CrossMark

ARTICLE

Systematic evaluation of combined automated NOE assignment and structure calculation with CYANA

Lena Buchner[1] · Peter Güntert[1,2,3]

Abstract The automated assignment of NOESY cross peaks has become a fundamental technique for NMR protein structure analysis. A widely used algorithm for this purpose is implemented in the program CYANA. It has been used for a large number of structure determinations of proteins in solution but a systematic evaluation of its performance has not yet been reported. In this paper we systematically analyze the reliability of combined automated NOESY assignment and structure calculation with CYANA under a variety of conditions on the basis of the experimental NMR data sets of ten proteins. To evaluate the robustness of the algorithm, the original high-quality experimental data sets were modified in different ways to simulate the effect of data imperfections, i.e. incomplete or erroneous chemical shift assignments, missing NOESY cross peaks, inaccurate peak positions, inaccurate peak intensities, lower dimensionality NOESY spectra, and higher tolerances for the matching of chemical shifts and peak positions. The results show that the algorithm is remarkably robust with regard to imperfections of the NOESY peak lists and the chemical shift tolerances but susceptible to lacking or erroneous resonance assignments, in particular for nuclei that are involved in many NOESY cross peaks.

Keywords Automated assignment · NOESY · Distance restraints · Structure calculation · CYANA

✉ Peter Güntert
guentert@em.uni-frankfurt.de

[1] Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, and Frankfurt Institute of Advanced Studies, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

[2] Laboratory of Physical Chemistry, ETH Zürich, Zurich, Switzerland

[3] Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

Introduction

The structure determination of biological macromolecules by NMR in solution relies primarily on distance restraints derived from cross peaks in NOESY spectra. A large number of assigned NOESY cross peaks are necessary to compute an accurate three-dimensional (3D) structure because many of the NOEs are short-range with respect to the sequence and thus carry little information about the tertiary structure and because NOEs are generally interpreted as loose upper bounds in order to implicitly take into account internal motions and spin diffusion [although, in principle, accurate distance measurements are possible with NOEs (Vögeli et al. 2012, 2009)]. Obtaining a comprehensive set of distance restraints from NOESY spectra is in practice not straightforward. The sheer amount of data, as well as resonance and peak overlap, spectral artifacts and noise, and the absence of expected signals because of fast relaxation turn interactive NOESY cross peak assignment into a laborious and error-prone task. Therefore, the development of computer algorithms for automating this often most time-consuming step of a protein structure determination by NMR has been pursued intensely and reviewed extensively (Altieri and Byrd 2004; Baran et al. 2004; Billeter et al. 2008; Gronwald and Kalbitzer 2004; Guerry and Herrmann 2011; Güntert 1998, 2003, 2009; Moseley and

Springer

Montelione 1999; Williamson and Craven 2009). Besides semi-automatic approaches (Duggan et al. 2001; Güntert et al. 1993; Meadows et al. 1994), several algorithms have been developed for the automated analysis of NOESY spectra given the chemical shift assignments, namely NOAH (Mumenthaler and Braun 1995; Mumenthaler et al. 1997), ARIA (Nilges et al. 1997; Rieping et al. 2007), ASDP (Huang et al. 2006), KNOWNOE (Gronwald et al. 2002), CANDID (Herrmann et al. 2002a), PASD (Kuszewski et al. 2004), AutoNOE-Rosetta (Zhang et al. 2014), and a Bayesian approach (Hung and Samudrala 2006). Automated NOESY peak picking has been integrated into the method (Herrmann et al. 2002b). Automated NOESY assignment can be combined with automated sequence-specific resonance assignment with the Garant (Bartels et al. 1997) or FLYA (Schmidt and Güntert 2012) algorithms in order to perform a complete NMR structure determination without manual interventions (López-Méndez and Güntert 2006). In favorable cases, this can even be achieved using exclusively experimental data from NOESY spectra (Ikeya et al. 2011; Schmidt and Güntert 2013).

The fundamental problem of NOESY assignment is the ambiguity of cross peak assignments. Assigning based solely on the match between cross peak positions and the chemical shift values of candidate resonances does in general not yield a sufficient number of unambiguously assigned distance restraints to obtain a structure (Mumenthaler et al. 1997). Ambiguous distance restraints make it possible to use also NOEs with multiple assignment possibilities in a structure calculation (Nilges 1995). Nevertheless, additional criteria have to be applied to resolve these ambiguities, such as using secondary structure information (Huang et al. 2006) or a preliminary structure that is refined iteratively in cycles of NOE assignment and structure calculation (Mumenthaler and Braun 1995). The CANDID automated NOESY assignment method introduced the concepts of network anchoring to reduce the initial ambiguity of NOE assignments and constraint combination to reduce the impact of erroneous restraints (Herrmann et al. 2002a). In CYANA, the conditions applied by CANDID for valid NOE assignments have been reformulated in a probabilistic framework that is conceptually more consistent and better capable to handle situations of high chemical shift-based ambiguity of the NOE assignments (Güntert 2004, 2009).

The aforementioned approaches can go wrong in two ways, especially with low-quality input data. Either the algorithm fails to ever assign enough NOE distance restraints to obtain a defined structure. This outcome, manifested by a divergent structure bundle with a high RMSD, is unfortunate but straightforward to detect. More problematic are failures of a second kind, where the

algorithm, possibly gradually over several cycles, discards part of the NOE cross peaks (by letting them unassigned) and selects a self-consistent but incomplete subset of the data to compute a well-defined but erroneous structure, i.e. a tight bundle of conformers with low RMSD to its mean coordinates that, however, differs significantly from the (unknown) correct structure of the protein. If this outcome goes unnoticed, it may result in the publication or PDB deposition of erroneous structures that cannot be detected easily by coordinate-based validation tools (Nabuurs et al. 2006).

Given the widespread use of automated NOESY assignment algorithms (Guerry and Herrmann 2011; Williamson and Craven 2009) it is important to give criteria for their safe application (Herrmann et al. 2002a) and to assess their reliability. It is known that the CANDID algorithm generally requires a high degree of completeness of the backbone and side chain chemical shift assignments (Jee and Güntert 2003). Recently, the CASD-NMR initiative (Rosato et al. 2009) has evaluated several NMR structure determination methods by blind testing. Using high-quality data sets of small proteins from a structural genomics project it was found that the NOESY-based methods included in the test yielded structures with an accuracy of 2 Å RMSD or better to the subsequently released reference structures (Rosato et al. 2012). However, the situation is less clear for more difficult cases, in which the resonance assignments may be incomplete, spectral crowding, overlap, and low signal-to-noise ratios prevent collecting a "complete" set of NOESY cross peaks, or the lack of isotope labeling may preclude the use of, intrinsically less ambiguous, 3D and 4D NOESY spectra. Further complications may arise with symmetric multimers or solid-state NMR data. In this paper, we address these questions by an extensive, systematic analysis of the combined automated NOESY assignment and structure calculation algorithm in CYANA under a variety of conditions mimicking data imperfections that may occur with challenging systems.

Materials and methods

Combined automated NOE assignment and structure calculation algorithm

The algorithm for automated NOE assignment in CYANA (Güntert 2004, 2009) is a re-implementation of principles of the former CANDID procedure (Herrmann et al. 2002a) on the basis of a probabilistic treatment of the NOE assignment process. The key features of the algorithm are network anchoring to reduce the initial ambiguity of NOESY peak assignments, ambiguous distance restraints to generate conformational restraints from NOESY cross

peaks with multiple possible assignments, and constraint combination to minimize the impact of erroneous distance restraints on the structure. Automated NOE assignment and the structure calculation are combined in an iterative process that comprises, typically, seven cycles of automated NOE assignment and structure calculation, followed by a final structure calculation using only unambiguously assigned distance restraints. Between subsequent cycles, information is transferred exclusively through the intermediary 3D structures. The molecular structure obtained in a given cycle is used to guide the NOE assignments in the following cycle. Otherwise, the same input data are used for all cycles, that is the amino acid sequence of the protein, one or several chemical shift lists from the sequence-specific resonance assignment, and one or several lists containing the positions and volumes of cross peaks in 2D, 3D, or 4D NOESY spectra. The input may further include previously assigned NOE upper distance bounds or other previously assigned conformational restraints for the structure calculation.

In each cycle, first all assignment possibilities of a peak are generated on the basis of the chemical shift values that match the peak position within given tolerance values, and the quality of the fit between the atomic chemical shifts and the peak position is expressed by a Gaussian probability, P_{shifts}. Second, the probability $P_{structure}$ for agreement with the preliminary structure from the preceding cycle (if available), represented by a bundle of conformers, is computed as the fraction of the conformers in which the corresponding distance is shorter than the upper distance bound plus the acceptable distance restraint violation cutoff. Assignment possibilities for which the product of these two probabilities is below the required probability threshold are discarded. Third, each remaining assignment possibility is evaluated for its network anchoring, i.e., its embedding in the network formed by the assignment possibilities of all the other peaks and the covalently constrained short-range distances. The network anchoring probability $P_{network}$ that the distance corresponding to an assignment possibility is shorter than the upper distance bound plus the acceptable violation is computed given the assignments of the other peaks but independent from knowledge of the three-dimensional structure. Contributions to the network anchoring probability for a given "current" assignment possibility result from other peaks with the same assignment (e.g. transposed peaks), from pairs of peaks that connect indirectly the two atoms of the current assignment possibility via a third atom, and from peaks that connect an atom in the vicinity of the first atom of the current assignment with an atom in the vicinity of the second atom of the current assignment. Short-range distances that are constrained by the covalent geometry can, for network anchoring, take the same role as an unambiguously assigned NOE. Individual contributions to the network anchoring of the current assignment possibility are expressed as probabilities, $P_1, P_2, \ldots$, that the distance corresponding to the current assignment possibility satisfies the upper distance bound. The network anchoring probability is obtained from the individual probabilities as $P_{network} = 1 - (1 - P_1)(1 - P_2)\ldots$, which is never smaller than the highest probability of an individual network anchoring contribution. Only assignment possibilities for which the product of the three probabilities is above a threshold, $P_{tot} = P_{shifts}P_{network}P_{structure} \geq P_{min}$, are accepted. Cross peaks with a single accepted assignment yield a conventional unambiguous distance restraint. Cross peaks with multiple accepted assignments result in an ambiguous distance restraint.

Spurious distance restraints may arise from the misinterpretation of noise and spectral artifacts, in particular at the outset of a structure determination before 3D structure-based filtering of the restraint assignments can be applied. CYANA uses "constraint combination" (Herrmann et al. 2002a) to reduce structural distortions from erroneous distance restraints. Medium-range and long-range distance restraints are incorporated into "combined distance restraints", which are ambiguous distance restraints with assignments taken from different, in general unrelated, cross peaks. A basic property of ambiguous distance restraints is that the restraint will be fulfilled by the correct structure whenever at least one of its assignments is correct, regardless of the presence of additional, erroneous assignments. This implies that such combined restraints have a lower probability of being erroneous than the corresponding original restraints, provided that the fraction of erroneous original restraints is smaller than 50 %. Constraint combination aims at minimizing the impact of erroneous NOE assignments on the resulting structure at the expense of a temporary loss of information. It is applied to medium- and long-range distance restraints in, by default, the first two cycles of combined automated NOE assignment and structure calculation with CYANA.

The distance restraints are then included in the input for the structure calculation with simulated annealing by the fast CYANA torsion angle dynamics algorithm (Güntert et al. 1997). A complete structure calculation with automated NOESY assignment typically comprises seven cycles. The second and subsequent cycles differ from the first cycle by the use of additional selection criteria for cross peaks and NOE assignments that are based on assessments relative to the protein 3D structure from the preceding cycle. The precision of the structure determination normally improves with each subsequent cycle. Accordingly, the cutoff for acceptable distance restraint violations in the calculation of $P_{structure}$ is tightened from cycle to cycle. In the final cycle, an additional filtering step ensures that all

NOEs have either unique assignments to a single pair of hydrogen atoms, or are eliminated from the input for the structure calculation. This facilitates the use of subsequent refinement and analysis programs that cannot handle ambiguous distance restraints.

Experimental NMR data sets

The performance of CYANA was assessed on the basis of the NMR structure bundles of ten proteins to which we refer in this paper by the four-letter acronyms given in Table 1: copz, the copper chaperone CopZ of *Enterococcus hirae* (Wimmer et al. 1999); cprp, the chicken prion protein fragment 128–242 (Calzolai et al. 2005); enth, the ENTH-VHS domain At3g16270 from *Arabidopsis thaliana* (López-Méndez and Güntert 2006; López-Méndez et al. 2004); fsh2, the Src homology two domain from the human feline sarcoma oncogene Fes (Scott et al. 2004, 2005); fspo, the F-spondin TSR domain 4 (Pääkkönen et al. 2006); pbpa, the *Bombyx mori* pheromone binding protein (Horst et al. 2001); rhod, the rhodanese homology domain At4g01050 from *Arabidopsis thaliana* (Pantoja-Uceda et al. 2005, 2004); wmkt, the *Williopsis mrakii* killer toxin (Antuch et al. 1996); scam, stereo-array isotope labeled (SAIL) calmodulin (Kainosho et al. 2006); ww2d, the second WW domain from mouse salvador homolog 1 protein (Ohnishi et al. 2007).

The proteins copz, cprp, enth, fsh2, pbpa, rhod and wmkt are proteins will a well-defined single-domain structure. The protein fspo has an unusual, less well-defined structure without regular secondary structure. The protein scam has two flexibly connected domains. The protein ww2d forms a symmetric dimer. For the original

structure determinations the proteins were uniformly labeled with ^{13}C and ^{15}N, except for copz that was only ^{15}N labeled, wmkt that was unlabeled, and scam that was stereo-array isotope labeled (Kainosho et al. 2006). The completeness of the resonance assignments and the type and amount of NOESY data are summarized in Table 1.

For most proteins the unassigned NOESY peak lists were the only source of conformational restraints. Exceptions are cprp and pbpa, whose data sets included 123 and 148 ϕ/ψ torsion angle restraints derived from C^α chemical shifts (Luginbühl et al. 1995), respectively, ww2d including 44 ϕ/ψ torsion angle restraints from TALOS (Cornilescu et al. 1999). In the data set of cprp the assignments of 18 NOESY cross peaks were kept fixed, as in the original structure determination (Calzolai et al. 2005). Disulfide bonds were restrained in cprp, fspo, pbpa, and wmkt. In scam the distances between the four calcium ions and their 16 ligands were restrained to the range 1.7–2.8 Å. No hydrogen bond restraints or other additional restraints were used.

The original experimental data sets were used to determine a reference structure for each protein using the same computational schedule as for the subsequent calculations with modified data. Seven cycles of combined automated NOESY assignment and structure calculation were performed, followed by a final structure calculation. In each cycle, structure calculations were started from 100 conformers with random values of the torsion angles, to which the standard CYANA simulated annealing schedule was applied with 10,000 torsion angle dynamics steps per conformer. The 20 conformers with the lowest final target function values were selected for analysis and are shown in Fig. 1.

Table 1 Overview of proteins and data sets used in this study

Acronym	PDB code	BMRB code	Amino acids	Chemical shifts	Assignment completeness (%)	NOESY spectra	NOESY peaks
copz	1CPZ	4344	68	425	88.8	2D, ^{15}N	1175, 1063
cprp	1U3M	6269	117	1093	97.8	2D, ^{15}N, ^{13}C, aro	94, 1498, 2850, 166
enth	1VDY	5928	140	1471	96.0	^{15}N, ^{13}C	1730, 4169
fsh2	1WQU	6331	114	1101	97.2	^{15}N, ^{13}C, aro	1313, 2979, 440
fspo	1VEX	10002	56	583	98.6	^{15}N, ^{13}C	494, 1398
pbpa	1GM0	4849	142	1409	99.3	^{15}N, ^{13}C, aro	1387, 3907, 320
rhod	1VEE	5929	134	1268	98.4	^{15}N, ^{13}C, aro	1801, 3683, 354
wmkt	1WKT	5255	88	455	97.0	2D	1998
scam	1X02	6541	148	1214	100.0	^{15}N, ^{13}C, 2D	1703, 3026, 85
ww2d	2DWV	10028	98	900	90.9	^{15}N, ^{13}C, inter	380, 1203, 62

The assignment completeness gives the percentage of the aliphatic and aromatic ^1H and backbone $^1H^N$ resonances that are assigned. Codes for NOESY spectra types are: 2D, 2D [^1H,^1H]-NOESY; ^{15}N, 3D ^{15}N-resolved [^1H,^1H]-NOESY; ^{13}C, 3D ^{13}C-resolved [^1H,^1H]-NOESY; aro, 3D aromatic ^{13}C-resolved [^1H,^1H]-NOESY; inter, 3D ^{13}C-filtered ^{13}C-edited [^1H,^1H]-NOESY for the detection of intermolecular NOEs (Zwahlen et al. 1997). The numbers of NOESY peaks are given in the order of the spectra in the preceding column

Fig. 1 Bundle representations of the ten proteins included in the present study (see Table 1). Secondary structure elements are highlighted in purple (α-helix) and yellow (β-sheet). Atomic coordinates originate from the Protein Data Bank (PDB) entries 1CPZ (copz), 1U3M (cprp), 1VDY (enth), 1WQU (fsh2), 1VEX (fspo), 1GM0 (pbpa), 1VEE (rhod), 1X02 (scam), 1WKT (wmkt), and 2DWV (ww2d). Two separate superpositions are presented for the two-domain protein scam
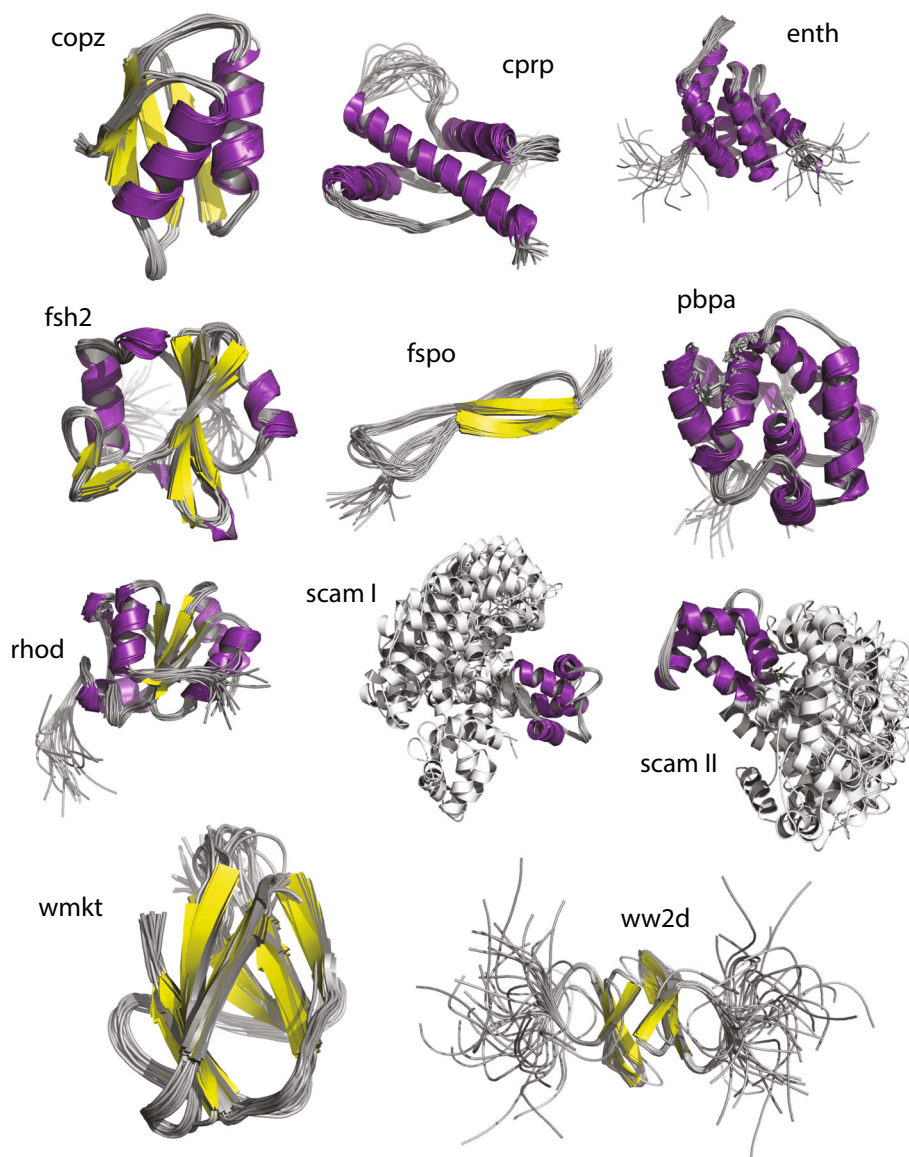
Modified input data sets

The experimental input data sets were modified in 14 different ways to mimic different kinds of data imperfections. All random data modifications were applied five times using different random numbers resulting in a total of 397 different data sets for each protein including the respective complete data set.

1. Missing chemical shift assignments
    A given percentage P between 0 and 40 % of randomly selected ^{1}H chemical shift assignments was deleted. Experimental NOESY peak lists were not changed.

    (a) Random shift deletion: The shifts to be deleted were chosen randomly among all assigned ^{1}H chemical shifts.

    (b) Deletion of side chain chemical shifts: The shifts to be deleted were chosen randomly among all side-chain ^{1}H chemical shift assignments.

    (c) Deletion of "important" chemical shift assignments: The shifts to be deleted were chosen among all assigned ^{1}H chemical shifts, but "important" shifts were deleted with higher probability. Importance was defined according to the number of NOEs in the reference calculation that involve a given atom. Chemical shifts were divided into eleven classes occurring in 0–1, 2–3, 4–5,…, and $\geq$20 peaks, with class indices $i = 0, 2, 4,…, 20$. Chemical shifts from class i were deleted with relative deletion probability $p_i = 1/(21-i)$, resulting in higher

deletion probabilities for more important chemical shifts.

(d)   Deletion of "unimportant" chemical shift assignments: As in (c), but "unimportant" ^1H shifts were deleted preferably. Chemical shifts from class i were deleted with relative deletion probability $p_i = 1/(i + 1)$.

2.   Erroneous chemical shift assignments
     A given percentage P between 0 and 40 % of randomly selected assigned ^1H chemical shift values were modified. Experimental NOESY peak lists were not changed.

(e)   Random new chemical shift values: The selected chemical shifts were set to randomly chosen values within fifteen times the assignment tolerance for a given atom.

(f)   Chemical shift permutations: Each selected chemical shift values was replaced with the chemical shift value of another atom from the set of selected atoms. Only atoms with a chemical shift value within 2.5 times the standard deviation of the corresponding chemical shift distribution from the BMRB were used for replacement.

(g)   Permuted locally with other chemical shifts: As in (f), but only atoms from the same or directly neighboring amino acid residues were used for replacement.

3.   Missing NOESY peaks
     A given percentage P between 0 and 75 % of the NOESY peaks was deleted. Chemical shift lists were not changed.

(h)   Random peak deletion: The peaks to be deleted were chosen randomly.

(i)   Deletion of weak peaks: The weakest peaks were (non-randomly) deleted.

4.   Inaccurate NOESY peaks
     The positions or volumes of all NOESY peaks were distorted. Chemical shift lists were not changed.

(j)   Inaccurate peak positions: Peak positions were modified in all spectral dimensions by adding a random number from a normal distribution with mean 0 and standard deviation equal to the corresponding assignment tolerance times a varying percentage P between 0 and 100 %.

(k)   Inaccurate peak volumes: Peak volumes were multiplied by a normally distributed random number with mean 1 and standard deviation P between 0 and 150 %.

5.   Projection to two dimensions

(l)   NOESY peak lists of all data sets were reduced to the two proton dimensions.

6.   Increased chemical shift tolerances

(m)   Chemical shift tolerance for NOESY peak assignment was increased from the standard value of 0.03 ppm to 0.04, 0.05, 0.06, 0.08, and 0.1 ppm for ^1H, and proportionally from 0.5 ppm to 0.67, 0.83, 1.0, 1.33, and 1.67 ppm for ^{15}N and ^{13}C. Chemical shift lists and NOESY peak lists were not changed.

7.   Increased number of random starting structures and annealing steps

(n)   The calculations with randomly deleted chemical shifts of modification (a) were repeated with 200 instead of 100 random starting structures and 20,000 instead of 10,000 torsion angle dynamics steps during the simulated annealing protocol.

Structure calculations

Automated NOESY peak assignment was performed with a chemical shift tolerance of 0.03 ppm for ^1H and 0.5 ppm for heavy atoms [except for modifications (m), see above]. Twenty independent structure calculation runs starting from different random structures were performed for each data set of each protein. Each of these structure calculations [except for modification (n), see above] started from 100 random conformers to which the standard CYANA simulated annealing protocol with 10,000 torsion angle dynamics steps was applied, and the 20 conformers with lowest target function values were chosen for the final structure bundle.

Analysis of results

For each protein, the solution NMR structure calculated from the complete data set was used as the reference structure (Fig. 1). The accuracy of a structure was measured by the RMSD bias (Güntert 1998), i.e. the backbone RMSD between the average structure of a given calculation and the average structure of the reference. The average structure of a structure bundle was obtained by optimally superimposing its individual conformers for minimal backbone RMSD of the ordered regions, and calculating the average coordinates. Ordered parts of each protein were determined by the program CYRANGE (Kirchner et al. 2011) applied to the reference structure. The average

RMSD bias for each type of input data modification was averaged over all ten proteins, five different random modifications and 20 independent structure calculation runs leading to averaging over 1000 structure calculations.

Important as well as unimportant chemical shifts were further analyzed by classification into six different ^{1}H classes: H$^\alpha$, H^N, methyl protons, aromatic ring protons, lysine and arginine side chain protons beyond H$^\beta$, and aliphatic protons. The number of NOE cross peaks involving a given atom was determined for each atom and the average was calculated for the different classes.

In de novo structure calculations there is usually no reference structure available. It is therefore necessary to have a measure independent from the RMSD bias to assess the quality of a structure calculation result. We analyzed two previously suggested criteria, i.e. the RMSD to the mean structure (RMSD radius) of cycle 1 (convergence) and the RMSD between the structure obtained in cycle 1 and in the final structure calculation (RMSD drift). The individual criteria were then combined into a weighted average calculated as $\sqrt{(1.5R)^2 + D^2}$, where R denotes the RMSD radius in cycle 1 and D the RMSD drift.

Results and discussion

The effect of missing, erroneous, or inaccurate structure calculation input data was investigated by random deletion and modification of chemical shifts as well as NOESY peaks. Structure calculations were performed using original and modified experimental data sets of ten different proteins (Table 1; Fig. 1) and the average RMSD bias was used as a measure of accuracy.

The consequence of random new chemical shifts in comparison to missing NOESY peaks is illustrated in Fig. 2 for the protein fsh2 as an example of the two principle kinds of structure calculation failures that were discussed in the Introduction. An incomplete set of NOESY peaks generally causes less well defined structure bundles indicative of a loss of long-range information. This is reflected in the RMSD radius, which increases from 1.15 Å at 30 % deleted peaks (Fig. 2a) to 2.08 Å at 60 % deleted peaks (Fig. 2b) and 10.13 Å at 75 % deleted peaks (Fig. 2c). This example illustrates the first category of structure calculation failure, namely the inability to ever assign enough distance restraints to converge to a well-defined structure bundle. This type of error is straightforward to detect and therefore less problematic. The results for erroneous chemical shifts show a different effect. The bundle remains rather well defined with a low RMSD radius of 0.82 Å (10 % modified chemical shifts, Fig. 2c), 1.04 Å (30 %, Fig. 2d) and 1.8 Å (40 %, Fig. 2e) whereas

the increasing RMSD bias of 2.07 Å (10 %), 7.64 Å (30 %) and 7.1 Å (40 %) shows that the structure calculation converges to an incorrect fold at a certain degree of erroneous shifts. This reflects the second kind of failure that can be attributed to the selection of a self-consistent, but incorrect subset of NOESY peak assignments. Due to the well-defined nature of the structure bundle, the error is more difficult to detect and hence potentially more dangerous.

For a systematic evaluation, the average RMSD bias was plotted against the percentage P of modified input data for the different types of modifications (Figs. 3, 4, 5). The dotted line indicates an RMSD value of 3 Å representing the threshold below which the global fold of the structure is still assumed to be correct. The results for each individual protein can be found in Fig. 4 and in the Supplementary Material (Figs. S1–S10).

The overall effect of chemical shift deletions is presented in Fig. 3a–d. Chemical shifts were deleted in four different ways: random deletion from the set of all shifts (Fig. 3a), random deletion only from side chain atoms (Fig. 3b), random deletion of "important" shifts (Fig. 3c) and random deletion of "unimportant" shifts (Fig. 3d). Omission rates were varied between 0 and 40 % in steps of 5 %. In all four cases the average RMSD bias increases at increasing omission rates P. In most cases, random deletion of 5 % of the chemical shifts results in structures with an RMSD bias below 3 Å, whereas 10 and 15 % missing chemical shifts raise the average RMSD bias slightly above 3 Å (Fig. 3a). Omission rates of more than 15 % increase the average RMSD including the standard deviation considerably above 3 Å indicating that structure calculations reproducibly fail to converge to the correct global fold when using severely incomplete chemical shift data. The outcome in the range between 10 and 15 % chemical shift omission strongly depends on the protein and the quality of the respective NOESY data, which becomes apparent when comparing the plots for the individual proteins presented in Fig. 4 and in the Supplementary Material. In favorable cases, the correct structure can still be found with 20 % chemical shifts missing, whereas rather unfavorable cases may fail at 5 % missing chemical shifts. Torsion angle restraints generated from chemical shifts with the program TALOS (Cornilescu et al. 1999) can in some cases slightly improve the structure calculation result. These improvements are predominantly observed in cases of higher deletion percentages (above 20 %) where the original calculations excluding TALOS restraints did not necessarily converge to the correct global fold (Fig. S11). Nearly no improvement is observed for the proteins copz and fspo.

It does not make any significant difference whether random chemical shifts or only side-chain shifts are missing (Fig. 3a, b). Deletion of "important" shifts causes a
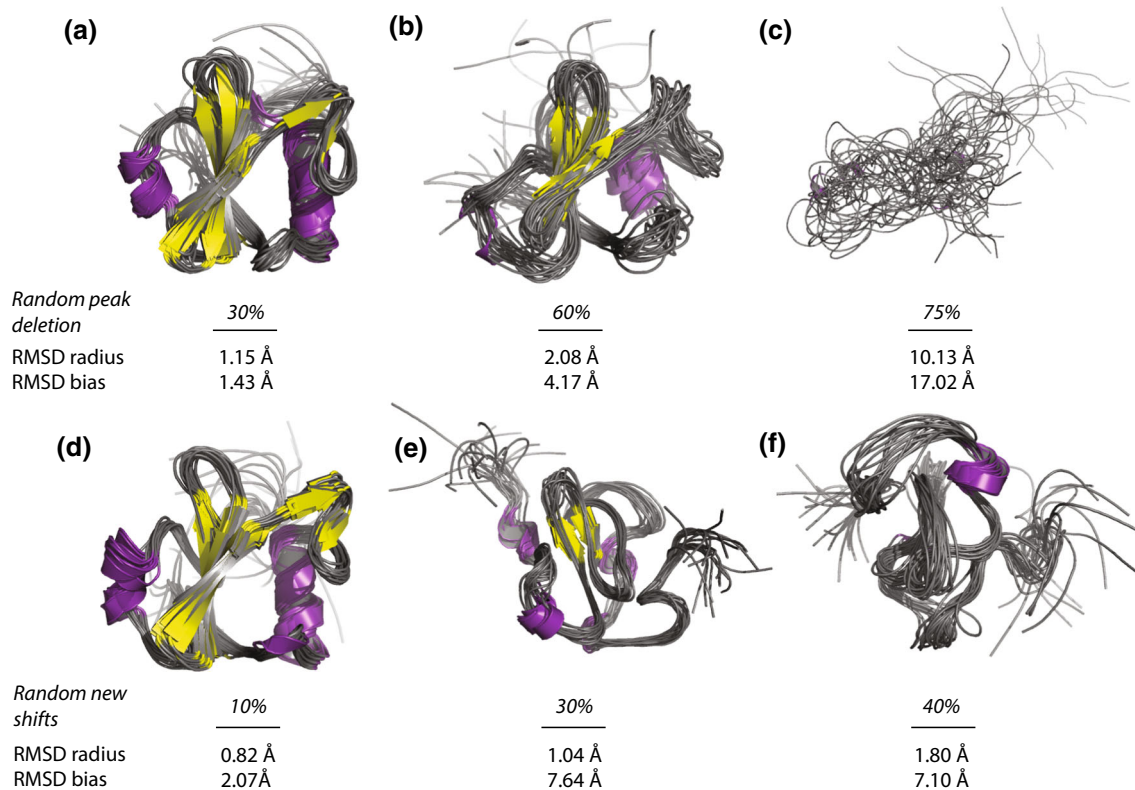
Fig. 2 Effect of **a** 30 %, **b** 60 %, **c** 75 % missing NOESY peaks [modification (h) in Methods] and **d** 10 %, **e** 30 %, **f** 40 % erroneous chemical shift assignments [modification (e) in Methods] on the structure calculation result of the protein fsh2. Structures were calculated using the standard CYANA protocol for combined automated NOE assignment and structure calculation based on 100 random starting structures and 10,000 annealing steps. The final structure bundles comprise the 20 conformers with lowest target function values. The ordered residues 8–108 in the reference structure (PDB 1WQU) were used for superposition and RMSD calculation. The RMSD bias is calculated as the RMSD between the mean structure of the bundle and the mean reference structure and represents the accuracy. The RMSD radius is calculated as the average RMSD of each conformer to the mean structure of the bundle and represents the precision

steeper increase in the average RMSD bias compared to random deletion, whereas the slope is less steep in the case of "unimportant" shifts (Fig. 3c, d). This shows that it can make a difference for structure calculation results which particular chemical shifts are missing. It is in practice more likely that "unimportant" shifts are missing, as they are typically more difficult to assign.

To further investigate the importance of individual types of protons, chemical shifts from all data sets were classified into six different classes: H^{α}, NH, methyl protons, aromatic protons, lysine and arginine side chain protons, and aliphatic protons. Importance is measured based on the amount of medium- and long-range NOESY peaks that involve the respective chemical shift (Fig. 6). Protons from methyl groups appear on average in 17.5 medium- and long-range NOE peaks; aromatic protons appear on average in 13.5 peaks, NH protons in 11.9 peaks, H^{α} protons in 10.3 peaks, aliphatic protons in 10.2 peaks and Lys/Arg sidechain protons in 9.0 peaks. Figure 6 suggests that methyls and aromatic protons are very important, which

can be attributed to their preferential occurrence in the hydrophobic, densely packed core of the protein enabling a large amount of NOE contacts.

Figure 3e–g shows the effect of modified chemical shift values. Different simulated sources of errors such as random new chemical shift values (Fig. 3e), randomly permuted chemical shift values (Fig. 3f), and locally permuted chemical shift values (Fig. 3g) result in very similar average RMSD values as random missing chemical shifts. Even local permutations show the same result.

Compared to missing chemical shifts, deletion of NOESY peaks shows a less steep increase of average RMSD (Fig. 5a). On average, the RMSD bias at 30 % deleted NOESY peaks is below 3 Å while the average RMSD rises slightly above 3 Å at 45 %. The much less pronounced increase can be explained by the fact that NOESY peaks firstly contain a large amount of signals that contain no or very limited structural information due to their sequential nature and secondly contain rather redundant information through the dense NOE network. In
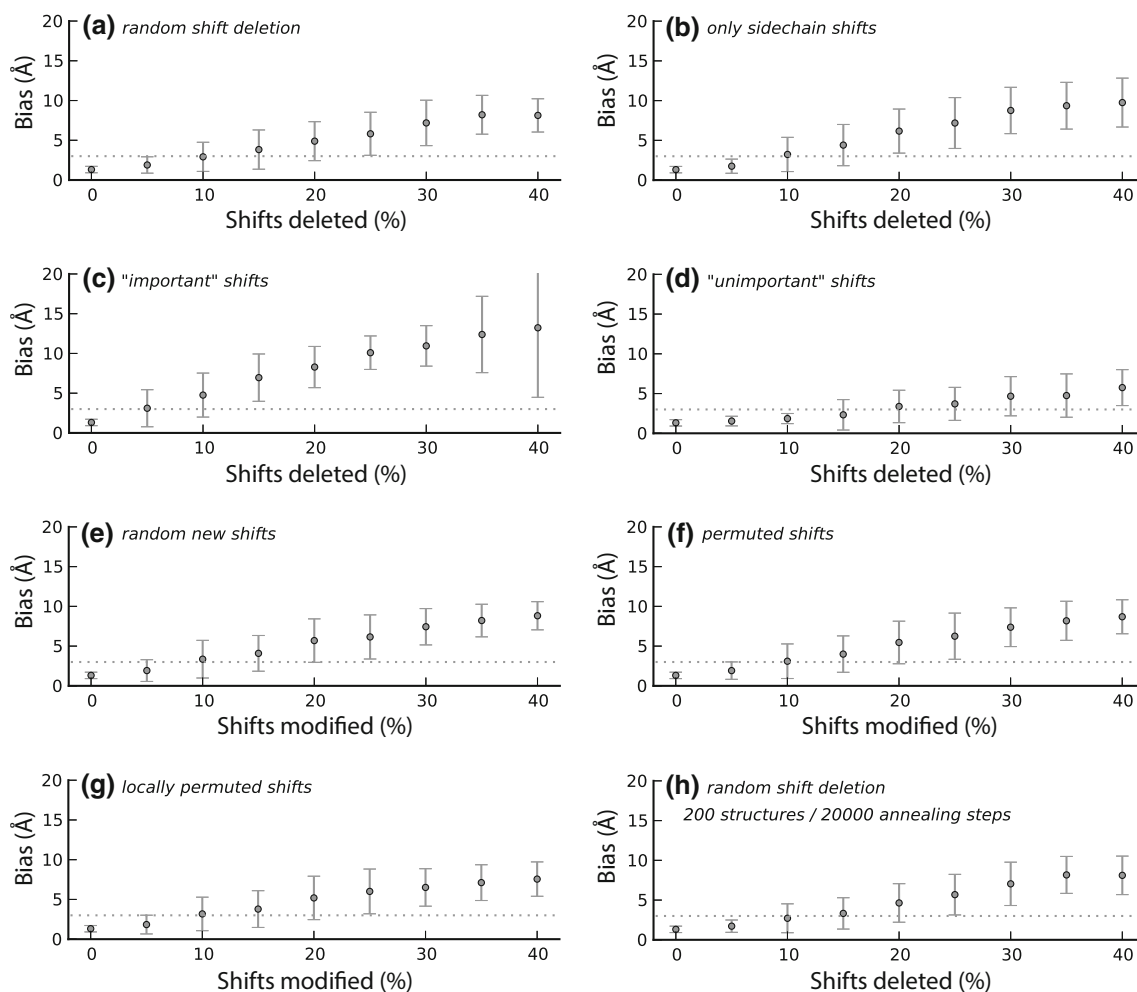
Fig. 3 RMSD to the reference structure for different types of simulated chemical shift imperfections. For each data point, twenty independent automated NOESY assignment and structure calculation runs were performed for each of five randomly modified data sets of the ten proteins of Table 1. The average RMSD to the reference structure is plotted against the percentage P of modified chemical shifts. See Methods for details. The data point at 0 % shift modification denotes the RMSD for 20 runs with the complete, unmodified experimental data. **a** Random deletion of chemical shift assignments, **b** random deletion of side-chain chemical shift assignments, **c** random deletion of "important" chemical shift assignments, **d** random deletion of "unimportant" chemical shift assignments, **e** random new chemical shift values, **f** random permutation of chemical shift values, **g** local permutation of chemical shift values, **h** doubled number of random starting structures and annealing steps for randomly deleted chemical shift assignments

contrast, one missing chemical shift leads to a whole set of NOESY peaks that remain unassigned in the more favorable case or get assigned incorrectly in the less favorable case. Figure 5b shows the result for deletion of weak peaks. The RMSD bias at 30 % deletion is comparable to random deletion, whereas deletion of 45 % of the weakest peaks results in a significant increase of 7 Å compared to 3 Å at 45 % randomly deleted peaks. A higher average RMSD for deletion of weak peaks is expected as they contain important long-range information.

Using the complete peak lists, but introducing errors in peak positions yields an average RMSD bias of 3 Å at 45 % error and of more than 5 Å at 60 % error (Fig. 5c). In contrast to errors in peak positions, errors in peak volumes

have largely no effect on the average RMSD for the complete range tested up to 150 % error (Fig. 5d). A larger influence from erroneous peak positions can be explained by the fact that the number of incorrect assignments increases, creating potentially distorting restraints, whereas erroneous peak volumes only affect the upper distance limit value. This erroneous effect on the upper distance limit value is furthermore greatly reduced by the r^{-6}-correlation between peak volume and calibrated distance.

Using only two-dimensional peak lists has almost no effect on the structure calculation result in the case of three proteins (copz, ww2d and wmkt). This result can be explained by the fact that a significant part of the peaks of the original data set comes from 2D NOESY spectra. Reducing
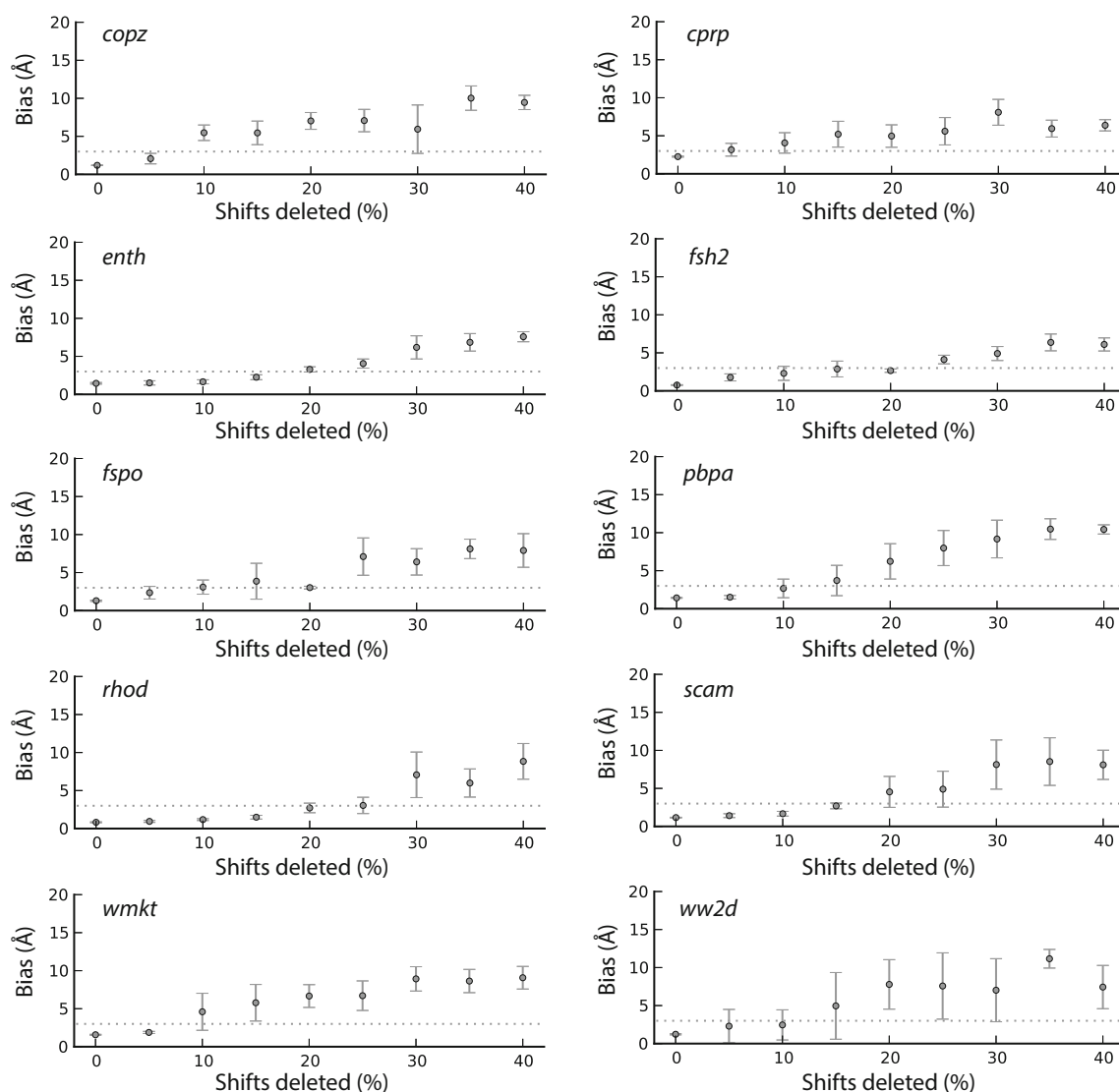
Fig. 4 RMSD to the reference structure for different percentages of randomly deleted chemical shifts [modification (a) in Methods]. Results are presented separately for each of the ten proteins of Table 1 and Fig. 1. For each data point, twenty independent automated NOESY assignment and structure calculation runs were performed for each of five randomly modified data sets. The RMSD bias from the reference structure is plotted against the percentage P of modified chemical shifts. The data point at 0 % shift deletion denotes the RMSD for 20 runs with the complete, unmodified experimental data

the remaining peaks to two dimensions has a less severe effect in these cases compared to other data sets, which contain mainly 3D data. For fsh2, fspo, rhod and scam the RMSD bias shows a slight increase but remains below 3 Å, and for cprp, enth and pbpa the RMSD bias increases above 5 Å (Fig. 5e).

Figure 5f shows the effect of increased chemical shift tolerances, which simulates spectra with less resolution resulting in higher assignment ambiguities. Chemical shift tolerances for NOESY peak assignments were raised up to 3.33 times their original value, which corresponds to 0.1 ppm for ^1H and 1.66 ppm for ^{15}N and ^{13}C. Up to 200 % increased tolerance, the average RMSD bias is still around 3 Å, whereas further increase results in RMSD bias

values of around 5 Å. Increased chemical shift tolerances have very diverse consequences on the different data sets (Supplementary Fig. S1-S10). The effect is most severe in cases where the data sets contain a large amount of two-dimensional data (copz, ww2d and wmkt) as well as in the case of the data set of cprp. Two-dimensional data are especially sensible to reduced resolution as the amount of assignment possibilities is much higher. It should, however, be noted that these two simulations (reduction to two spectral dimensions and increased chemical shift tolerance) might give a somewhat too optimistic picture of the situation encountered in NMR spectra with poor resolution. In severely overlapped spectra, several peaks may be fused into one single peak with a biased peak position, or peaks
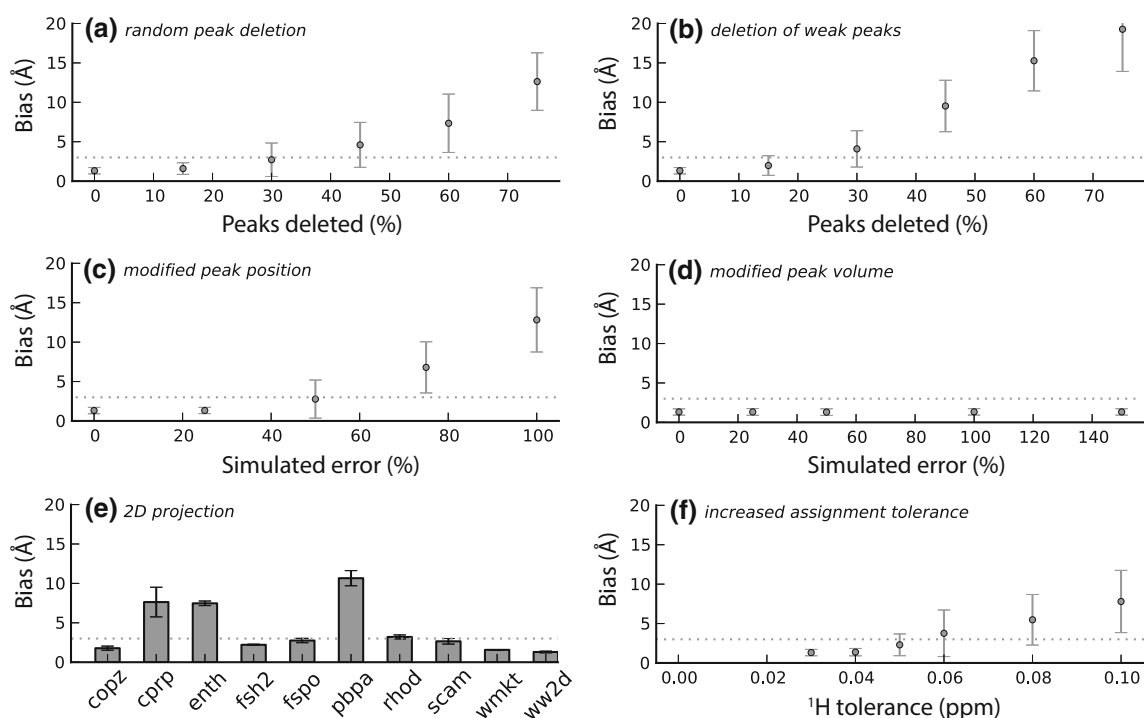
Fig. 5 RMSD to the reference structure for different types of simulated peak list imperfections. For each data point, twenty independent automated NOESY assignment and structure calculation runs were performed for each of five randomly modified data sets of the ten proteins of Table 1. The average RMSD to the reference structure is plotted against the percentage P of modified data, where applicable. See Methods for details. The data point at 0 % peak modification denotes the RMSD for 20 runs with the complete, unmodified experimental data. **a** Random deletion of NOESY peaks, **b** random deletion of the weakest NOESY peaks, **c** erroneous peak positions, **d** erroneous peak volumes, **e** 2D projection of NOESY peaks, **f** increased assignment tolerances

may no longer be recognizable at all. In our simulation, all peaks are still considered individually at the correct peak position.

Finally, we tested whether the effect of missing data can be compensated by performing more annealing steps during structure calculation and using more random starting structures. For this purpose, we repeated all calculations with randomly deleted chemical shifts with 200 instead of 100 random starting structures and with 20,000 instead of 10,000 annealing steps. The calculation results show only marginal overall improvement (Fig. 3h), indicating that data imperfections can in general not be compensated by longer computation times. The only exception is the homodimeric protein ww2d, for which longer simulated annealing yielded significantly lower RMSD bias values for the data sets with 5–15 % deleted chemical shifts.

These results show that data imperfections of various natures can dramatically reduce the quality of NMR structures. In case of de novo structure determination with lack of a reference structure, it is important to be able to evaluate the structure calculation result based on a measure independent of the RMSD bias. Several criteria have been suggested previously. Two of these criteria are the convergence (RMSD to the mean structure) of the initial structure calculation cycle and the RMSD drift (RMSD between the first and the last cycle). If the initial cycle converges to an RMSD radius below 3 Å and the RMSD drift is simultaneously below 2 Å, the result is considered reliable (Herrmann et al. 2002a; Jee and Güntert 2003). We have investigated these criteria using all aforementioned structure calculations and summarized the results in Fig. 7.

Figure 7a and b show the accuracy plotted against the RMSD in cycle 1 and the RMSD drift. Especially dangerous are false positives, i.e. cases, where the evaluation parameters meet the required criteria (convergence < 3.0 Å, drift < 2.0 Å) but the structure is misfolded. Considering both criteria individually, the number of false positives is 2 % (convergence) and 0.4 % (drift), respectively. Calculation of a weighted average from both values (Fig. 7c) further reduces the number of false positives to 0.01 %. The correlation of the weighted average and the accuracy shows a significantly reduced number of data points above the diagonal (accuracy exceeding the criterion) which therefore allows it to be used as an upper limit on the accuracy. The distribution for the homodimeric protein ww2d is presented separately in Fig. 7d. In contrast

Fig. 6 Number of ^1H resonances that are involved in a given number of medium- and long-range NOESY peaks. Proton chemical shifts were separated into six disjoint classes: HA, NH, aliphatic, methyl, aromatic, and lysine and arginine sidechain atoms. Chemical shifts were taken from the data sets of ten different proteins (Table 1). Peaks were counted in the final assigned peak lists of the combined automatic NOE assignment and structure calculation run that yielded the reference structure
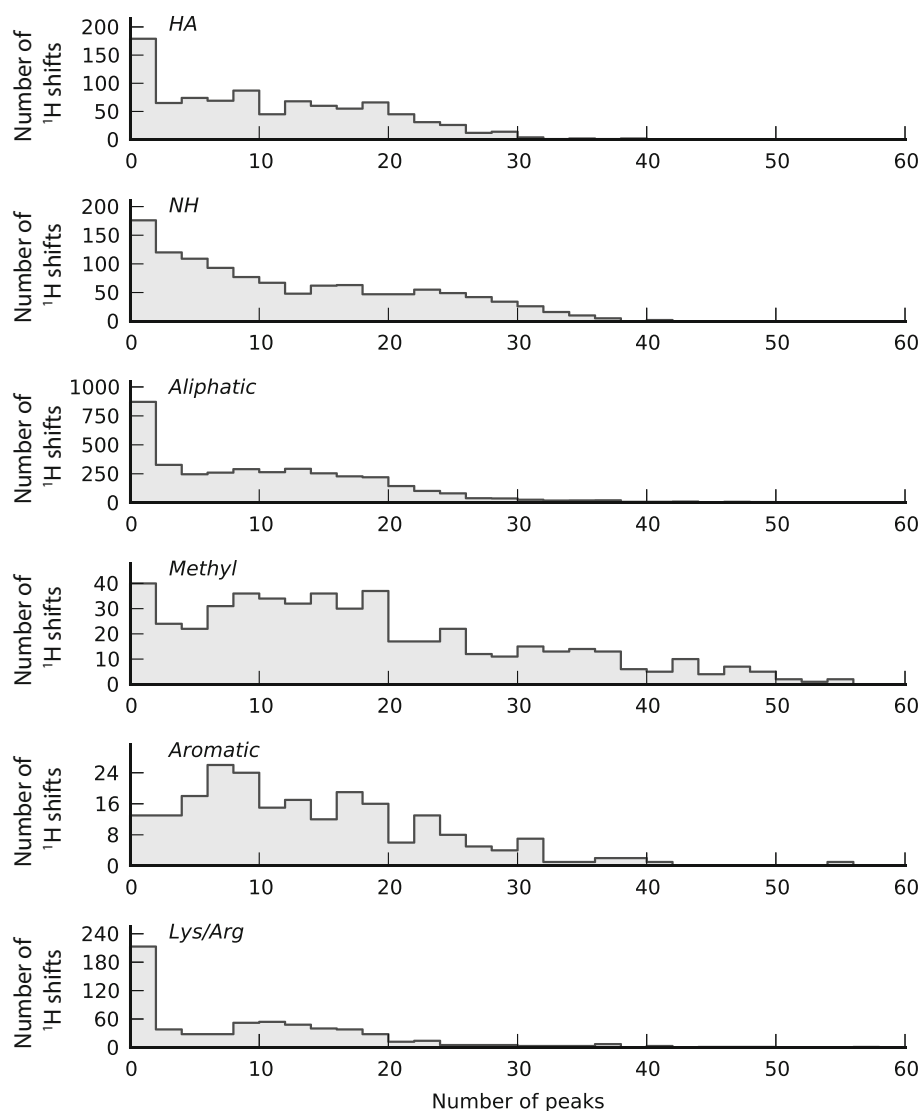
to the monomeric proteins, it shows multiple clusters that are presumably due to different ways of dimer formation. On the one hand, there are a large number of cases of structures with a high accuracy around 1 Å for which the combined criterion varies over a large range of 1–10 Å. On the other hand, there is a narrow cluster of structures with an RMSD bias of about 10 Å and values of 2–10 Å for the combined criterion.

In order to investigate the influence of artifacts such as water signals or baseline distortions on the structure calculation result, we have recalculated the structures of the three proteins enth, fsh2, and rhod based on peaks lists from automatic peak picking without subsequent refinement (López-Méndez and Güntert 2006). Results are summarized in Table 2. Only slight differences between the results obtained with refined and unrefined sets of peak lists can be observed in the case of enth with respect to the

RMSD bias, the final CYANA target function, as well as the aforementioned evaluation parameters (RMSD radius in cycle 1, RMSD drift, and the combination thereof). This is in good agreement with the results obtained from the modified data sets, where enth is one of the rather stable structure calculations which yields an accurate structure bundle up to 15 % missing chemical shifts (Fig. 4). In the two other cases, the structural quality drops significantly when compared to the results obtained from refined peak lists, however, the RMSD bias is still below 3.0 Å and the global fold is thus considered correct. In all three cases, the final CYANA target function increases and the RMSD radius decreases when using unrefined peak lists. This can be attributed to an increased number of potentially incorrect long-range restraints that result from artifact peaks. The combined criterion gives a good indication about the structural quality.
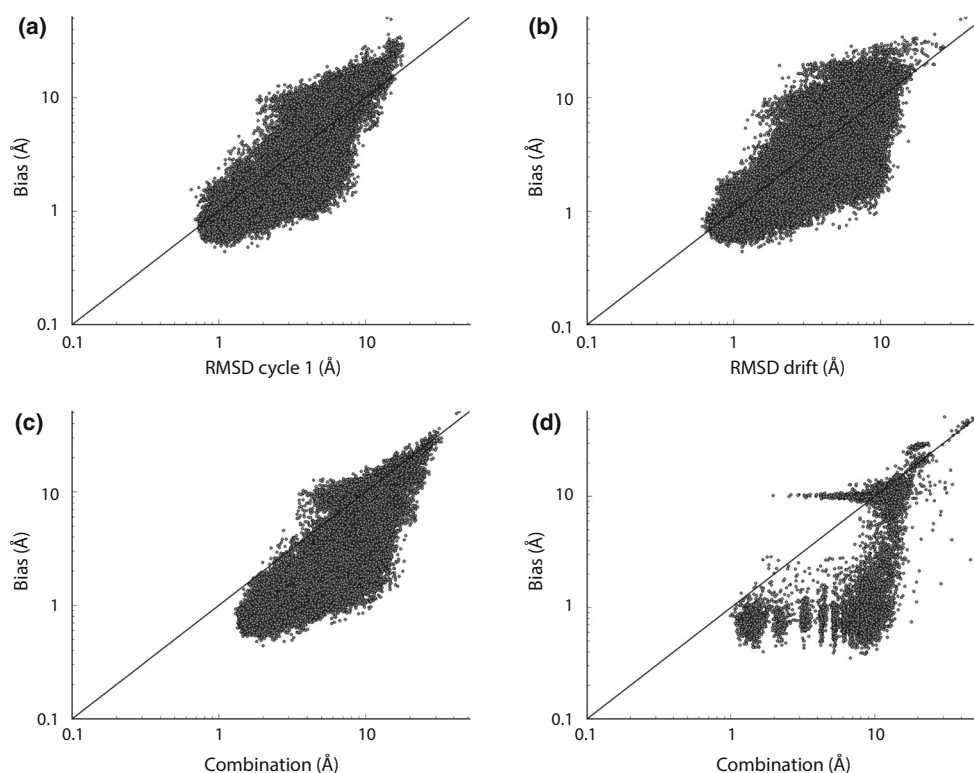
Fig. 7 Structural accuracy plotted against commonly applied evaluation criteria for combined automated NOESY assignment and structure calculation runs. The accuracy is represented by the RMSD bias, i.e. the RMSD between the mean structure of the bundle and the mean reference structure. Every data point represents one combined automated NOESY assignment and structure calculation run. **a** Initial convergence measured by the RMSD to the mean structure of the structure bundle from the first structure calculation cycle, **b** RMSD drift measured by the RMSD between the final structure bundle and the structure bundle of the first cycle, **c** a combination of the two criteria calculated as $\sqrt{(1.5R)^2 + D^2}$, where R denotes the RMSD radius in cycle 1 and D the RMSD drift, for all proteins except the homodimeric ww2d, and **d** same as in **c** for the structure calculations of the protein ww2d

Table 2 Structure calculation results using refined and unrefined NOESY peak lists

	enth		fsh2		rhod	
	Refined	Unrefined	Refined	Unrefined	Refined	Unrefined
RMSD bias (Å)	0.90	1.00	0.67	2.11	0.40	1.09
RMSD radius (Å)	0.53	0.25	0.52	0.22	0.76	0.26
CYANA target function (Å^2)	1.18	3.95	1.84	15.99	0.95	6.17
RMSD radius in cycle 1 (Å)	1.21	0.75	1.10	1.60	0.89	1.48
RMSD drift (Å)	1.01	0.96	1.10	2.46	0.82	2.26
Combined criterion	2.08	1.48	1.99	3.44	1.57	3.18

The combined criterion is given as $\sqrt{(1.5R)^2 + D^2}$, where R denotes the RMSD radius in cycle 1 and D the RMSD drift

Conclusions

The results presented in this study clearly show that imperfections within the chemical shift assignment can cause severe problems during NOE assignment and structure calculation. In most of the data sets tested 10 % of missing or erroneous chemical shifts result in inaccurate structures with RMSD bias values above 3 Å. In some cases of high quality data and large amounts of 3D peaks, higher percentages of missing or erroneous chemical shifts can be tolerated. Less severe problems arise from missing peaks, errors in peak positions and volumes as well as lower resolution simulated by using higher assignment tolerances. Furthermore, it was shown that data imperfections

cannot be overcome by longer computation times. The convergence of the initial structure calculation cycle and the RMSD drift between the first and the last cycle can be combined in a weighted average and used as an indication for the reliability of a structure calculation result.

References

Altieri AS, Byrd RA (2004) Automation of NMR structure determination of proteins. Curr Opin Struct Biol 14:547–553

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

Baran MC, Huang YJ, Moseley HNB, Montelione GT (2004) Automated analysis of protein NMR assignments and structures. Chem Rev 104:3541–3555

Bartels C, Güntert P, Billeter M, Wüthrich K (1997) GARANT—a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. J Comput Chem 18:139–149

Billeter M, Wagner G, Wüthrich K (2008) Solution NMR structure determination of proteins revisited. J Biomol NMR 42:155–158

Calzolai L, Lysek DA, Perez DR, Güntert P, Wüthrich K (2005) Prion protein NMR structures of chickens, turtles, and frogs. Proc Natl Acad Sci USA 102:651–655

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Duggan BM, Legge GB, Dyson HJ, Wright PE (2001) SANE (Structure assisted NOE evaluation): an automated model-based approach for NOE assignment. J Biomol NMR 19:321–329

Gronwald W, Kalbitzer HR (2004) Automated structure determination of proteins by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 44:33–96

Gronwald W, Moussa S, Elsner R, Jung A, Ganslmeier B, Trenner J, Kremer W, Neidig KP, Kalbitzer HR (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287

Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. Q Rev Biophys 44:257–309

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

Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125

Güntert P (2004) Automated NMR structure calculation with CYANA. Meth Mol Biol 278:353–378

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

Güntert P, Berndt KD, Wüthrich K (1993) The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. J Biomol NMR 3:601–606

Güntert P, Mumenthaler C, 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, Wüthrich K (2002a) 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

Herrmann T, Güntert P, Wüthrich K (2002b) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24:171–189

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

Huang YJ, Tejero R, Powers R, Montelione GT (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins 62:587–603

Hung LH, Samudrala R (2006) An automated assignment-free Bayesian approach for accurately identifying proton contacts from NOESY data. J Biomol NMR 36:189–198

Ikeya T, Jee J-G, Shigemitsu Y, Hamatsu J, Mishima M, Ito Y, Kainosho M, Güntert P (2011) Exclusively NOESY-based automated NMR assignment and structure determination of proteins. J Biomol NMR 50:137–146

Jee J, 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 AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57

Kirchner DK, Güntert P (2011) Objective identification of residue ranges for the superposition of protein structures. BMC Bioinformatics 12:170

Kuszewski J, Schwieters CD, Garrett DS, Byrd RA, Tjandra N, Clore GM (2004) Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments. J Am Chem Soc 126:6258–6273

López-Méndez B, 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, Seki E, Matsuda T, Hirota H, Yoshida M, Tanaka A, Osanai T, Seki M, Shinozaki K, Yokoyama S, Güntert P (2004) NMR assignment of the hypothetical ENTH-VHS domain At3g16270 from *Arabidopsis thaliana*. J Biomol NMR 29:205–206

Luginbühl P, Szyperski T, Wüthrich K (1995) Statistical basis for the use of $^{13}C^a$ chemical shifts in protein structure determination. J Magn Reson B 109:229–233

Meadows RP, Olejniczak ET, Fesik SW (1994) A computer-based protocol for semiautomated assignments and 3D structure determination of proteins. J Biomol NMR 4:79–96

Moseley HNB, Montelione GT (1999) Automated analysis of NMR assignments and structures for proteins. Curr Opin Struct Biol 9:635–642

Mumenthaler C, Braun W (1995) Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. J Mol Biol 254:465–480

Mumenthaler C, Güntert P, Braun W, Wüthrich K (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. J Biomol NMR 10:351–362

Nabuurs SB, Spronk CAEM, Vuister GW, Vriend G (2006) Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. PLoS Comp Biol 2:71–79

Nilges M (1995) Calculation of protein structures with ambiguous distance restraints—automated assignment of ambiguous NOE crosspeaks and disulfide connectivities. J Mol Biol 245:645–660

Nilges M, Macias MJ, ODonoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. J Mol Biol 269:408–422

Ohnishi S, Güntert P, Koshiba S, Tomizawa T, Akasaka R, Tochio N, Sato M, Inoue M, Harada T, Watanabe S, Tanaka A, Shirouzu M, Kigawa T, Yokoyama S (2007) Solution structure of an atypical WW domain in a novel b-clam-like dimeric form. FEBS Lett 581:462–468

Pääkkönen K, Tossavainen H, Permi P, Rakkolainen H, Rauvala H, Raulo E, Kilpeläinen I, 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, Matsuda T, Hirota H, Yoshida M, Tanaka A, Osanai T, Seki M, Shinozaki K, Yokoyama S, Güntert P (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, Yokoyama S, Güntert P (2005) Solution structure of the rhodanese homology domain At4g01050 (175-295) from *Arabidopsis thaliana*. Protein Sci 14:224–230

Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, 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 JF, Giachetti A, Guerry P, Güntert P, Herrmann T, Huang YJ, Jonker HRA, Mao B, Malliavin TE, Montelione GT, Nilges M, Raman S, van der Schot G, Vranken WF, Vuister GW, Bonvin AMJJ (2009) CASD-NMR: critical assessment of automated structure determination by NMR. Nat Methods 6:625–626

Rosato A, Aramini JM, Arrowsmith C, Bagaria A, Baker D, Cavalli A, Doreleijers JF, Eletsky A, Giachetti A, Guerry P, Gutmanas A, Güntert P, He YF, Herrmann T, Huang YPJ, Jaravine V, Jonker HRA, Kennedy MA, Lange OF, Liu GH, Malliavin TE, Mani R, Mao BC, Montelione GT, Nilges M, Rossi P, van der Schot G, Schwalbe H, Szyperski TA, Vendruscolo M, Vernon R, Vranken WF, de Vries S, Vuister GW, Wu B, Yang YH, Bonvin AMJJ (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20:227–236

Schmidt E, Güntert P (2012) A new algorithm for reliable and general NMR resonance assignment. J Am Chem Soc 134:12817–12829

Schmidt E, 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, Güntert P (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, Güntert P (2005) Solution structure of the Src homology 2 domain from the human feline sarcoma oncogene Fes. J Biomol NMR 31:357–361

Vögeli B, Segawa TF, Leitz D, Sobol A, Choutko A, Trzesniak D, van Gunsteren W, Riek R (2009) Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. J Am Chem Soc 131:17215–17225

Vögeli B, Kazemi S, Güntert P, Riek R (2012) Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. Nat Struct Mol Biol 19:1053–1057

Williamson MP, Craven CJ (2009) Automated protein structure calculation from NMR data. J Biomol NMR 43:131–143

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

Zhang Z, Porter J, Tripsianes K, Lange OF (2014) Robust and highly accurate automatic NOESY assignment and structure determination with Rosetta. J Biomol NMR 59:135–145

Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage l N-peptide/*boxB* RNA complex. J Am Chem Soc 119:6711–6721