

Influence of NMR Data Completeness on Structure Determinations of Homodimeric Proteins

Yi-Jan Lin,^{a,*} Teppei Ikeya,^b Donata K. Kirchner^{c,d} and Peter Güntert^{b,c,d}^aGraduate Institute of Natural Products, Center for Research Resources and Development, Center of Excellence for Environmental Medicine, Center for Infectious Disease and Cancer Research, Kaohsiung Medical University, No.100, Shi-Chuan 1st Road, San-Ming District, Kaohsiung 807, Taiwan^bDepartment of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, Tokyo 192-0397, Japan^cInstitute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany^dFrankfurt Institute for Advanced Studies, Goethe University Frankfurt am Main, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

(Received: Feb. 21, 2014; Accepted: Jun. 25, 2014; Published Online: Aug. 27, 2014; DOI: 10.1002/jccs.201400095)

The structure determination of homodimeric proteins by NMR using conventional NOESY experiments is still challenging due to the degeneracy of the chemical shifts in the identical monomers, which causes ambiguity in the NOE assignments. Residues involved in the interface between two monomers provide essential intermolecular NOEs for the structure determinations of homodimeric proteins. Hence NMR data, such as NOE peak lists and chemical shift assignments of these interface residues, play a crucial role for the successful structure determination of homodimeric proteins. This paper extends our previous report (Lin, Y.-J.; Kirchner, D. K.; Güntert, P. *J. Magn. Reson.* **2012**, *222*, 96) and investigates the influence of incomplete NOESY peak lists combined with incomplete ¹H chemical shift assignments of the interface residues on the structure determination of homodimeric proteins using the program CYANA. Data incompleteness was simulated by random omission of both NOESY cross peaks and interface ¹H chemical shifts. Our results for three proteins with different percentages of interface residues reveal that the algorithm can tolerate about 40–50% NOESY peak omission with complete interface chemical shift assignments, which indicates that partial NOESY peak omission does not cause severe problems when the interface chemical shifts are completely assigned. Combining NOESY peak omission with incomplete interface chemical shift assignments, the tolerance for interface chemical shift omission decreases with the extent of omitted NOESY peaks. The tolerance for unassigned interface side chain, methyl and aromatic chemical shifts is affected more strongly by NOESY peak omission than that for the omission of general interface ¹H chemical shifts including the backbone. In general about 10–30% peaks omission is tolerated in conjunction with missing chemical shift assignments. If more NOESY peaks are omitted calculations gradually become unstable and tend not to tolerate any missing interface chemical shifts. A large amount of omitted NOESY peaks, for instance 30% omission in our calculations, could decrease the tolerance for missing aromatic or methyl interface ¹H chemical shifts to as few as 2–4 missing chemical shifts, suggesting that complete aromatic and methyl ¹H chemical shift assignments are important when the NOESY peak data is significantly incomplete. Finally, for homodimeric proteins with a low percentage of interface residues, our results reveal that the omission of NOESY peaks, even at an extent of only 10%, can result in no tolerance against the omission of interface ¹H chemical shifts, suggesting that the completeness of both interface ¹H chemical shift assignments and NOESY peaks are important for the successful structure determination of proteins with a small homodimer interface.

Keywords: Homodimeric protein; NMR structure determination; NOESY; Resonance assignment; CYANA.

INTRODUCTION

Two-thirds of human enzymes are oligomers, and in *E. coli* the average oligomerization state of proteins is

four.¹ Although the efficiency of NMR assignments, such as automated chemical shift assignment and NOE assignment, and automated structure calculation methods for

* Corresponding author. Tel.: +886-7-312-1101 ext. 6922; Fax: +886-7-269-3467; Email: yjlin@kmu.edu.tw

monomeric proteins have been improved and become reliable and routine,²⁻⁴ the determination of homodimeric protein structures by NMR is still challenging and hence the number of NMR structures of homodimeric proteins remains small. Only about 4.4% of all solution NMR protein structures and 2.9% of all solid-state NMR protein structures in the PDB are symmetric dimmers,⁵ which is a small fraction compared to the 27.0% homodimers among all X-ray protein structures. Nevertheless, several approaches and investigations for combined automated NOESY assignment and structure calculation have been developed and applied to symmetric oligomers, especially symmetric homodimers.⁵⁻⁷ Some factors still limit the application of the automated methods for combined automatic NOESY assignment and structural calculation⁸⁻¹⁷ to homodimeric proteins, e.g. the availability of inter-monomeric restraints or information on the interface between two monomers, the completeness of the chemical shift assignments, especially for the important interface chemical shifts,¹⁸ and the difficulty to distinguish inter-monomeric from intra-monomeric NOEs due to the degeneracy of the chemical shifts in the symmetric dimers.

The structure determination of proteins, including homodimeric proteins, is based on assigned chemical shifts and NOE peak lists. In order to discuss how these data affect the structure calculation, we investigated the influence of incomplete NOESY peak picking with and without missing ¹H chemical shift assignments of interface residues on the NMR structure determination of three homodimeric proteins using the CYANA program,^{3,12} as described in the experimental section.

RESULTS

Random omission of NOESY peaks without chemical shifts omission from interface residues

In the structure calculations, NOESY peaks were removed in steps of 10% until severely incorrect structures (either two separated monomers or dimers with an interface radically different from the reference structure) were obtained. The maximal tolerances for NOESY peak omission are similar for the three test proteins NikA, WW, and REI (see experimental section), when NOE assignments and structure calculations were carried out with complete interface chemical shift assignments. The maximal allowed NOESY peak omission is 40% in WW, and 50% in NikA and REI. The RMSD values between the mean structures and the mean reference structures at various different per-

centages of NOESY peak omission are within 1.5 Å for NikA, 1.1 Å for WW, and 1.8 Å for REI. These data show that the random omission of NOESY peaks, which corresponds to incomplete NOESY peak picking, does not cause severe problems if the interface chemical shifts are completely assigned, even though the three proteins contain different percentages of interface residues, i.e. 67% in NikA, 63% in WW, and 27% in REI.¹⁸

Random omission of NOESY peaks and incomplete ¹H chemical shift assignments of interface residues

For NikA, successful structure calculations could be achieved for up to 20% interface ¹H chemical shift omission provided that the NOESY peaks were complete.¹⁸ With up to 5% ¹H interface chemical shift omission, successful calculations can tolerate up to 50% NOESY peak omission (Fig. 1a), and the RMSD value between the mean structures and the mean reference structure (accuracy) remains reasonably small (Fig. 1b). At 10% and 15% ¹H interface chemical shift omission, the tolerance decreases to 20% NOESY peak omission (Fig. 1a). In some of these calculations β -sheets were disordered, which nevertheless could be corrected by further refinement. At 20% ¹H interface chemical shift omission, the tolerance for NOESY peak omission is only 10% (Fig. 1a). The relationship between the maximally allowed peak omission and maximally allowed interface ¹H chemical shift omission is shown in Fig. 1g. For calculations where the NOESY peak omission is above the maximal tolerance, the deviations from the reference structure increased significantly (Fig. 1a) and incorrect structures were obtained, although some of them still showed low bundle RMSDs and reasonable target function values (data not shown), which was also observed in the previous report.¹⁸

For WW, the maximally tolerated interface chemical shift omission is 25% without peak omission. At 5–15% interface chemical shift omission, the maximally tolerated NOESY peak omission is 40%. The tolerated NOESY peak omission decreased to $\leq 20\%$ peak omission for 20% interface chemical shift omission (Fig. 1c). At 25% interface chemical shift omission, even 10% peak omission was not tolerated (Fig. 1c). For calculations with the NOESY peak omission within the maximal tolerance, the RMSD values between the mean structures and the mean reference structure (accuracy) are in a reasonable range (Fig. 1d). The relationship between the maximally tolerated peak omission and interface chemical shift omission is shown in Fig. 1g.

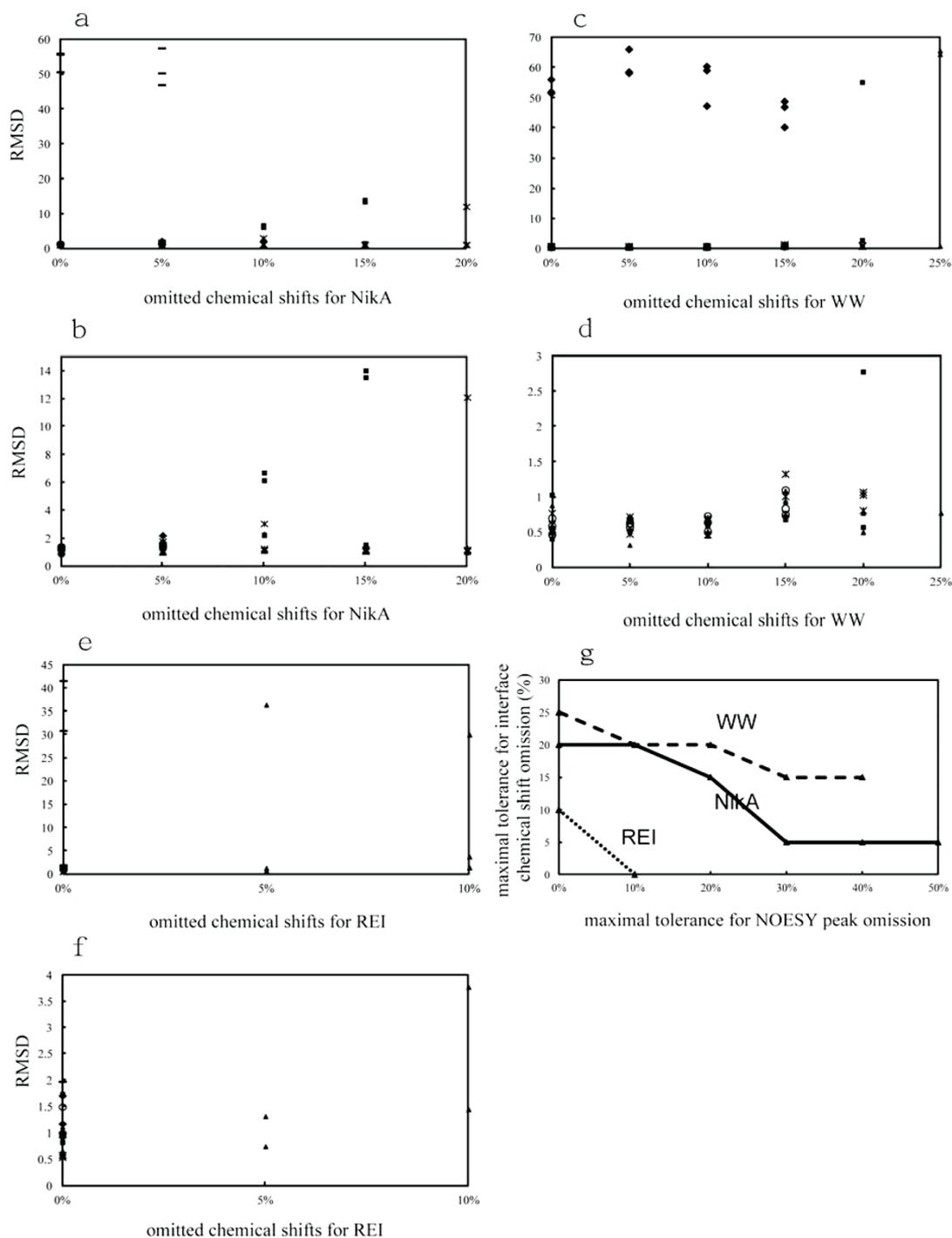


Fig. 1. Results of structure calculations of three homodimeric proteins, NikA (a, b), WW (c, d), and REI (e, f), at different percentages of omitted interface chemical shift assignments combined with 10% (triangle), 20% (star), 30% (square), 40% (circle), 50% (diamond), and 60% (minus) NOESY peak omission. Three independent runs were performed for each percentage of omitted chemical shifts and NOESY peaks. The horizontal axis indicates the percentage of randomly omitted interface chemical shifts. The vertical axis indicates the RMSD value between the mean structure from each calculation and the mean reference structure. (b), (d), and (f) are vertical expansions of (a), (c), and (e), respectively, for better visibility of the low RMSD range. RMSD values were calculated for the backbone atoms N, C $^{\alpha}$, and C' of residues 16–50 in NikA, 18–39 in WW, and 4–106 in REI. (g) Relationship between the maximally allowed peak omission and the maximally allowed missing interface ^1H chemical shifts.

For REI, at 5% ^1H interface chemical shift omission, even 10% NOESY peak omission caused incorrect structures in the form of an erroneous ensemble with two separated monomers and hence significantly increased RMSD to the reference structure (Fig. 1e). The relationship between the maximally tolerated peak omission and interface chemical shift omission is shown in Fig. 1g.

Random omission of NOESY peaks and incomplete side chain ^1H chemical shift assignments of interface residues

Since the backbone and H^β resonances can be assigned using spectra for backbone resonance assignment, the test calculations were performed with omission of side-chain resonances beyond H^β , which are harder to assign than the backbone resonances. For Nika, successful calculations without NOESY peak omission could be achieved with up to 45% omission of side chain interface chemical shifts (H^β not included). With 10% and 20% side chain interface chemical shift omission, the tolerance for NOESY peak omission is 50% (Fig. 2a), although in some calculations β -sheets were disordered, which could be corrected by further refinement. At 30% side chain interface chemical shift omission, the tolerance for NOESY peak omission decreased to 20%, whereas at 40% side chain interface chemical shift omission, the tolerance for NOESY peak omission increased again to 30%, which suggests that calculations already became unstable under large omission of both interface side chain chemical shift omission and NOESY peak omission (Fig. 2a). Therefore we conclude that with up to 40% side chain interface chemical shift omission, the tolerance for NOESY peak omission is 20%. At 45% side chain interface chemical shift omission, already 10% NOESY peak omission caused incorrect structures that deviated strongly from the reference structure, although the ensembles showed low RMSD values to their mean coordinates and low target function values.

For WW, the maximal tolerance for NOESY peak omission is 30% with up to 30% side chain interface chemical shift omission (Fig. 2c). Although calculations with 20% side chain interface chemical shift omission can tolerate 40% NOESY peak omission (Fig. 2c), one of three repeats starting from different random structures showed separated structures with two monomers in cycle 1, suggesting that calculations under this condition may become unstable. Therefore the tolerance for NOESY peak omission at 20% side chain interface chemical shift omission is

regarded as 30%. Even 10% peak omission was not tolerated at 40% and 50% interface chemical shift omission (Fig. 2c).

For REI, the tolerance for NOESY peak omission is 40% at 10–20% side chain interface chemical shift omission. It decreased to 10% at 30% and 40% side chain interface chemical shift omission.

For the calculations within the tolerance for all three proteins, the RMSD values between the mean structures and the mean reference structure are in a reasonable range (Figs. 2b, 2d, 2f) and the relationship between the maximal allowed peak tolerance and interface side chain chemical shift omission is shown in Fig. 2g.

Random omission of NOESY peaks and incomplete methyl ^1H chemical shift assignments of interface residues

For Nika, tolerance of NOESY peak omission with 10–20% interface methyl ^1H chemical shift omission is 50%. The tolerance for peak omission gradually decreased to 40% NOESY peak omission for 30% interface methyl ^1H chemical shift omission and then decreased to 20% and 10% NOESY peak omission for 40% and 50% interface methyl ^1H chemical shift omission, respectively. The tolerance for NOESY peak omission is 10–20% at 60–80% interface methyl ^1H chemical shift omission (Fig. 3a). In the calculations with 60–80% interface methyl ^1H chemical shift omission the helices of Nika remained essentially unchanged, but the β -sheets were often disordered, which resulted in an increase of the RMSD values to the reference structure (Fig. 3b). The disorder of the β -sheet could be corrected by further refinement. With 90% interface methyl ^1H chemical shift omission, even 10% NOESY peak omission caused wrong ensembles with low RMSDs to their mean coordinates (Fig. 3b) and reasonable target function values.

For WW, the tolerance of NOESY peak omission is 30% with 20% (corresponding to 1 omitted methyl ^1H chemical shift) and 40% interface methyl ^1H chemical shift omission and then decreased to 20% NOESY peak omission with 60% interface methyl ^1H chemical shift omission (Fig. 3c, 3d).

For REI, the tolerance is 40% NOESY peak omission with up to 40% interface methyl ^1H chemical shift omission, and the tolerance slightly decreased to 30% NOESY peak omission at 50% interface methyl ^1H chemical shift omission (Figs. 3e, 3f). The relationship between the maxi-

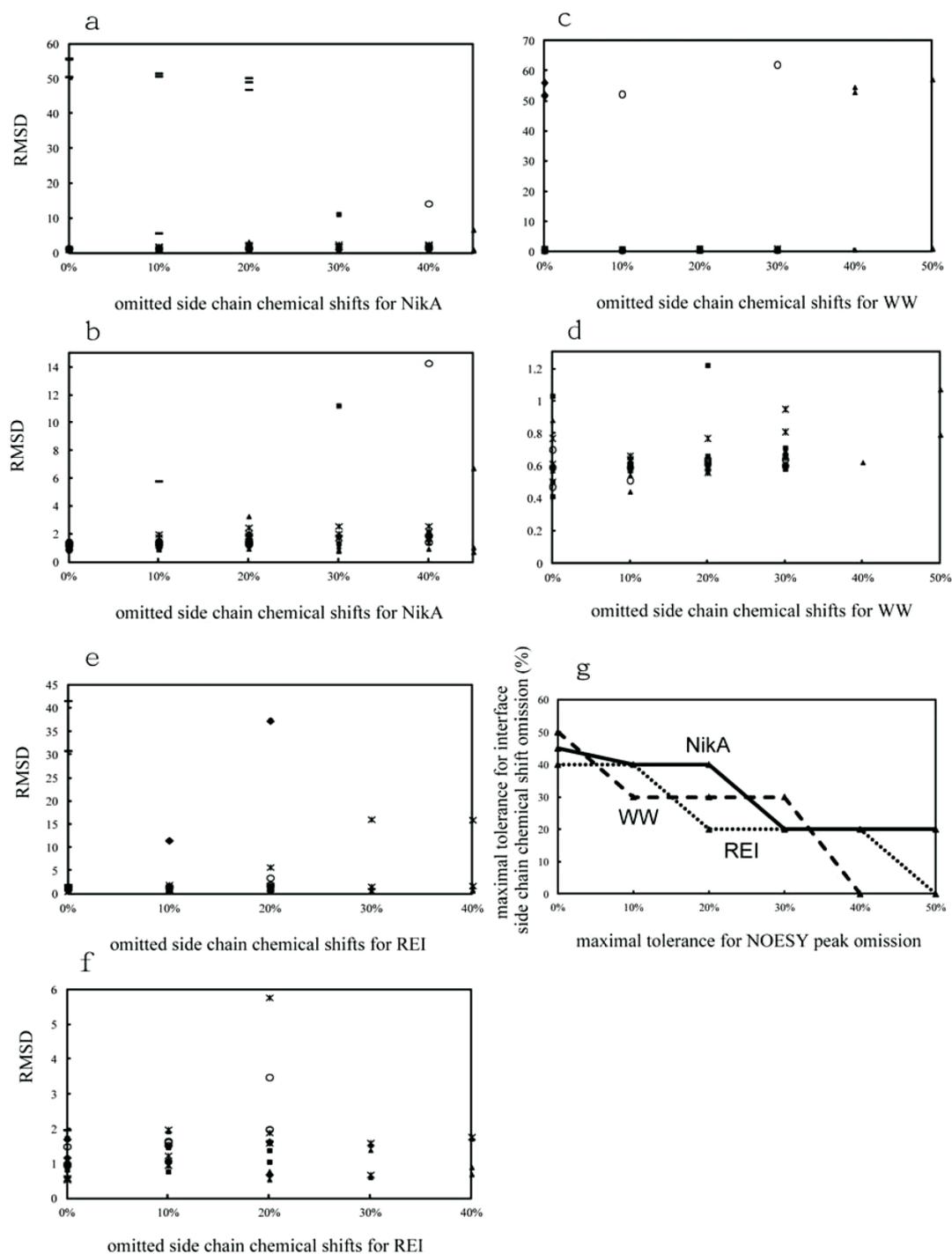


Fig. 2. Results of structure calculations of three homodimeric proteins, NikA (a, b), WW (c, d), and REI (e, f), at different percentages of omitted interface side chain chemical shift assignments combined with 10% (triangle), 20% (star), 30% (square), 40% (circle), 50% (diamond), and 60% (minus) NOESY peak omission. Three independent runs were performed for each percentage of omitted chemical shifts and NOESY peaks. The horizontal axis indicates the percentage of randomly omitted chemical shifts. The vertical axis indicates the RMSD value between the mean structure from each calculation and the mean reference structure. (b), (d), and (f) are expansions of (a), (c), and (e), respectively, for better vision. RMSD values were calculated for the backbone atoms, N, C $^{\alpha}$ and C $^{\prime}$ of residues 16–50 in NikA, 18–39 in WW and 4–106 in REI, respectively. (g) Relationship between the maximally allowed peak omission and the maximally allowed missing interface side chain chemical shifts.

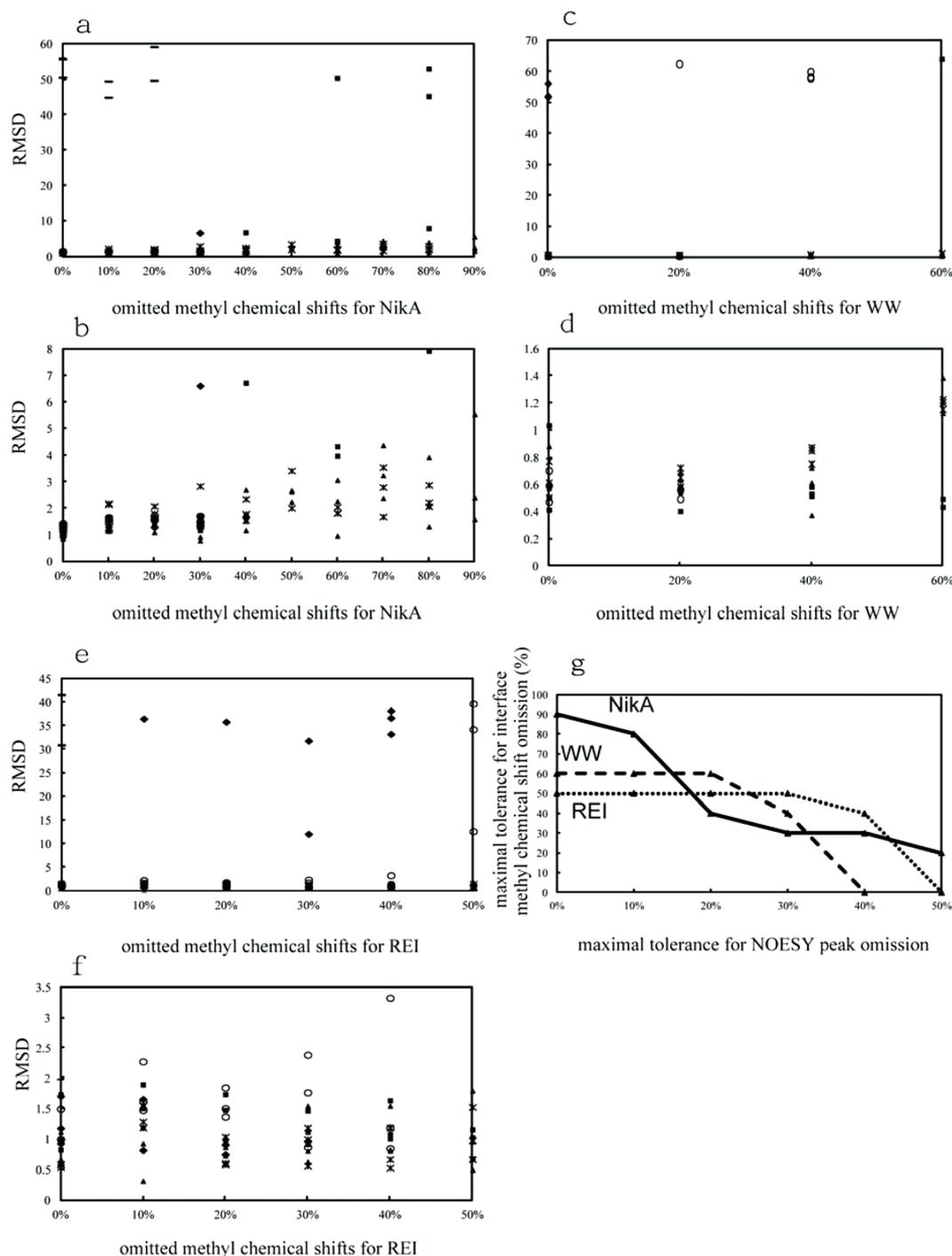


Fig. 3. Results of structure calculations of three homodimeric proteins, Nika (a, b), WW (c, d), and REI (e, f), at different percentages of omitted interface methyl chemical shift assignments combined with 10% (triangle), 20% (star), 30% (square), 40% (circle), 50% (diamond), and 60% (minus) NOESY peak omission. Three independent runs were performed for each omitted percentage of chemical shifts and NOESY peaks. The horizontal axis indicates the percentage of randomly omitted chemical shifts. The vertical axis indicates the RMSD value between the mean structure from each calculation and the mean reference structure. (b), (d), and (f) are expansions of (a), (c), and (e), respectively, for better visibility. RMSD values were calculated for the backbone atoms, N, C $^{\alpha}$ and C' of residues 16–50 in Nika, 18–39 in WW, and 4–106 in REI. (g) Relationship between the maximally allowed peak omission and maximally allowed missing interface methyl chemical shifts.

mal.

Random omission of 40% NOESY peaks with incomplete ^1H chemical shift assignments of non-interface residues

Results showed that calculations tolerate 50% omission of ^1H chemical shifts of the non-interface residues in REI, whereas calculations can tolerate 100% omission of the ^1H chemical shifts of the non-interface residues in NikA and WW. The high tolerance for ^1H chemical shift omission of the non-interface residues in NikA and WW is probably due to the low ratio of non-interface residues in the structured regions excluding the non-interface residues in the flexible chain termini, i.e. residues 1–15 in NikA, and 1–13 and 44–49 in WW, which do almost not participate in any long-range intra-monomeric NOEs. NikA contains 24 dimer interface residues out of 51 residues per monomer, and WW contains 19 interface residues out of 49 residues per monomer. This means that NikA and WW contain only $51 - 24 - 15 = 12$ and $49 - 13 - 6 = 11$ non-interface residues in their structured regions, respectively. In contrast, REI contains 29 interface residues out of a total of 107 residues in one monomer and 78 non-interface residues in the structured regions. The percentages of non-interface residues in the structured regions are 23.5%, 22.4%, and 72.9% in NikA, WW, and REI, respectively.

DISCUSSION

As reported in a previous paper,¹⁸ particular attention should be paid to those runs that show a high RMSD bias to the reference structure but low RMSDs within the ensemble and low target function values, which lets them appear as well-converged structures. Without interface chemical shift omission, the maximally allowed NOESY peak omission is 40% in WW and 50% in NikA and REI. This indicates that for all three proteins the random omission of NOESY peaks does not cause severe problems and very different results when the interface chemical shifts are completely assigned, in spite of the different percentages of interface residues in the three proteins, i.e. 67% in NikA, 63% in WW, and 27% in REI.

Overall, the tolerance for all ^1H , side chain, methyl, and aromatic proton chemical shift omission is approximately inversely proportional to the percentage of omitted NOESY peaks. Especially, 10–30% omitted NOESY peaks significantly affected the tolerance against interface side chain chemical shift omission, for which percentages dropped from 40% to 20% in NikA, 50% to 30% in WW,

and 40% to 20% in REI. This significant decrease is probably due to the fact that side chain protons, including methyl and aromatic protons, are often involved in long-range intermolecular interactions and therefore more sensitive to incomplete NOESY peak picking. These data suggest the importance of identifying the NOESY peak as completely as possible when interface side chain proton chemical shift assignments are incomplete in homodimeric proteins. A similar effect is also observed for interface methyl and aromatic chemical shift omission, especially in NikA, where the tolerance for unassigned interface methyl chemical shifts dropped significantly from 80% to 40% with 20% omitted NOESY peaks (Fig. 3g). For the tolerance of interface ^1H (containing backbone and side chain protons) chemical shift omission, the tolerance decreased smoothly in NikA and WW, whereas in REI, 10% omitted NOESY peaks caused wrong ensembles even at 5% interface ^1H chemical shift omission. This indicates that REI cannot tolerate any omission of interface ^1H chemical shift together with NOESY peak omission, or reversely, cannot tolerate any NOESY peak omission with simultaneous omission of interface ^1H chemical shifts. It was reported that for REI the tolerance for unassigned interface ^1H chemical shifts is only 10% with complete NOESY peak lists.¹⁸ This may suggest that for homodimeric proteins with a low percentage of interface residues, such as REI, the completeness of both interface ^1H chemical shift assignments and NOESY peaks is particularly important for successful NMR structure determination, whereas proteins with a larger interface can tolerate more incompleteness.

The maximally allowed percentage of NOESY peak omission is 40% in WW and 50% in NikA and REI with complete interface chemical shift assignments. Interestingly, at 50% NOESY peak omission combined with various different kinds of interface chemical shift incompleteness, calculations of NikA can still tolerate 5% interface ^1H chemical shift omission, 20% interface side chain ^1H chemical shift omission, 20% interface methyl ^1H chemical shift omission, and 40% interface aromatic ^1H chemical shift omission (Figs. 1g, 2g, 3g, 4g), whereas for REI, 50% NOESY peak omission can only tolerate 20% interface aromatic ^1H chemical shift omission (Figs. 1g, 2g, 3g, 4g). For WW, with 40% NOESY peak omission calculations only tolerate 15% missing interface ^1H chemical shifts (Figs. 1g, 2g, 3g, 4g). Among these three proteins, NikA can tolerate a higher degree of missing chemical shifts than WW and REI if the NOESY peak lists are incomplete.

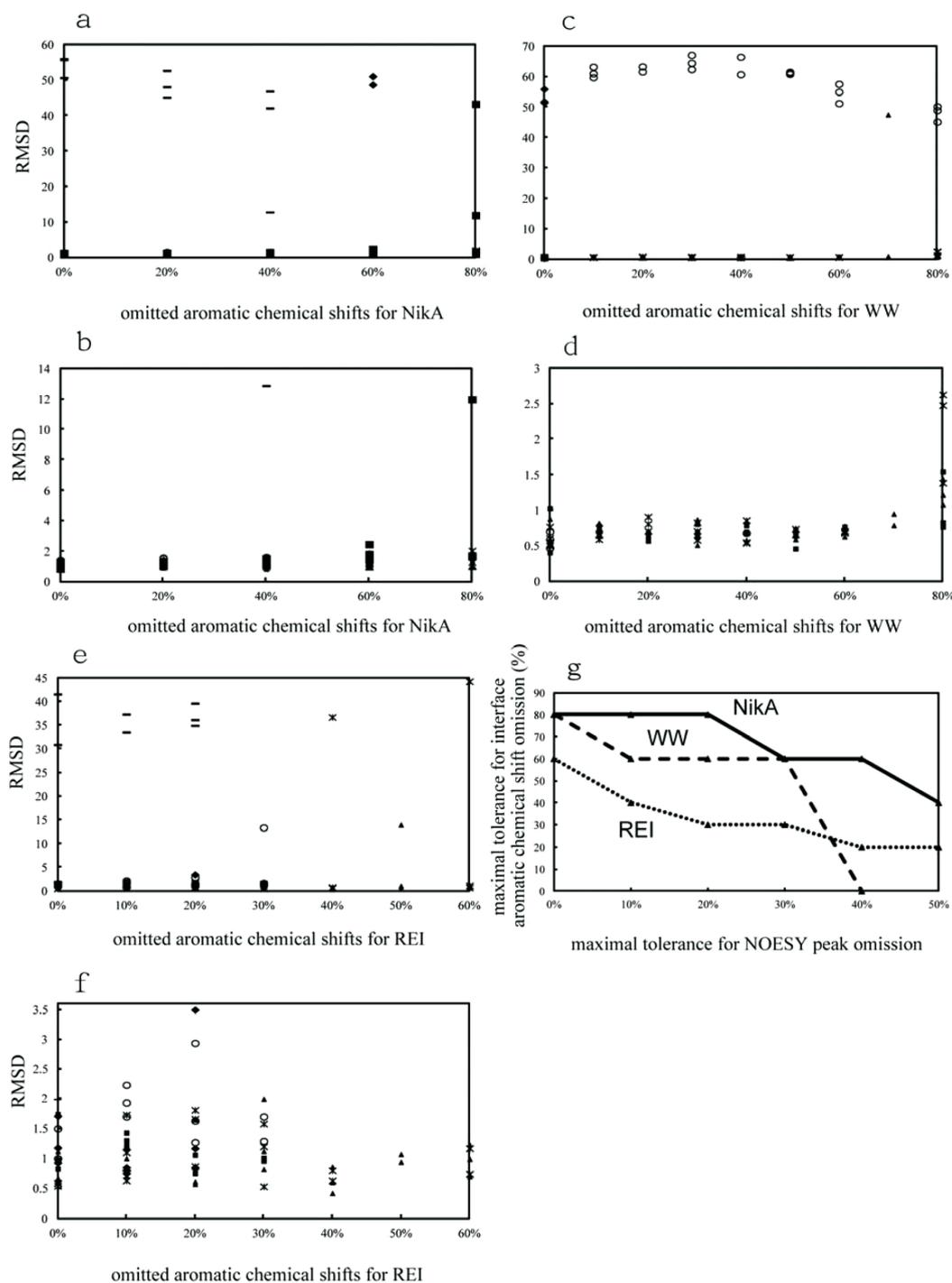


Fig. 4. Results of structure calculations of three homodimeric proteins, Nika (a, b), WW (c, d), and REI (e, f), at different percentages of omitted interface aromatic chemical shift assignments combined with 10% (triangle), 20% (star), 30% (square), 40% (circle), 50% (diamond), and 60% (minus) NOESY peak omission. Three independent runs were performed for each percentage of omitted chemical shifts and NOESY peaks. The horizontal axis indicates the percentage of randomly omitted chemical shifts. The vertical axis indicates the RMSD value between the mean structure from each calculation and the mean reference structure. (b), (d), and (f) are expansions of (a), (c), and (e), respectively, for better visibility. RMSD values were calculated for the backbone atoms, N, C $^{\alpha}$ and C' of residues 16–50 in Nika, 18–39 in WW, and 4–106 in REI. (g) Relationship between the maximally allowed peak omission and the maximally allowed missing interface aromatic chemical shifts.

A previous report¹⁸ revealed that with complete NOESY peak lists incorrect structures were obtained with 80% omission, corresponding to 4 methyl ¹H chemical shifts, of all assigned methyl ¹H chemical shifts in WW or 70%, corresponding to 9 aromatic ¹H chemical shifts, of all assigned aromatic ¹H chemical shifts in REI. This means that successful calculations only tolerate 60% omission of all assigned methyl ¹H chemical shifts (corresponding to 3 missing methyl ¹H chemical shifts) in WW or 60% omission of aromatic ¹H chemical shift assignments (corresponding to 8 aromatic ¹H chemical shifts) in REI, which indicates that aromatic and methyl interface proton chemical shifts play important roles in the structure determination of homodimeric proteins. The lack of a small number of “essential” chemical shifts can lead to incorrect structures. When combined with NOESY peak omission, for example, with 30% NOESY peak omission, the tolerance for successful calculations dropped to 40% omission (corresponding to 2 methyl ¹H chemical shifts) of all assigned methyl ¹H chemical shifts for WW and to 30% (corresponding to 4 aromatic ¹H chemical shifts) for REI, respectively (Figs. 3g, 4g). Already a few missing methyl or aromatic ¹H chemical shifts out of total ¹H chemical shifts combined with a high percentage of missing NOE peaks could cause severe problems for the structure calculation of homodimeric proteins.

EXPERIMENTAL

Structure calculations of homodimeric proteins were carried out with the program CYANA using a probabilistic algorithm for automated NOESY assignment³ and structure calculations with torsion angle dynamics.¹⁹ As described earlier,¹⁸ the homodimer symmetry was explicitly taken into account for the network-anchoring of the NOE assignments, for ensuring an identical conformation of the two monomers by dihedral angle difference restraints for all corresponding torsion angles, and for maintaining a symmetric relative orientation of the two monomers by distance difference restraints between symmetry-related intermolecular C^α-C^α distances.^{5,20}

The test calculations were performed by randomly omitting different percentages of NOESY peaks combined with random omission of assigned chemical shifts of interface residues. The experimental NMR data of the DNA binding protein Nika (PDB code: 2BA3),^{20,21} the second WW domain from the mouse salvador homolog 1 protein (PDB code: 2DWV),²² and data simulated from the X-ray structure of the variable portions of the Bence-Jones protein REI (PDB code: 1REI)²³ were used for homodimer

structure calculations with CYANA. Nika is an α/β protein of 51 amino acid residues with one β -strand and two α -helices per monomer. Two monomers form a dimer with a two-stranded β -sheet and a four-helix bundle. The 49 residue WW domain forms a homodimer with a β -clam-like motif. REI is an all β -sheet protein with 107 amino acid residues in each monomer. The dimer interface residues were defined as reported previously.¹⁸ residues 14–24, 26–27, 29–30, 41–42, 44–46, and 48–51 in Nika, residues 16–21, 23, 25–27, 29, 31–33, and 36–40 in WW, and residues 1, 6, 34, 36, 38, 41–46, 49–50, 53, 55–56, 85, 87, 89, 91, 94–101, and 103 in REI. The input data and methods for test calculations with chemical shift omission of interface residues and complete NOESY peak lists were described in a previous report.¹⁸ Based on the previously reported maximal tolerances for interface chemical shift omission without NOESY peak omission, calculations were carried out within these reported maximal tolerances, but combined with NOESY peak omission. In these calculations, NOESY peaks were removed in steps of 10% and interface chemical shifts were removed in steps of 5% for all interface ¹H chemical shifts, 10% for interface side chain, methyl and aromatic ¹H chemical shifts, or in steps of 20% when this corresponded to 1 chemical shift. Test calculations were performed by randomly omitting different percentages of NOESY peaks and interface chemical shifts until severely incorrect structures, i.e. either two separated monomers or dimers with an interface radically different from the reference structure, were obtained.¹⁸ Each calculation was repeated three times with different random number generator seed values. The test calculations were performed first without omission of interface chemical shifts or NOESY peaks. The average structure for each of the three proteins obtained without chemical shift or NOESY peak omission was shown to be very similar to the corresponding original NMR and X-ray PDB structures¹⁸ and used as the reference structure for structural comparisons with the average structures calculated with chemical shift and peak omission. The RMSD values of structural comparisons were always calculated for the backbone atoms N, C^α, and C^γ in the ordered regions starting with the first and ending with the last secondary structure element in the sequence, i.e. residues 16–50 for Nika, 18–39 for WW, and 4–106 for REI.

In addition, test calculations for missing chemical shifts from non-interface residues were also performed by omitting 40% NOESY peaks combined with random omission of assigned chemical shifts of the non-interface residues in step of 10%, assuming that all ¹H chemical shifts at the interface are assigned.

The percentages of homodimeric structures in the PDB (see Introduction) reflect the state of May 5, 2014. The search was performed on all files obtained by a single experimental method

(X-ray, solution NMR, or solid-state NMR) and containing protein but neither DNA nor RNA. The respective number of homodimers was taken to be equal to the size of the subset of structures classified by the PDB as 'Homo 2-mer - A2'.

CONCLUSIONS

As reported previously,¹⁸ particular attention should be paid to ensembles of conformers with significant deviation from the reference structure but low RMSD within the ensemble and reasonable target function values, which let them appear as correct structures. Our results reveal that NOESY peak omission does not cause severe problems for the three test proteins containing different percentages of interface residues, provided that the interface chemical shifts are completely assigned. When NOESY peak omission is combined with interface chemical shift omission, the tolerance for interface chemical shift omission in the three proteins decreases further with a tendency approximately inversely proportional to the percentage of omitted NOESY peaks. Overall, the maximal allowed NOESY peak omission for the three test proteins is 40–50% with complete interface chemical shift assignments. The influence of NOESY peak omission on the tolerance for missing interface chemical shifts is mainly below 10–30% peak omission, above which calculations gradually tend to not tolerate any missing interface chemical shifts. The impact of unassigned interface side chain, methyl or aromatic chemical shifts on the susceptibility to NOESY peak omission is more pronounced than that of the same amount of unassigned general interface ¹H (including also backbone and side chain protons). Complete aromatic and methyl ¹H chemical shift assignments are thus important when NOESY peak lists are incomplete. In combination with NOESY peak omission, for example at 30% omission, calculations can tolerate only very few missing aromatic or methyl ¹H chemical shifts. In addition, for homodimeric proteins with a low percentage of interface residues, such as REI, completeness of both interface ¹H chemical shift assignment and NOESY peaks are important for a successful structure determination, since the results revealed that calculations fail to tolerate any combined omission of both interface ¹H chemical shifts and NOESY peaks together in REI.

ACKNOWLEDGEMENTS

We gratefully acknowledge financial support by the National Science Council of Taiwan, the Lichtenberg pro-

gram of the Volkswagen Foundation, and the Japanese Society for the Promotion of Science (JSPS) through Grants-in-Aid for Scientific Research C (25440032), Young Scientists B (24770101), and Scientific Research on Innovative Areas (24113720, 26102538 and 25120003).

REFERENCES

1. Williamson, M. *How Proteins Work*; Garland: New York, 2012; p 99.
2. Guerry, P.; Herrmann, T. *Q. Rev. Biophys.* **2011**, *44*, 257.
3. Güntert, P. *Eur. Biophys. J.* **2009**, *38*, 129.
4. Rosato, A.; Aramini, J.; Arrowsmith, C.; Bagaria, A.; Baker, D.; Cavalli, A.; Doreleijers, J. F.; Eletsky, A.; Giachetti, A.; Guerry, P. et al. *Structure* **2012**, *8*, 227.
5. Nilges, M. *Proteins* **1993**, *17*, 297.
6. O'Donoghue, S. I.; Nilges, M. In *Structure Computation and Dynamics in Protein NMR*; Krishna, N. R.; Berliner, L. J., Eds.; Kluwer Academic/Plenum Publishers: New York, 1999; Vol. 17, p 131.
7. Bardiaux, B.; Bernard, A.; Rieping, W.; Habeck, M.; Malliavin, T. E.; Nilges, M. *Proteins* **2009**, *75*, 569.
8. Mumenthaler, C.; Güntert, P.; Braun, W.; Wüthrich, K. *J. Biomol. NMR* **1997**, *10*, 351.
9. Mumenthaler, C.; Braun, W. *J. Mol. Biol.* **1995**, *254*, 465.
10. Nilges, M.; Macias, M. J.; O'Donoghue, S. I.; Oschkinat, H. *J. Mol. Biol.* **1997**, *269*, 408.
11. Linge, J. P.; O'Donoghue, S. I.; Nilges, M. *Methods Enzymol.* **2001**, *339*, 71.
12. Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209.
13. Kuszewski, J.; Schwieters, C. D.; Garrett, D. S.; Byrd, R. A.; Tjandra, N.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 6258.
14. Huang, Y. J.; Tejero, R.; Powers, R.; Montelione, G. T. *Proteins* **2006**, *62*, 587.
15. Greenfield, N. J.; Huang, Y. J.; Palm, T.; Swapna, G. V. T.; Monleon, D.; Montelione, G. T.; Hitchcock-DeGregori, S. E. *J. Mol. Biol.* **2001**, *312*, 833.
16. Gronwald, W.; Moussa, S.; Elsner, R.; Jung, A.; Ganslmeier, B.; Trenner, J.; Kremer, W.; Neidig, K. P.; Kalbitzer, H. R. *J. Biomol. NMR* **2002**, *23*, 271.
17. Savarin, P.; Zinn-Justin, S.; Gilquin, B. *J. Biomol. NMR* **2001**, *19*, 49.
18. Lin, Y.-J.; Kirchner, D. K.; Güntert, P. *J. Magn. Reson.* **2012**, *222*, 96.
19. Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283.
20. Yoshida, H.; Furuya, N.; Lin, Y. J.; Güntert, P.; Komano, T.; Kainosho, M. *J. Mol. Biol.* **2008**, *384*, 690.
21. Furuya, N.; Komano, T. *J. Bacteriol.* **1995**, *177*, 46.
22. Ohnishi, S.; Güntert, P.; Koshihara, S.; Tomizawa, T.; Akasaka, R.; Tochio, N.; Sato, M.; Inoue, M.; Harada, T.; Watanabe, S. et al. *FEBS Lett.* **2007**, *581*, 462.
23. Epp, O.; Lattman, E. E.; Schiffer, M.; Huber, R.; Palm, W. *Biochemistry* **1975**, *14*, 4943.