Figure 3. Accuracy Overestimation by Structure Bundle Precision

The degree of overestimation is quantified by the accuracy-to-precision ratio, where the structural accuracy is given by the rmsd between the mean coordinates of a structure bundle and the NMR reference structure from the PDB, which was determined and refined by experienced scientists; and the precision of a structure bundle is given by the average rmsd of the individual conformers of the structure bundle to its mean coordinates. Only ordered residues (see text) were used for rmsd calculation. The accuracy-to-precision ratio is presented for eight proteins from the CASD-NMR project and two different data sets for each protein (i.e., raw peak lists and refined peak lists). Results are given for the classic structure calculation process as the average from 20 independent runs (red), for the combined structure bundle consisting of the lowest target function conformer from each of the 20 independent runs (blue), and the consensus structure bundle based on the consensus set of distance restraints (green).

are very similar to those of the combined structure bundle. This shows that the consensus structure bundle and the corresponding combined set of distance restraints are as well suited to represent the experimental data as the combined structure bundle. Due to the significantly increased reliability, we recommend the use of the consensus structure bundle and the corresponding combined restraints when presenting structure bundles determined by NMR spectroscopy. The precision of the consensus structure bundle can be used as a strong measure of the data quality, and thus be compared with the resolution in X-ray crystallography.

In order to investigate reproducibility, we applied our method to a second test data set comprising ten proteins including various types of simulated data imperfections for each protein, i.e., randomly deleted chemical shifts, randomly modified chemical shifts, deleted NOESY peaks, etc. These modifications resulted in a total of 4,050 restraint data sets covering a very large range of input data quality. A description of the data sets is given in the Experimental Procedures, and more details will be published elsewhere. For every structure calculation, the overestimation ratio between the rmsd to the reference structure and the rmsd to the mean structure was analyzed and plotted as a histogram for the individual structure bundles (Figure 4, top), the combined structure bundle (Figure 4, center), and the consensus structure bundle (Figure 4, bottom). The histograms show clearly that the overestimation ratios are in general significantly higher in the case of individual structure bundles (i.e., 71% of the structure bundles have ratios above 2.0) with a median ratio of 2.9, whereas the median ratio for the combined structure bundles is 1.2 (5% of ratios above 2.0) and for the consensus structure bundles 1.4 (15.8% of ratios above 2.0). These results are in very good agreement with those from the CASD-NMR data set. Due to the large amount of structure calculations and the large variety of input data qualities resulting in large variations among the calculated structures, it can be concluded that the presented method works reproducibly and can thus be applied routinely.

The CYANA target function measures the agreement between the structure bundle and the experimental restraints and is

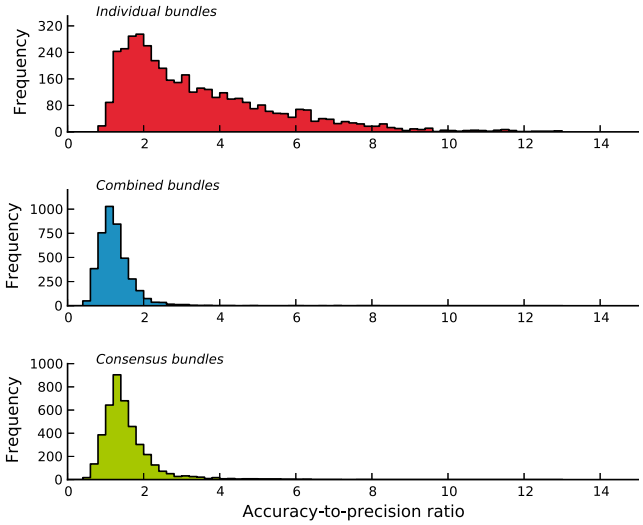


Figure 4. Accuracy-to-Precision Ratios for 4050 NMR Data Sets

Frequency distributions of the accuracy-to-precision ratios are given for the classic structure calculation process as the average from 20 independent runs (red), for the combined structure bundle consisting of the lowest target function conformer from each of the 20 independent runs (blue), and the consensus structure bundle based on the consensus set of distance restraints (green). The test data consist of 4050 solution NMR data sets from ten different proteins. It includes for each protein the original experimental data set and modifications that simulate a large variety of data imperfections (see Experimental Procedures).

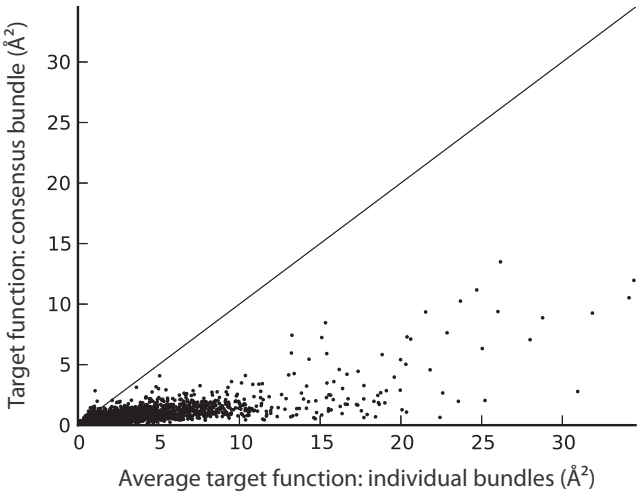


Figure 5. Target Function Values for 4050 Conventional and Consensus Structure Bundles

Each data point correlates the average target function value for a consensus structure bundle with the average target function value for the conformers of the corresponding conventional structure bundles. The CYANA target function measures the agreement between the structure bundle and the experimental and steric conformational restraints from which it was calculated. It is defined such that it is zero if all conformational restraints are fulfilled.

defined such that it is zero if all restraints are fulfilled. High target function values indicate problems during the structure calculation and need to be avoided by closer inspection of the experimental data. In order to show that our method yields distance restraints that are still fulfilled by the recalculated structure bundle, Figure 5 compares the target function of the consensus structure bundles with the average final target function of the respective 20 individual structure calculations. The target function values of the consensus structure bundles are in almost all cases lower than those of the individual calculations. This shows that no inconsistencies or convergence problems result from using the consensus set of distance restraints.

The traditional criterion for evaluating the outcome of a CYANA structure calculation with automated NOE assignment is that an rmsd radius of less than 3 Å in the first cycle of automated NOE assignment and structure calculation is indicative of a final structure with low rmsd bias (Herrmann et al., 2002). Cycle 1 rmsd radii above 3 Å indicate that the resulting final structure may (but does not have to) be inaccurate. Therefore, the cycle 1 rmsd radius is not a direct measure of accuracy but rather provides a criterion to recognize potentially unreliable calculations. For comparison, Table 1 includes the cycle 1 rmsd radii and Figure S4 shows the correlation between the rmsd radii of the cycle 1 structure bundles and the consensus structure bundles.

Conclusions

We present a method for combined automated NOE assignment and structure calculation implemented using the software package CYANA. The principal advantage of our method over simply repeating full calculations (referred to as combined structure bundles in this article) is that all conformers of the consensus structure bundle are calculated from the same restraint data, i.e., the consensus restraint list, in a single CYANA structure calculation. In the case of repeated full calculations, each calculation will lead to somewhat different NOESY peak assignments and restraints. Hence, the resulting structures will in general not fulfill a single set of restraints. This can be problematic if, e.g., a combined structure bundle is submitted to the Protein Data Bank (PDB) along with the restraints from one of its individual NOE assignment/structure calculations, because in this case a later evaluation of the agreement between the coordinates and the conformational restraints in the PDB will in general show additional restraint violations that have not been reported in the original publication. In contrast, we propose to deposit in the PDB the consensus structure bundle together with the consensus restraint list from which the consensus structure bundle was computed. NOESY peak lists containing the consensus peak assignments can also be produced by the program.

We have tested our method using optimized and raw input peak lists of eight different proteins provided by the CASD-NMR project in 2011–2012 as well as a data set based on ten different proteins including various simulated data imperfections. We have measured the reliability of structure bundles as the ratio between accuracy (rmsd to the reference structure) and precision (rmsd to the mean structure) and have compared the results from the classical structure calculation procedure with the results from our method. The results clearly show that our protocol for automatic structure calculation produces very reliable structure bundles where the precision can be used as a very good indication for the structure quality without having any prior information about the correct protein fold. It should be noted that the precision of the consensus structure bundles

is not strictly equal to the accuracy but proportional with a median proportionality factor of 1.4 (Figure 3). For a conservative estimate, an upper bound on the accuracy, given by the rmsd bias, can be approximated as twice the precision, given by the rmsd radius, of the consensus structure bundle. The precision of the consensus bundle gives an estimate of the input data quality; however, additional criteria such as the assignment completeness of the assigned consensus peak lists as well as the average ambiguity of the latter can be used to assess experimental uncertainties (e.g., a large number of discarded peaks and high ambiguity indicate inconsistencies within the input data).

Our method is helpful for input data optimization in the course of NMR structure determinations, and we recommend it especially for routine use in the final structure calculation, since the consensus bundle reflects the experimental data much better.

EXPERIMENTAL PROCEDURES

Generation of Consensus Distance Restraint Set

The algorithm performs 20 independent automated NOESY assignment and structure calculation runs using the same input data and different random number generation seeds, resulting in 20 individual structure bundles. The lowest energy structure of each of these 20 structure bundles is combined to obtain a new combined structure bundle. The precision of the combined structure bundle is a measure of the extent to which individual calculations differ from each other.

Each of the 20 individual structure calculations leads to a different set of distance restraints as a result of the seven cycles of NOE assignment and structure calculation. These individual final sets of distance restraints are in optimal agreement with the respective structure bundle; however, do not represent the aforementioned combined structure bundle. The combination of the individual sets of distance restraints yields a consensus set of distance restraints that represents the combined structure bundle and thus results in a structure bundle similar to the combined structure bundle when used as input for a further structure calculation. This final structure calculation is a simple standard CYANA structure calculation without automatic NOE assignment. It uses the consensus NOE distance restraints (and other conformational restraints, if available) as input and yields the consensus structure bundle as output.

The combination makes use of the fact that each distance restraint is the result of a NOESY peak assignment. During the seven cycles of NOESY peak assignment and structure calculation, peaks may have unambiguous assignments, ambiguous assignments, or remain unassigned. In the final structure calculation, all remaining ambiguities are resolved and nonstereospecifically assigned methyl- or methylene-protons are treated by symmetrization and pseudo-atom correction (Güntert et al., 1991). During the combination process, every distance restraint assignment originating from the same peak in all individual restraint sets is combined to obtain one ambiguous (or unambiguous) restraint.

Individual peaks may be assigned to different atom pairs in different structure calculation runs or may remain unassigned in individual calculations. To form a consensus distance restraint data set that is suitable for recalculating the consensus structure bundle, it is necessary to choose only those restraints that represent the combined structure bundle in a sufficient manner. Consequently, restraints are only chosen if the corresponding peak could be assigned (with any assignment[s]) in a specified minimal number of individual structure calculations, otherwise the complete peak will be discarded. If a restraint is chosen, then all atom pairs appearing in any of the respective peak assignments of the individual structure calculations are combined to obtain one ambiguous or unambiguous distance restraint. The threshold on the minimal number of individual structure calculations in which a peak must be assigned in order to be chosen for consensus restraint generation can be chosen by the user; however, after having tested the complete range of cutoff values, we recommend a threshold of 60% of the individual structure calculations in which a peak needs to be assigned to any atom pair in order to be selected. Higher threshold values lead in a few cases to an unacceptably large loss of information by a very large number of discarded peaks, resulting in a severe underestimation of the achievable accuracy. Low threshold values, on the other hand, again increase the apparent precision, due to a large number of restraints that are selected, even though they represent only a small fraction of the conformers in the combined bundle.

Our choice of 60% for the peak selection cutoff percentage can be rationalized from Figure S2, Table S2, and Figure S3. Overall, the results depend only weakly on the choice of the cutoff percentage. On the one hand, increasing the cutoff value slightly decreases the accuracy-to-precision ratio toward the ideal value of 1 (Figure S2, left panels). On the other hand, lowering the cutoff increases the occurrence of structures with high accuracy (low rmsd to reference; Figure S2, right panels). The median accuracy-to-precision ratios and accuracy values at different cutoff values are summarized in Table S2, which shows that a cutoff of 0.6 provides a good compromise between the two opposite trends. Figure S3 shows the precision and accuracy as a function of the cutoff value for two examples from the CASD-NMR data set. In the first example, the results are almost independent from the cutoff, whereas in the second example there is a loss of accuracy with increasing cutoff.

Combined structure bundles with low precision generally show large differences among the NOESY peak assignments from the individual runs. This in turn results in an increased ambiguity of the restraints as well as a larger number of discarded peaks in the combined data set. Altogether this reduces the information content of the combined restraint set, which in turn decreases the precision of the consensus structure bundle. Combined structure bundles with high precision, on the other hand, have very similar individual distance restraint sets, which leads to the fact that most of the restraints are selected and have low ambiguity in the consensus restraint set. The combined restraint list has thus more information content and, consequently, the consensus structure bundle will have high precision.

The method generates a consensus set of distance restraints that essentially reproduces the combined structure bundle when used in a conventional structure calculation based on distance restraints. This is achieved because the precision of a structure bundle depends on the information content of the data set, which in turn is determined by the amount of meaningful long-range restraints as well as their ambiguity.

Individual Structure Calculations

Individual structure calculations are performed using the standard structure calculation procedure of the CYANA software (Güntert, 2009; Güntert et al., 1997; Herrmann et al., 2002). The protein sequence as well as chemical shifts and unassigned peak lists from NOESY spectra are used as input for the structure calculations. Chemical shift assignments were taken from the Biological Magnetic Resonance Bank (BMRB), whereas torsion angle restraints were taken from the PDB. The chemical shift tolerance for NOESY peak assignments was set to 0.03 ppm for ^{1}H, and 0.3 ppm for ^{15}N and ^{13}C. The standard CYANA protocol was applied using 200 random starting structures and 15,000 annealing steps during torsion angle dynamics. The 20 structures with the lowest target function values were used as the final structure bundle. Details are described in Schmidt and Güntert (2013).

Data Sets from CASD-NMR

In order to evaluate the advantages of consensus set distance restraints and consensus structure bundles for the reliability of protein structure calculations, we used input data of eight different proteins that were provided as test data sets for the CASD-NMR project in 2011–2012 (Rosato et al., 2009, 2012). The same data set had already been used for the analysis of automatic chemical shift assignment based solely on NOESY spectra using the FLYA automated resonance assignment algorithm and subsequent structure calculations (Schmidt and Güntert, 2013). The present data set was of particular interest for our study since the structure calculation results presented by Schmidt and Güntert (2013) revealed a wide range of structural qualities. The most problematic cases are those that yield a structure with high precision but low accuracy, where problems generally remain hidden when performing just a single NOESY assignment and structure calculation run.

The eight proteins are the human NFU1 iron-sulfur cluster scaffold homologue, Northeast Structural Genomics Consortium (NESG) target HR2876B

(PDB accession code 2LTM, 107 amino acid [aa] residues, ordered residues 13–104); the human mitotic checkpoint serine/threonine-protein kinase BUB1 N-terminal domain, HR5460A (2LAH, 160 aa, 12–160); the RRM domain of RNA-binding protein FUS, HR6430A (2LA6, 99 aa, 12–99); the homeobox domain of the human homeobox protein Nkx-3.1, HR6470A (2L9R, 69 aa, 12–55); a de novo designed protein with IF3-like fold, OR135 (2LN3, 83 aa, 5–75) (Koga et al., 2012); a de novo designed protein with P-loop NRPase fold, OR36 (2LCI, 134 aa, 3–125); TSTM1273 from *Salmonella typhimurium* LT2, StT322 (2LOJ, 63 aa, 23–63); and the NifI-like protein from *Saccharomyces cerevisiae* YR313A (2LTL, 119 aa, 16–116). The corresponding NMR structures deposited in PDB were used as reference structures in this study. In principle, X-ray structures could also be used as independently determined reference structures (but were not available for these proteins).

The NMR data provided for this project were prepared according to standard NESG procedures (http://www.nesg.org). Two data sets were available for each protein, one containing "refined" NOESY peak lists that were used for the final structure calculations of the reference structures and one containing the "raw" NOESY peak lists from an early stage of spectral analysis. Peak lists were generated from ^{15}N-resolved NOESY spectra as well as ^{13}C-resolved NOESY spectra. Chemical shift assignments were performed manually by experienced scientists and have been provided in addition to the NOESY peak lists.

Second Test Data Set

The second test data set is composed of solution NMR data sets from ten different proteins. It includes for each protein the original experimental data set and modifications thereof that simulate a large variety of data imperfections.

The ten proteins are copper chaperone of *Enterococcus hirae* (PDB accession code 1CPZ, BMRB accession code 4344, 68 aa residues) (Wimmer et al., 1999); chicken prion protein fragment 128–242 (PDB 1U3M, BMRB 6269, 117 aa) (Calzolai et al., 2005); *Arabidopsis thaliana* ENTH-VHS domain At3g16270 (PDB 1VDY, BMRB 5928, 140 aa) (López-Méndez and Güntert, 2006; López-Méndez et al., 2004); Src homology 2 domain from the human feline sarcoma oncogene Fes (PDB 1WQU, BMRB 6331, 114 aa) (Scott et al., 2004, 2005); F-spondin TSR domain 4 (PDB 1VEX, BMRB 10002, 56 aa) (Pääkkönen et al., 2006); *Bombyx mori* pheromone binding protein (PDB 1GM0, BMRB 4849, 142 aa) (Horst et al., 2001); *Arabidopsis thaliana* rhodanese domain At4g01050 (PDB 1VEE, BMRB 5929, 134 aa) (Pantoja-Uceda et al., 2004, 2005); *Williopsis mrakii* killer toxin (PDB 1WKT, BMRB 5255, 88 aa) (Antuch et al., 1996); stereo-array isotope labeled (SAIL) calmodulin (PDB 1X02, BMRB 6541, 293 aa) (Kainosho et al., 2006); and second WW domain from mouse Salvador homolog 1 protein (mWW45) (PDB 2DWV, BMRB 10028, 98 aa) (Ohnishi et al., 2007).

Each data set contains chemical shift lists, peak lists from 2D and/or 3D ^{15}N-resolved and ^{13}C-resolved NOESY spectra, and TALOS-generated angle restraints. Modifications include, among others, various percentages of randomly deleted chemical shifts, randomly permuted chemical shifts, or randomly deleted NOESY peaks. A detailed description of the proteins as well as the 81 types of data set modifications will be given elsewhere. Every type of modification was performed five times using a different seed for random number generation, resulting in a total of 10 × 81 × 5 = 4,050 data sets that were used to evaluate our structure calculation protocol. Parameter values identical to the CASD data set were chosen for computing the consensus structure bundles.

Analysis of Structure Calculation Results

Evaluation of structure calculation results was mainly based on rmsd values that were calculated with respect to the reference structure (accuracy) and to the mean coordinates of the bundle (precision). Rmsd values were calculated for the 20 individual structure bundles, for the combined structure bundle consisting of the lowest target function structure from the 20 individual calculations, and for the consensus structure bundle based on the consensus set of distance restraints. Only backbone atoms N, C^{α}, C′ in the structured regions of each protein were considered for rmsd calculations.

Structure Validation

One of the 20 individual structure calculation results of each of the CASD data sets was validated using the "validate" script of the CYANA software package that calls various validation software tools and summarizes their respective results into one file. Structure validation parameters were computed for 1 of the 20 individual structure calculation results and the following parameters were chosen: (1) zp-comb-score from the software ProSa2003 (Sippl, 1993); (2) the Verify3D score (Bowie et al., 1991; Lüthy et al., 1992); (3) the clashscore calculated by MolProbity, and the MolProbity score, which considers steric clashes, and Ramachandran plot and staggered rotamer outliers (Chen et al., 2010; Davis et al., 2004, 2007); (4) the packing, the Ramachandran plot appearance, the χ^1/χ^2 rotamer normality, and the backbone conformation quality scores calculated by the WHAT_CHECK program (Hooft et al., 1996); (5) the percentage of residues in the most favored region of the Ramachandran plot (Rama G-factor), and the χ^1 rotamer normality (Chi-1 G-factor), as defined by the program PROCHECK-NMR (Laskowski et al., 1996; Morris et al., 1992).

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2014.11.014.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support by the Lichtenberg program of the Volkswagen Foundation and the Japan Society for the Promotion of Science.

Received: September 24, 2014
Revised: November 17, 2014
Accepted: November 17, 2014
Published: January 8, 2015

REFERENCES

Antuch, W., Güntert, P., and Wüthrich, K. (1996). Ancestral βγ-crystallin precursor structure in a yeast killer toxin. Nat. Struct. Biol. *3*, 662–665.

Bagaria, A., Jaravine, V., Huang, Y.P.J., Montelione, G.T., and Güntert, P. (2012). Protein structure validation by generalized linear model root-mean-square deviation prediction. Protein Sci. *21*, 229–238.

Bagaria, A., Jaravine, V., and Güntert, P. (2013). Estimating structure quality trends in the Protein Data Bank by equivalent resolution. Comput. Biol. Chem. *46*, 8–15.

Bowie, J.U., Lüthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known 3-dimensional structure. Science *253*, 164–170.

Calzolai, L., Lysek, D.A., Perez, D.R., Güntert, P., and Wüthrich, K. (2005). Prion protein NMR structures of chickens, turtles, and frogs. Proc. Natl. Acad. Sci. U S A *102*, 651–655.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D *66*, 12–21.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. *32*, W615–W619.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375–W383.

Güntert, P. (1998). Structure calculation of biological macromolecules from NMR data. Q. Rev. Biophys. *31*, 145–237.

Güntert, P. (2009). Automated structure determination from NMR spectra. Eur. Biophys. J. *38*, 129–143.

Güntert, P., Braun, W., and Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance

data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J. Mol. Biol. *217*, 517–530.

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. *319*, 209–227.

Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E.E. (1996). Errors in protein structures. Nature *381*, 272.

Horst, R., Damberger, F., Luginbühl, P., Güntert, P., Peng, G., Nikonova, L., Leal, W.S., and Wüthrich, K. (2001). NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. Proc. Natl. Acad. Sci. U S A *98*, 14374–14379.

Huang, Y.J., Tejero, R., Powers, R., and Montelione, G.T. (2006). A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins *62*, 587–603.

Jee, J., and Güntert, P. (2003). Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. J. Struct. Funct. Genom. *4*, 179–189.

Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A.M., and Güntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. Nature *440*, 52–57.

Kirchner, D.K., and Güntert, P. (2011). Objective identification of residue ranges for the superposition of protein structures. BMC Bioinformatics *12*, 170.

Koga, N., Tatsumi-Koga, R., Liu, G.H., Xiao, R., Acton, T.B., Montelione, G.T., and Baker, D. (2012). Principles for designing ideal protein structures. Nature *491*, 222–227.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.

López-Méndez, B., and Güntert, P. (2006). Automated protein structure determination from NMR spectra. J. Am. Chem. Soc. *128*, 13112–13122.

López-Méndez, B., Pantoja-Uceda, D., Tomizawa, T., Koshiba, S., Kigawa, T., Shirouzu, M., Terada, T., Inoue, M., Yabuki, T., Aoki, M., et al. (2004). NMR assignment of the hypothetical ENTH-VHS domain At3g16270 from Arabidopsis thaliana. J. Biomol. NMR *29*, 205–206.

Lüthy, R., Bowie, J.U., and Eisenberg, D. (1992). Assessment of protein models with 3-dimensional profiles. Nature *356*, 83–85.

Morris, A.L., Macarthur, M.W., Hutchinson, E.G., and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. Proteins *12*, 345–364.

Nabuurs, S.B., Spronk, C., Krieger, E., Maassen, H., Vriend, G., and Vuister, G.W. (2003). Quantitative evaluation of experimental NMR restraints. J. Am. Chem. Soc. *125*, 12026–12034.

Nabuurs, S.B., Spronk, C.A.E.M., Vuister, G.W., and Vriend, G. (2006). Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. PLoS Comp. Biol. *2*, 71–79.

Ohnishi, S., Güntert, P., Koshiba, S., Tomizawa, T., Akasaka, R., Tochio, N., Sato, M., Inoue, M., Harada, T., Watanabe, S., et al. (2007). Solution structure of an atypical WW domain in a novel beta-clam-like dimeric form. FEBS Lett. *581*, 462–468.

Pääkkönen, K., Tossavainen, H., Permi, P., Rakkolainen, H., Rauvala, H., Raulo, E., Kilpeläinen, I., and Güntert, P. (2006). Solution structures of the first and fourth TSR domains of F-spondin. Proteins *64*, 665–672.

Pantoja-Uceda, D., López-Méndez, B., Koshiba, S., Kigawa, T., Shirouzu, M., Terada, T., Inoue, M., Yabuki, T., Aoki, M., Seki, E., et al. (2004). NMR assignment of the hypothetical rhodanese domain At4g01050 from Arabidopsis thaliana. J. Biomol. NMR *29*, 207–208.

Pantoja-Uceda, D., López-Méndez, B., Koshiba, S., Inoue, M., Kigawa, T., Terada, T., Shirouzu, M., Tanaka, A., Seki, M., Shinozaki, K., et al. (2005). Solution structure of the rhodanese homology domain At4g01050(175-295) from Arabidopsis thaliana. Protein Sci. *14*, 224–230.

Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. Science *309*, 303–306.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics *23*, 381–382.

Rosato, A., Bagaria, A., Baker, D., Bardiaux, B., Cavalli, A., Doreleijers, J.F., Giachetti, A., Guerry, P., Güntert, P., Herrmann, T., et al. (2009). CASD-NMR: critical assessment of automated structure determination by NMR. Nat. Methods *6*, 625–626.

Rosato, A., Aramini, J.M., Arrowsmith, C., Bagaria, A., Baker, D., Cavalli, A., Doreleijers, J.F., Eletsky, A., Giachetti, A., Guerry, P., et al. (2012). Blind testing of routine, fully automated determination of protein structures from NMR data. Structure *20*, 227–236.

Saccenti, E., and Rosato, A. (2008). The war of tools: how can NMR spectroscopists detect errors in their structures? J. Biomol. NMR *40*, 251–261.

Schmidt, E., and Güntert, P. (2013). Reliability of exclusively NOESY-based automated resonance assignment and structure determination of proteins. J. Biomol. NMR *57*, 193–204.

Scott, A., Pantoja-Uceda, D., Koshiba, S., Inoue, M., Kigawa, T., Terada, T., Shirouzu, M., Tanaka, A., Sugano, S., Yokoyama, S., et al. (2004). NMR assignment of the SH2 domain from the human feline sarcoma oncogene FES. J. Biomol. NMR *30*, 463–464.

Scott, A., Pantoja-Uceda, D., Koshiba, S., Inoue, M., Kigawa, T., Terada, T., Shirouzu, M., Tanaka, A., Sugano, S., Yokoyama, S., et al. (2005). Solution structure of the Src homology 2 domain from the human feline sarcoma oncogene Fes. J. Biomol. NMR *31*, 357–361.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Sippl, M.J. (1993). Recognition of errors in 3-dimensional structures of proteins. Proteins *17*, 355–362.

Spronk, C.A.E.M., Nabuurs, S.B., Bonvin, A.M.J.J., Krieger, E., Vuister, G.W., and Vriend, G. (2003). The precision of NMR structure ensembles revisited. J. Biomol. NMR *25*, 225–234.

Spronk, C., Nabuurs, S.B., Krieger, E., Vriend, G., and Vuister, G.W. (2004). Validation of protein structures derived by NMR spectroscopy. Prog. Nucl. Magn. Reson. Spectrosc. *45*, 315–337.

Wimmer, R., Herrmann, T., Solioz, M., and Wüthrich, K. (1999). NMR structure and metal interactions of the CopZ copper chaperone. J. Biol. Chem. *274*, 22597–22603.

Zhao, D.Q., and Jardetzky, O. (1994). An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors. J. Mol. Biol. *239*, 601–607.

