Journal of Magnetic Resonance 222 (2012) 96–104

Contents lists available at SciVerse ScienceDirect

Journal of Magnetic Resonance

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

Influence of ^{1}H chemical shift assignments of the interface residues on structure determinations of homodimeric proteins

Yi-Jan Lin [a,*], Donata K. Kirchner [b,c], Peter Güntert [b,c,d,*]

[a] Graduate Institute of Natural Products and Center of Excellence for Environmental Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan
[b] Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
[c] Frankfurt Institute for Advanced Studies, Goethe University Frankfurt am Main, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany
[d] Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

ARTICLE INFO

Article history:
Received 22 March 2012
Revised 1 July 2012
Available online 14 July 2012

Keywords:
Homodimeric proteins
Resonance assignment
Structure determination
CYANA

ABSTRACT

Homodimeric proteins pose a difficulty for NMR structure determination because the degeneracy of the chemical shifts in the two identical monomers implies an ambiguity in all assignments of distance restraints. For homodimeric proteins, residues involved in the interface between two monomers provide essential intermolecular NOEs. The structure determination of homodimeric proteins hence relies strongly on chemical shift assignments of these interface residues. Our paper discusses the influence of the extent of ^{1}H chemical shift assignments of interface residues on the structure determinations of homodimeric proteins using the CYANA program. The results reveal that successful structure determinations of homodimeric proteins with automated NOE assignment depend on the percentage of assigned interface residues and that a high completeness of around 80–90% of the ^{1}H chemical shift assignment in the interface is needed for reliable NMR structure determinations of homodimeric proteins for which no experimental distinction between intra- and intermolecular NOEs, e.g. by filtered NOESY experiments, is available. Our results also show that RMSD and target function values are insufficient to judge the quality of homodimeric structures determined using automated NOE assignment. Structure determinations of homodimeric proteins by NMR using conventional NOESY experiments are thus possible but more challenging than for monomeric proteins.

1. Introduction

Oligomeric proteins are ubiquitous in nature. Approximately two-thirds of the human enzymes are oligomers and in *Escherichia coli* the average oligomerization state of proteins is four [1]. However, the number of homodimeric protein structures determined by NMR remains small compared to that of monomeric proteins, for which automated NOE assignment and structure calculation methods have become reliable and routine [2–4]. Nevertheless, the NMR technique for protein structure determination in solution has been increasingly applied to investigating symmetric oligomers, especially symmetric dimers [5–7]. Table 1 shows that 406 structures of symmetric multimers have been determined by NMR in solution, comprising 4.2% of all 8105 solution NMR protein structures currently in the PDB. Of these 406 entries 344 are for symmetric dimers. A different picture emerges for solid state NMR. Out of the small number of 53 solid state NMR protein structures in the PDB, 22 are for symmetric multimers, of which only 4 are homodimers, while the others consist of 3–18 identical monomers.

Homodimeric proteins pose a difficulty for NMR structure determination. The corresponding nuclei in both monomers have equivalent magnetic environments and therefore their chemical shifts are degenerate. Only one set of signals, as from one monomer, is observed in the spectra. Due to the degeneracy of chemical shifts in the symmetric dimers, spectral overlap is reduced and only chemical shifts of one monomer have to be assigned, i.e. about half the number than for a monomeric protein of the same size. However, at the same time the NOE assignment and structure calculation become more complicated because a priori every NOE has an intramolecular and an intermolecular assignment. Although several fully automated approaches for combined automatic NOESY assignment and structural calculation have been developed [8–17], a limiting factor for the application of these automated NOE assignment procedures to symmetric dimers is the difficulty to distinguish inter-monomeric from intra-monomeric NOEs (and from NOEs with simultaneous intra- and inter-monomeric

* Corresponding authors. Addresses: Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany. Fax: +49 69 79829643 (P. Güntert). Graduate Institute of Natural Products and Center of Excellence for Environmental Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan (Y.-J. Lin).

E-mail addresses: yjlin@kmu.edu.tw (Y.-J. Lin), guentert@em.uni-frankfurt.de (P. Güntert).

1090-7807/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmr.2012.07.001

Table 1
NMR structures of symmetric multimers in the Protein Data Bank[a].

Monomers	Solution NMR structures	Solid state NMR structures
2	344[b]	4 (1NAG, 1ZY6, 2KLR, 2RLZ)
3	18 (1AQ5, 1IIE, 1J7H, 1M7L, 1QCE, 1RFO, 1WCR, 2EZO, 2EZP, 2EZQ, 2EZR, 2EZS, 2FXP, 2KA3, 2KDC, 2KO8, 2KP8, 2OII)	1 (2KJ3)
4	36 (1F6G, 1JQ1, 1JQ2, 1KQE, 1MPE, 1OLG, 1OLH, 1P23, 1PES, 1PET, 1PFM, 1PFN, 1QEY, 1SAE, 1SAF, 1SAK, 1SAL, 1TXP, 2J0Z, 2J10, 2J11, 2JO4, 2JO5, 2K1E, 2KB1, 2KBY, 2KIH, 2KIX, 2KJ1, 2KWX, 2L9H, 2LJB, 2LJC, 2RLF, 2RP4, 3SAK)	7 (1NYJ, 2E8D, 2H95, 2KAD, 2KQT, 2L0J, 2NNT)
5	6 (1ZLL, 2BEG, 2HYN, 2JZ7, 2KYV[c], 2RNM)	4 (1EQ8, 2KWD, 2KYV[c], 2LBU)
7	2 (2KU1, 2KU2)	–
8	–	2 (2KIB, 2LNQ)
12	–	2 (2LMN, 2LMO)
18	–	2 (2LMP, 2LMQ)
⩾2	406	22
Total	8105	53

[a] As of February 14, 2012. The rows for 2–18 monomers list all PDB entries determined by solution or solid state NMR that contain symmetric multimer protein structures with the given number of monomers of identical sequence, excluding entries containing any DNA or RNA. The row for ⩾2 monomers gives the sum over the preceding rows, i.e. the total number of NMR structures of symmetric multimers in the PDB. There are no NMR structures of symmetric multimers with more than 18 identical monomers in the PDB. The last row gives the total numbers of corresponding PDB entries regardless of the multimeric state, i.e. including also monomeric proteins and asymmetric protein complexes (without DNA or RNA).

[b] PDB entries 1A03, 1A1U, 1AFO, 1AJY, 1ARQ, 1ARR, 1AUU, 1B28, 1B4C, 1B50, 1B53, 1BFM, 1BUQ, 1BVE, 1BVG, 1CFP, 1CNP, 1COP, 1CTA, 1CTD, 1DBD, 1DHM, 1DIP, 1DOM, 1DON, 1DUM, 1DZ1, 1E0E, 1E52, 1EJP, 1EJQ, 1EV0, 1EXE, 1F3C, 1GJZ, 1GRM, 1HBW, 1HF9, 1HI7, 1HRJ, 1HS5, 1HUE, 1HUM, 1HUN, 1HZE, 1I18, 1I4V, 1IHQ, 1IHV, 1IHW, 1IL8, 1ISK, 1J4V, 1J9I, 1JNO, 1JO3, 1JO4, 1JOY, 1JR5, 1JUN, 1JWD, 1JY4, 1JY6, 1K2H, 1KLA, 1KLC, 1KLD, 1L3N, 1L5E, 1L6E, 1LR1, 1M31, 1MGS, 1MH6, 1MI2, 1MIC, 1MNT, 1MSG, 1MSH, 1MV4, 1N3J, 1N9J, 1NCV, 1NEI, 1NG7, 1NG8, 1NI8, 1NIQ, 1NLA, 1NRM, 1NRU, 1NS1, 1NSH, 1NT5, 1NT6, 1NTC, 1NVO, 1OVX, 1OZO, 1P94, 1PBZ, 1PE3, 1PFS, 1PZQ, 1PZR, 1Q10, 1Q53, 1Q6A, 1Q6B, 1QCK, 1QLK, 1QMC, 1QNK, 1QP6, 1QTG, 1QXN, 1R05, 1R2A, 1R48, 1R6R, 1R8P, 1RE6, 1RJJ, 1ROD, 1RPR, 1RQT, 1RQU, 1RQV, 1RTN, 1RTO, 1S1O, 1S4A, 1SYM, 1TKV, 1TMZ, 1TRL, 1U7J, 1U7M, 1UTR, 1UWO, 1VL3, 1WJA, 1WJB, 1WJC, 1WJD, 1WJE, 1WJF, 1WRS, 1WRT, 1WTU, 1X93, 1X9A, 1X9V, 1X9X, 1XDX, 1XOX, 1XSX, 1XYD, 1Y00, 1Y4O, 1Y7Q, 1Y9X, 1YFB, 1YSF, 1YUR, 1YUS, 1YUT, 1YUU, 1Z09, 1Z0R, 1ZAE, 1ZFS, 1ZXA, 1ZZF, 2A2Y, 2ADL, 2ADN, 2AF2, 2AN7, 2AQ0, 2ASY, 2B95, 2B9Z, 2BA3, 2BGF, 2BZB, 2CNP, 2CXJ, 2D9S, 2DO6, 2DSM, 2DWV, 2E8J, 2EZW, 2EZX, 2EZY, 2EZZ, 2F8B, 2FI2, 2FY9, 2GD7, 2GJF, 2GJH, 2GVA, 2GVB, 2GW6, 2GZU, 2HAC, 2HDP, 2HQO, 2HQR, 2HV1, 2I7U, 2IL8, 2J5D, 2J8J, 2J8L, 2JNA, 2JNJ, 2JO8, 2JP1, 2JPQ, 2JPT, 2JQF, 2JR2, 2JRA, 2JRL, 2JRX, 2JS1, 2JS3, 2JS5, 2JSC, 2JST, 2JUG, 2JUW, 2JUZ, 2JV7, 2JW8, 2JWA, 2JWE, 2JWK, 2JWL, 2JWN, 2JXG, 2JXH, 2JZ0, 2JZ1, 2K01, 2K1K, 2K1L, 2K1O, 2K29, 2K5J, 2K6L, 2K6S, 2K7I, 2K7O, 2K7V, 2K8S, 2K8X, 2K9I, 2K9Q, 2K9Y, 2KA1, 2KA2, 2KAX, 2KAY, 2KDD, 2KEL, 2KHM, 2KHZ, 2KIK, 2KJB, 2KJC, 2KJZ, 2KKE, 2KKO, 2KLH, 2KLK, 2KM2, 2KMT, 2KNV, 2KO1, 2KOD, 2KOT, 2KPE, 2KPF, 2KQM, 2KQN, 2KRF, 2KS0, 2KUJ, 2KVJ, 2KW6, 2KX9, 2KXO, 2KYI, 2L01, 2L02, 2L0P, 2L2T, 2L34, 2L3A, 2L48, 2L49, 2L50, 2L51, 2L5H, 2L66, 2L7H, 2L7I, 2L8T, 2L8X, 2L9U, 2LE2, 2LFH, 2LFR, 2LFS, 2LH0, 2LHK, 2LI9, 2LK5, 2LK6, 2LNZ, 2LO0, 2NBT, 2NLU, 2NWT, 2OSG, 2PE9, 2PEA, 2PRU, 2RMM, 2RO3, 2RO4, 2RO5, 2RP5, 2W1O, 2WC2, 2XA6, 2XDI, 3MSP, 3ZTG, 4AAI.

[c] PDB entry 2KYV was determined by a combination of solution and solid state NMR.

contributions). In homodimeric proteins, residues located in the interface between two monomers have inter-monomeric interactions and hence provide inter-monomeric NOEs that play an important role for the NMR structure determination of homodimeric proteins. The NOE assignments strongly rely on the chemical shift assignments of the ^{1}H hydrogens and their attached ^{15}N and ^{13}C heavy atoms. In order to discuss the importance of the chemical shift assignment of these interface residues in homodimeric proteins, we investigated the influence of missing ^{1}H chemical shift assignments of interface residues on the NMR structure determination of homodimeric proteins using automated NOE assignment and structure calculation with the program CYANA [3,12].

2. Materials and methods

Structure calculations of homodimeric proteins were carried out with the program CYANA for combined automated NOESY assignment using a probabilistic algorithm [3] and structure calculations with torsion angle dynamics [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 [5,19]. Dihedral angle difference restraints were implemented by adding a term,

$$T_{ide} = k_{ide}\sum_i (\theta_i^{(1)} - \theta_i^{(2)})^2,$$

to the CYANA target function where the sum runs over all dihedral angles, $\theta_i^{(1)}$ and $\theta_i^{(2)}$ are the values of the dihedral angle i in the two monomers, and k_{ide} is a force constant that was set to 0.03 Å^2 for all calculations in this paper. Minimizing T_{ide} favors an identical conformation of the two monomers without restricting their relative position and orientation. Differences between symmetry-related distances were restrained by a target function term

$$T_{sym} = k_{sym}\sum_{(\alpha,\beta)} (d(\alpha^{(1)},\beta^{(2)}) - d(\alpha^{(2)},\beta^{(1)}))^2,$$

where the sum runs over all restrained atom pairs (α,β), $\alpha^{(1)}$, $\beta^{(1)}$, $\alpha^{(2)}$, and $\beta^{(2)}$ denote atoms α and β in monomers 1 and 2, respectively, and k_{sym} is a force constant. For all calculations in this paper the distances between all pairs of C^α atoms were restrained, and k_{sym} = 0.0025. Minimizing T_{sym} favors a symmetric relative orientation of the two monomers without imposing a specific symmetry.

For the automated assignment of NOEs, the symmetry is taken into account by restricting possible assignments to one (the "principal distance") out of every group of symmetry-related equivalent distances. In the case of a homodimer, these are, for instance, the intramolecular contacts within the first monomer, and intermolecular contacts from a first atom in the first monomer to a second atom in the second monomer. This reduces the initial ambiguity of an NOE cross peak assignment in a homodimer from 4 to 2 assignment possibilities for every atom pair whose chemical shifts match the peak position, and eliminates artificial degeneracies that would be present if the two monomers are treated as independent unrelated monomers in which all chemical shifts are "accidentally" degenerate with those of the other monomer, which would dilute the information content of the NOEs during network-anchoring. After the NOE assignment has been made, the restraints for equivalent symmetry-related distances are generated from those for the principal distances. In a homodimer, each intramolecular restraint within monomer 1 is complemented by an identical restraint within monomer 2, and each intermolecular restraint from monomer 1 to monomer 2 is complemented by a corresponding restraint from monomer 2 to monomer 1.

It is straightforward to generalize this treatment of symmetric homodimers to higher order multimers of known symmetry type by adding corresponding terms to T_{ide} and T_{sym} and applying the concept of one principal distance for every group of symmetry-related equivalent distances, which may then comprise more than two distances.

The iterative procedure of NOE assignments and structure calculations for homodimeric proteins consists of seven cycles. 100 conformers were calculated in each cycle and the 20 structures with the lowest target function values were used as input structures for the next cycle. The structure calculation starts by generating structures with random torsion angle values and uses the standard simulated annealing schedule [18]. Backbone hydrogen bond restraints in helices were used as additional input, since helices can be identified readily from the NMR data, e.g. by the chemical shift index [20]. No hydrogen bond restraints between β-strands were used. In order to investigate the influence of α-helical hydrogen bond restraints on the homodimer structure determination, we also performed calculations without input of hydrogen bond restraints.

The experimental NMR data of the DNA binding protein NikA (PDB code: 2BA3) [19,21], 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 containing 51 amino acid residues with one β-strand and two α-helices in each monomer. Two monomers form a dimer consisting of 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 those with protons within 6 Å of protons of the other monomer. The dimer interface residues thus included 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 included chemical shift assignments and NOESY peak lists. For NikA 89.5% of the chemical shifts of the backbone and non-labile side-chain ^{1}H were assigned (BMRB code 15784). In total, there were 228 assigned ^{1}H chemical shifts, 163 for amide, H^{α} and H^{β} protons, and 65 for side-chain protons (excluding H^{β}). We counted H^{β} as a backbone resonance because generally H^{β} chemical shifts can be obtained from spectra for the backbone resonance assignment. The side-chain chemical shifts included 5 aromatic ^{1}H and 31 methyl ^{1}H chemical shifts. 122 (54%) ^{1}H chemical shifts, including 79 (35%) H^{N}, H^{α} and H^{β}, and 43 (19%) side-chain proton chemical shifts comprising 5 aromatic ^{1}H and 19 methyl ^{1}H, belong to dimer interface residues. A total of 1839 NOE peaks containing peak positions and volumes from three three-dimensional NOESY spectra, one ^{15}N-, one aliphatic ^{13}C- and one aromatic ^{13}C-resolved NOESY spectrum, were used as input for the calculations. Hydrogen bonds in the α-helical regions, whose locations were determined using the chemical shift index, were constrained by a set of upper and lower distance restraints.

For the WW domain 90.9% of the chemical shifts of the backbone and non-labile side-chain ^{1}H were assigned (BMRB code 10028). In total, there were 227 assigned ^{1}H chemical shifts, 147 for amide, H^{α} and H^{β} protons, and 80 for side-chain protons (excluding H^{β}) including 15 aromatic ^{1}H and 11 methyl ^{1}H chemical shifts. 113 (49.8%) ^{1}H chemical shifts, including 69 (30.4%) amide, H^{α} and H^{β} proton chemical shifts, and 44 (19.4%) side-chain proton chemical shifts, comprising 15 aromatic and 5 methyl protons, belong to dimer interface residues. A total of 1583 NOE peaks containing peak positions and volumes from two three-dimensional NOESY spectra, one ^{15}N- and one ^{13}C-resolved NOESY spectrum, were used as input for the calculations. Hydrogen bond restraints were not used since there are no α-helices in this protein.

For REI, the ^{1}H, ^{13}C, and ^{15}N chemical shifts were simulated by setting their values to the average value from the BMRB statistics [24] plus a normally distributed random number with a standard deviation of 0.4 ppm for ^{1}H or 3.0 ppm for ^{13}C and ^{15}N. H^{γ} of CYS 23 and CYS 88 were excluded. There was a total of 559 simulated ^{1}H chemical shifts, comprising 371 amide, H^{α} and H^{β} protons and 188 side-chain protons (excluding H^{β}). The side-chain chemical shifts included 31 aromatic ^{1}H and 71 methyl ^{1}H chemical shifts. Interface residues accounted for 151 (27%) proton chemical shifts, containing 100 (18%) amide, H^{α} and H^{β} protons and 51 (9%) side-chain protons, including 13 aromatic ^{1}H and 15 methyl ^{1}H chemical shifts. The ^{15}N- and ^{13}C-resolved 3D NOESY peak lists containing 7724 peaks were simulated using the X-ray structure assuming that NOEs can be seen up to 6 Å. The NOE assignments for residues outside the interface between two monomers were treated as intra-monomeric NOEs, whereas the NOEs from residues located in the interface between two monomers were treated as both inter- and intra-monomeric NOEs.

Test calculations were performed by omitting different percentages of the assigned chemical shifts of interface residues:

1. Random omission of ^{1}H proton chemical shifts: given percentages (in steps of 5% until severely incorrect structures were obtained) of the ^{1}H chemical shifts were randomly selected and removed from the chemical shift list.
2. Random omission of aromatic or methyl ^{1}H chemical shifts: aromatic or methyl ^{1}H chemical shifts were randomly selected and removed (in steps of 10%, except for the aromatic shifts in NikA and methyl shifts in WW that were randomly selected and removed in steps of 20%, because every 20% corresponds to 1 ^{1}H chemical shift).
3. Random omission of side-chain ^{1}H chemical shifts: side-chain ^{1}H chemical shifts (H^{β} not included) were randomly selected and removed (in steps of 5%) from the chemical shift lists.

In addition, to take into account that not all theoretically expected NOEs can be detected in experimental spectra we also prepared by random selection a reduced data set for REI comprising only 60% of the complete NOESY peak lists. Structure calculations for REI were also performed using 60% of the NOEs by omitting ^{1}H, side-chain ^{1}H, aromatic and methyl ^{1}H chemical shifts as described above.

Each calculation was repeated three times with different random number generator seed values, and the averaged RMSD and target function values were used for comparison and discussion. The test calculations were performed first without omission of chemical shifts. The average structure obtained without chemical shift omission was used as the reference structure for structural comparisons with the average structures calculated with chemical shift omission. The RMSD values of structural comparisons were always calculated for the backbone atoms N, C^{α}, and C′ 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. For drawing structure figures these ranges were extended to include additional residues that are either involved in intermolecular NOEs or have NOEs to the core structure, i.e. residues 16–51 for NikA, 14–43 for WW, and 1–107 for REI.

3. Results

3.1. Test calculations without omission of the chemical shifts from interface residues

The RMSD values between the mean structures from test calculations without chemical shift omission and from the PDB were 1.15 Å for NikA, 0.43 Å for WW and 0.92 Å for REI (Figs. 1a, 2a, and 3a), indicating that the structures of NikA, WW and REI (Figs. 1b, 2b, and 3b) determined without chemical shift omission

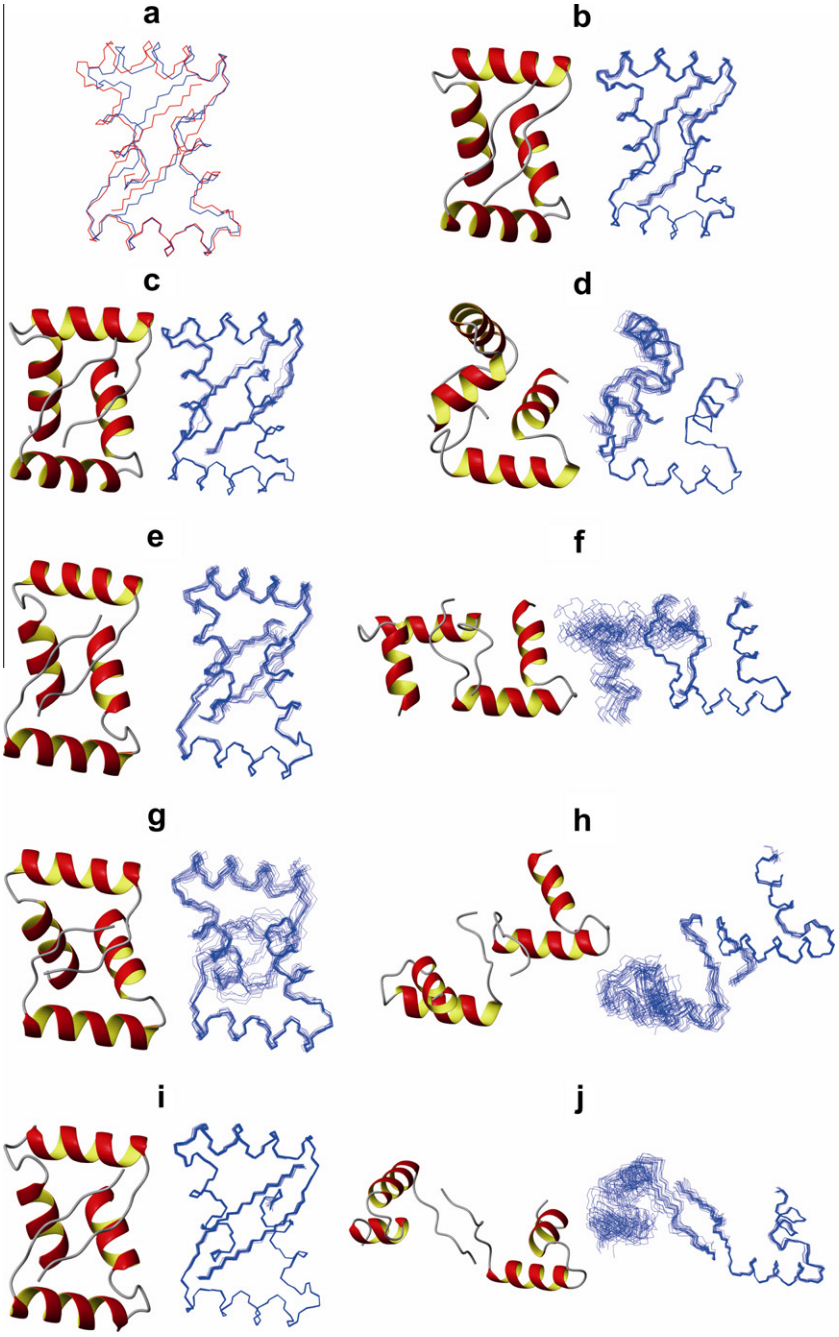

Fig. 1. Results of structure calculations of the homodimeric protein NikA at different levels of completeness of the chemical shift assignment. Only the structured region of residues 16–51 is shown. (a) Structural comparison between the NMR structure deposited in the PDB (red) and the mean structure determined without chemical shift omission from interface residues (blue). (b) Ribbon and structure ensembles determined without chemical shift omission from interface residues. (c) and (d) are ribbon and structure ensembles determined with ^{1}H chemical shift omission from interface residues at maximal tolerated omission (20%) and over-tolerated omission (25%), respectively. (e) and (f) are ribbon and structure ensembles determined with side-chain ^{1}H chemical shift omission from interface residues at maximal tolerated omission (45%) and over-tolerated omission (50%), respectively. g and h are ribbon and structure ensembles determined with methyl ^{1}H chemical shift omission from interface residues at maximal tolerated omission (90%) and over-tolerated omission (100%), respectively. (i) and (j) are ribbon and structure ensembles determined with aromatic ^{1}H chemical shift omission from interface residues at maximal tolerated omission (80%) and over-tolerated omission (100%), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from interface residues are very similar to the corresponding original NMR and X-ray PDB structures.

3.2. Random omission of ^{1}H chemical shifts of interface residues

For NikA, with up to 20% interface chemical shift omission, the structure bundles (Fig. 1c) showed low RMSD (high precision) and the RMSD values between the mean structures and the mean reference structure (accuracy) increased only slowly (Fig. 4a). Deviations due to chemical shift omission occurred mainly in the β-strands. The conformations of the β-strands became gradually disordered with increasing percentage of omitted chemical shifts. In contrast, the conformation of the α-helices, in which chemical shifts were also omitted, remained virtually unchanged, presumably because of the α-helical hydrogen bond restraints. At 25% interface chemical shift omission, the deviations from the reference

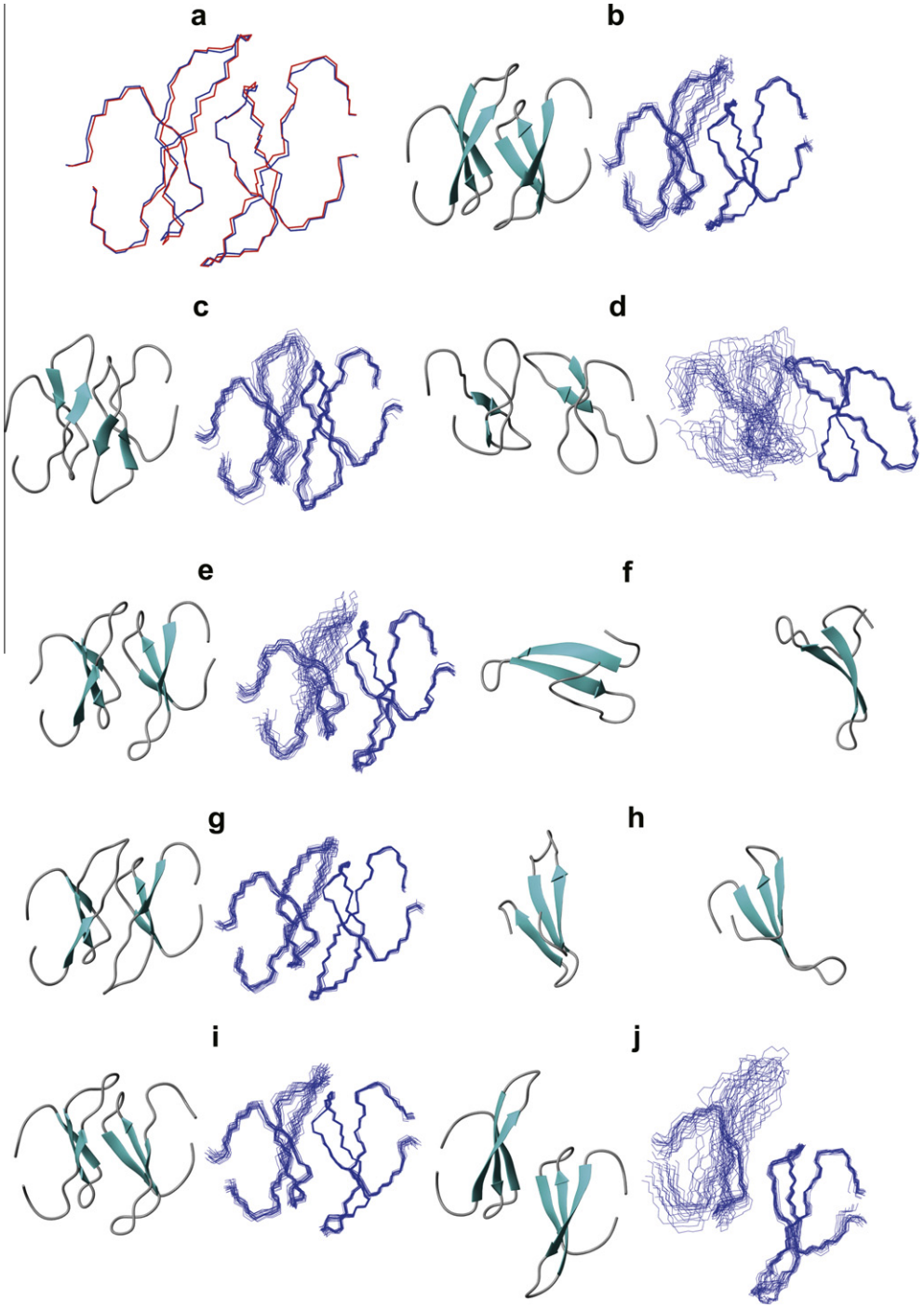

Fig. 2. Results of structure calculations of the homodimeric protein WW at different levels of completeness of the chemical shift assignment. Only the structured region of residues 14–43 is shown. (a) Structural comparison between the NMR structure deposited in the PDB (red) and the mean structure determined without chemical shift omission from interface residues (blue). (b) Ribbon and structure ensembles determined without chemical shift omission from interface residues. (c) and (d) are ribbon and structure ensembles determined with ^{1}H chemical shift omission from interface residues at maximal tolerated omission (25%) and over-tolerated omission (30%), respectively. (e) and (f) are ribbon and structure ensembles determined with side-chain ^{1}H chemical shift omission from interface residues at maximal tolerated omission (50%) and over-tolerated omission (55%), respectively. (g) and (h) are ribbon and structure ensembles determined with methyl ^{1}H chemical shift omission from interface residues at maximal tolerated omission (60%) and over-tolerated omission (80%), respectively. No bundles are shown in (f) and (h) where the monomers do not contact each other. (i) and (j) are ribbon and structure ensembles determined with aromatic ^{1}H chemical shift omission from interface residues at maximal tolerated omission (80%) and over-tolerated omission (90%), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure increased significantly (Fig. 4a), indicating that severely disordered homodimeric structures or incorrect structures (Fig. 1d) had been obtained, although the ensembles still showed low RMSDs and reasonable target function values (Fig. 4a). For WW, 25% omission (Fig. 2c) led to an RMSD deviation from the reference structure of 2.26 Å (Fig. 4b), whereas 30% omission (Fig. 2d) resulted in distorted structures with RMSD to the reference structure of 3.22 Å. Hence the maximally tolerated shift omission for WW is 25%. For REI, until 10% omission, the ensembles (Fig. 3c) showed low RMSD and the deviations from the reference structure increased slowly (Fig. 4c). At 15% omission, the RMSD to the reference structure increased significantly (Fig. 4c) and the structures

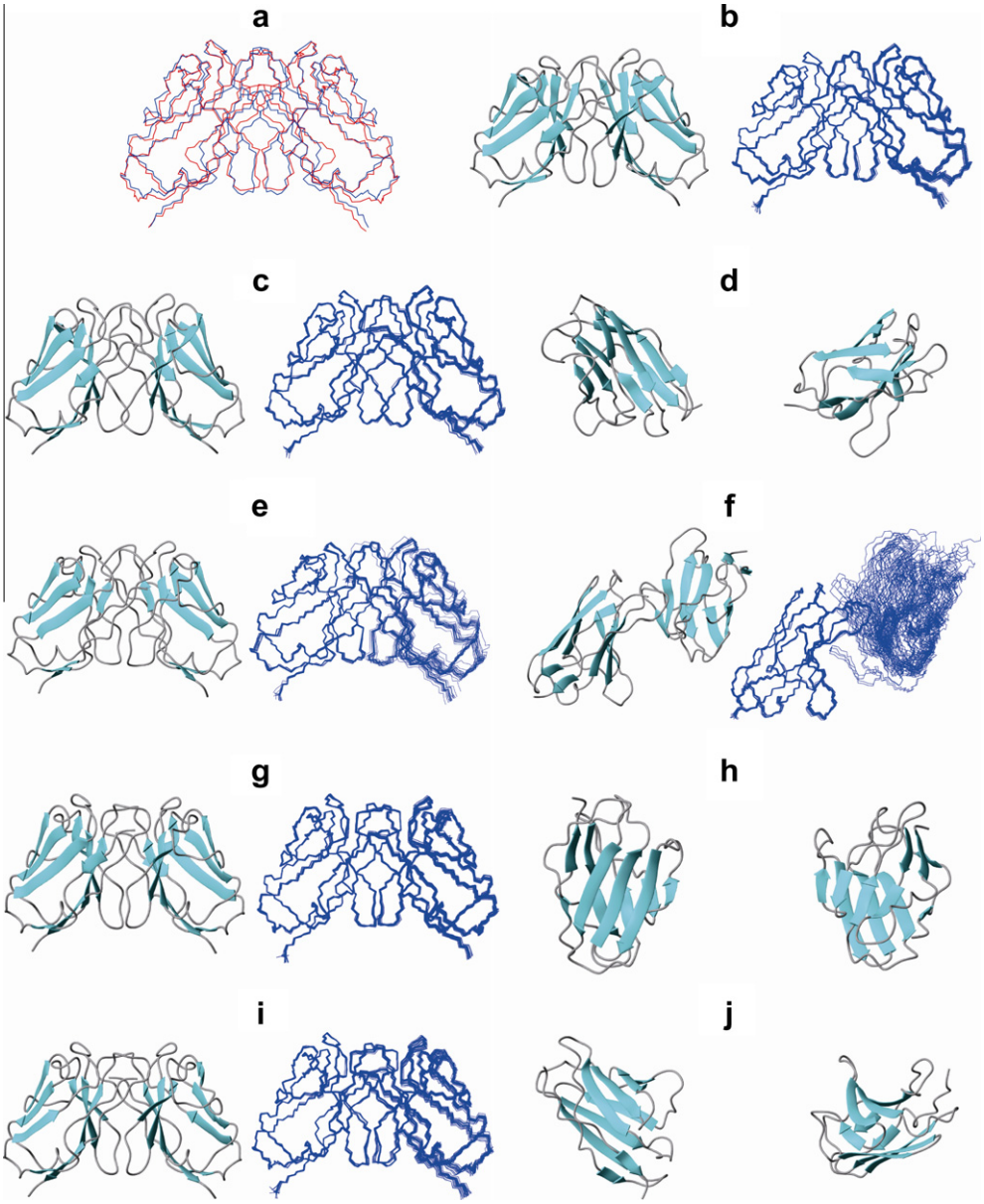

Fig. 3. Results of structure calculations of the homodimeric protein REI at different levels of completeness of the chemical shift assignment. (a) Structural comparison between X-ray PDB structure (red) and mean structure determined without chemical shift omission from interface residues (blue). (b) Ribbon and structure ensembles determined without chemical shift omission from interface residues. (c) and (d) are ribbon and structure ensembles determined with ^{1}H chemical shift omission from interface residues at maximal tolerated omission (10%) and over-tolerated omission (15%), respectively. (e) and (f) are ribbon and structure ensembles determined with side-chain ^{1}H chemical shift omission from interface residues at maximal tolerated omission (40%) and over-tolerated omission (45%), respectively. (g) and (h) are ribbon and structure ensembles determined with methyl ^{1}H chemical shift omission from interface residues at maximal tolerated omission (50%) and over-tolerated omission (60%), respectively. (i) and (j) are ribbon and structure ensembles determined with aromatic ^{1}H chemical shift omission from interface residues at maximal tolerated omission (60%) and over-tolerated omission (70%), respectively. For clarity, in cases where the structure calculations yielded separated monomers (i.e., panels d, h, and j) the structure ensemble is not shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

showed an erroneous ensemble with two separated monomers (Fig. 3d).

3.3. Random omission of side-chain ^{1}H chemical shifts of interface residues

Generally, the backbone and H^{β} resonances can be assigned completely using spectra for backbone resonance assignment. Therefore we also investigated the effect of unassigned side-chain resonances beyond H^{β} on homodimer structure calculations. For NikA, until 45% omission of interface side-chain ^{1}H chemical shifts (H^{β} not included), the deviations from the reference structure did not increase significantly (Fig. 4d), although β-strands became disordered with increasing percentage of unassigned interface chemical shifts, whereas the four-helix bundle, which is also involved in interface interactions, did not change significantly. The ensembles obtained with 5–45% omission (Fig. 1e) showed low RMSD values to the reference structure (Fig. 4d). At 50% omission, the ensembles (Fig. 1f) from two of the three calculations deviated strongly from the reference structure, although the ensembles showed low RMSD values to their mean coordinates (0.81 Å) and target function values (1.08 $Å^2$) (Fig. 4d). For WW and REI, until 50% (Fig. 2e) and 40% omission (Fig. 3e), the RMSD to the reference structures varied slowly with low ensemble RMSD values, whereas with 55% and 45% omission of interface side-chain chemical shifts for WW and REI, respectively, the deviation from the reference

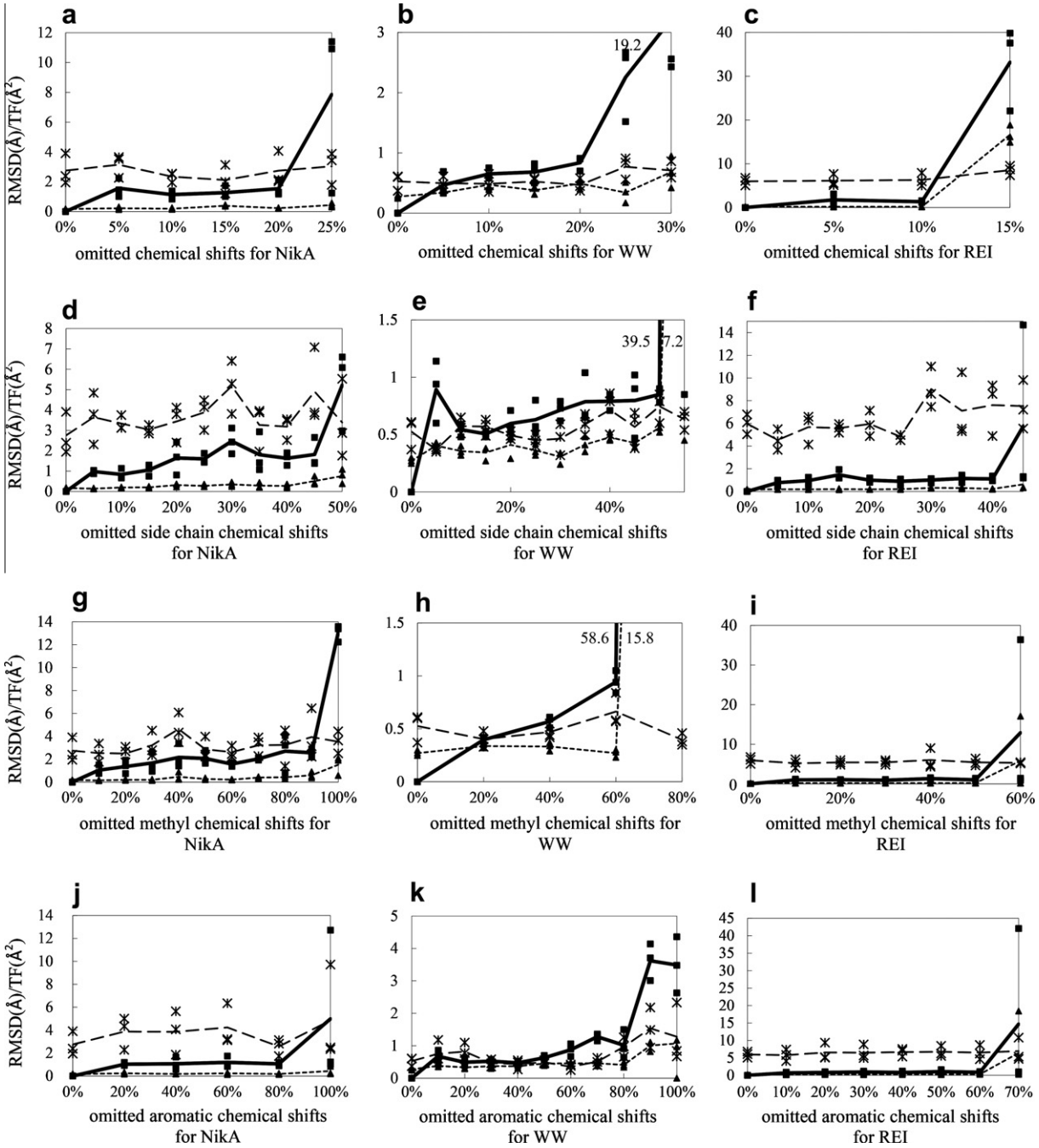

Fig. 4. Results of structure calculations of three homodimeric proteins, NikA, WW and REI, at different levels of completeness of the chemical shift assignment. Three independent runs were performed for each omitted percentage of chemical shifts. The horizontal axis indicates the percentage of randomly omitted chemical shifts. The vertical axis indicates the RMSD value of the ensemble relative to its mean coordinates in each run (triangles connected by a dotted line), the RMSD value between the mean structure from each calculation and the mean reference structure (squares connected by a solid line), and the average final target function value in each run (stars connected by a dashed line), respectively. The average values over the three runs are connected by a dotted, thick, and dashed line, respectively. The RMSD values were always calculated for the backbone atoms, N, C^{α} and C' of residues 16–50 in NikA, 18–39 in WW, and 4–106 in REI.

structure increased significantly (Fig. 4e and f) and erroneous structures were obtained (Figs. 2f and 3f).

3.4. Random omission of methyl ^{1}H chemical shifts of interface residues

For NikA, until 90% omission (Fig. 1g) of interface methyl ^{1}H chemical shifts, the RMSD to the reference structure remained on average below 2.7 Å (Fig. 4g). The fold of NikA remained correct in spite of disorder in the β-strands. The complete omission of all interface methyl ^{1}H chemical shifts yielded wrong ensembles (Fig. 1h) with low RMSD to their mean coordinates and reasonable target function values (Fig. 4g). For WW, 80% omission, corresponding to 4 omitted methyl ^{1}H chemical shifts, already caused significant deviations to the reference structure and wrong ensembles (Figs. 2h and 4h). For REI, up to 50% omission (Fig. 3g) of the interface methyl ^{1}H chemical shifts the target function and RMSD to the reference structure varied only slightly (Fig. 4i). With 60% omission, the deviations increased significantly (Fig. 4i) and incorrect structures were obtained (Fig. 3h).

3.5. Random omission of aromatic ^{1}H chemical shifts of interface residues

For NikA, up to 80% omission (Fig. 1i) of interface aromatic proton chemical shift assignments the ensemble RMSD and the RMSD to the reference structure remained almost unchanged (Fig. 4j). The omission of all interface aromatic chemical shifts yielded structures (Fig. 1j) with low ensemble RMSD to the mean coordinates and reasonable target function but with significantly increased RMSD to the reference structure in one of three calculations (Fig. 4j). For WW, 80% omission of aromatic proton chemical shifts can be tolerated (Figs. 2i and 4k). At 90% and 100% omission, structures with RMSD deviations to the reference structure of 4.14 Å and 4.36 Å were obtained, respectively (Figs. 2j and 4k). For REI, until 60% omission (Fig. 3i), the RMSD to the reference structure varied little (Fig. 4l). At more than 70% omission of the interface aromatic ^{1}H chemical shifts, incorrect structures were obtained (Fig. 3j) with both a large ensemble RMSD value and significantly increased RMSD to the reference structure in one of the three calculations (Fig. 4l).

3.6. Random omission of ^{1}H chemical shifts of interface residues without hydrogen bond restraints in α-helical regions

Repeating the structure calculations without hydrogen bond restraints for the α-helices revealed that the tolerance percentage for unassigned ^{1}H chemical shifts in NikA was essentially the same as with hydrogen bond restraints, except that α-helices became slightly disordered and that the RMSD to the reference structure rose by around 1.5 Å (data not shown).

3.7. Random omission of 40% NOE peaks for REI

Results obtained with the random omission of 40% of the NOESY peaks for REI showed that the tolerance for missing methyl ^{1}H chemical shifts stayed unchanged (50%), for missing aromatic ^{1}H chemical shifts it decreased from 60% to 50%, for missing side-chain ^{1}H chemical shifts decreased even more from 40% to 20%, and for missing ^{1}H chemical shifts it increased slightly from 10% to 15%. However, the latter increase is not significant, since a more detailed investigation showed that without NOE omission calculations with up to 14% omitted shifts were successful, whereas unsuccessful calculations with 60% NOEs were obtained with 16% omission. Overall, we conclude that omitting 40% of the NOESY cross peaks for REI has no striking influence on the percentage of missing chemical shifts that can be tolerated for successful dimer structure calculation.

4. Discussion

Omitting 25% of all assigned ^{1}H chemical shifts in the dimer interface of NikA, corresponding to 31 chemical shifts, or 15% of all assigned ^{1}H chemical shifts in the dimer interface of REI, corresponding to 23 chemical shifts, resulted in incorrect structures. The higher tolerance percentage for NikA than REI is probably due to the high percentage of interface residues relative to the entire protein, even though the simulated data in REI were more perfect, i.e. the NOE peaks were simulated to be perfectly picked in ^{15}N- and ^{13}C-resolved 3D NOESY spectra, and almost all chemical shifts of the backbone and non-labile side-chains ^{1}H were assigned, whereas in NikA 10.5% of the chemical shifts for the backbone and non-labile side-chains ^{1}H had not been assigned experimentally.

In NikA, 24 residues (47%) out of the total of 51 residues in one monomer are involved in inter-monomeric interactions. In the case of REI, 29 residues out of a total of 107 residues in one monomer, 27%, are involved in inter-monomeric interactions. This indicates that the percentage of unassigned interface chemical shifts that can be tolerated for the determination of a correct homodimeric structure depends on the percentage of residues involved in dimer interface interactions. In WW, 19 residues (38.8%) out of the total of 49 residues in a monomer are involved in inter-monomeric interactions. The percentage of the interface residues in WW is lower than the percentage of interface residues in NikA (47%). However, excluding the residues in the flexible chain termini, i.e. 1–15 in NikA, and 1–13 and 44–49 in WW, the percentage of interface residues becomes similar, i.e. 66.7% (24 out of 36 residues) in NikA and 63.3% (19 out of 30 residues) in WW, respectively. WW can tolerate a slightly higher percentage (25%; 28 chemical shifts) of unassigned interface ^{1}H chemical shifts than NikA (20%; 24 chemical shifts). The fact that for REI the tolerance for unassigned interface ^{1}H chemical shifts is only 10% indicates that the structure determination of homodimeric proteins with automated NOESY assignment still remains difficult if the interface region is small compared to the rest of the protein and if no dimer-specific information (e.g. from NOESY experiments filtered for only inter- or intra-monomeric interactions, or from manually assigned restraints) is available.

Structure calculations of dimeric proteins can fail in two ways. Either the two monomers end up separated from each other, which can be detected easily; or the two monomers are brought together in an incorrect structure whose target function values and RMSDs to the mean coordinates, however, seem reasonable. This latter case is more difficult to detect. The RMSD to the mean coordinates and the target function values of the final ensemble are not suitable criteria for judging the quality of the homodimeric structures.

NikA tolerates 20% randomly unassigned interface ^{1}H chemical shifts (including backbone and side-chain chemical shifts), corresponding to 24 chemical shifts, which is comparable to the 19 chemical shifts that correspond to 45% omission of all assigned interface side-chain proton chemical shifts (excluding H^N, H^α, and H^β), which is also tolerated by NikA. This indicates that backbone and side-chain chemical shifts are similarly important in obtaining the correct fold of NikA in our test calculations, although backbone chemical shifts are conventionally thought to be less important than side-chain chemical shifts that provide predominantly long range NOEs and hence are much more important for tertiary structure determination. However, backbone chemical shifts are important for restraining contacts between β-strands in β-sheets. The dimer interface of NikA contains one β-sheet with two β-strands, and therefore the backbone chemical shifts are important for the structure determination of the NikA homodimer. The same phenomenon is even more evident in REI which is a pure β-sheet protein and uses only β-strands as interface. REI tolerances 40% unassigned interface side-chain proton chemical shifts, corresponding to 20 chemical shifts, but only 10% unassigned interface backbone and side-chain ^{1}H chemical shifts, corresponding to 15 chemical shifts. REI thus tolerates more omission of side-chain chemical shifts than omission of all ^{1}H chemical shifts.

With regard to the random omission of methyl and aromatic ^{1}H chemical shifts, incorrect structures were obtained by omitting all assigned aromatic chemical shifts in NikA (5 out of all 122 assigned interface proton chemical shifts, corresponding to 4.1%), all assigned methyl ^{1}H chemical shifts in NikA (19 out of 122, 15.6%), 90% of all assigned aromatic ^{1}H chemical shifts in WW (14 out of 113, 12.4%), or 80% omission of all assigned methyl ^{1}H chemical shifts in WW (4 out of 113, 3.5%), and 70% of all assigned aromatic ^{1}H chemical shifts in REI (9/151, 6.0%), or 60% omission of all assigned methyl ^{1}H chemical shifts in REI (9/151, 6.0%). These results indicate that the omission of aromatic and methyl proton chemical shifts causes more severe problems than the omission of the same number of chemical shifts out of all assigned proton chemical

shifts and hence aromatic and methyl proton chemical shifts play important roles in the structure determination of homodimeric proteins, as conventionally thought for the structure determination of monomeric proteins. The lack of a small number of "essential" chemical shifts can lead to incorrect structures. NikA tolerates up to 80% unassigned aromatic protons and 90% unassigned methyl protons, which is higher than the corresponding percentages of 60% and 50% for REI, respectively. This is probably due to the fact that more than half (66.7%, excluding the flexible N-terminus) of the NikA protein is involved in the dimer interface, as discussed above for the omission of all ^{1}H chemical shifts. In our test calculations, aromatic and methyl groups were also important for the interface interactions between two monomers in homodimeric proteins, especially for proteins with a relatively low percentage of interface residues, such as REI.

In the structure calculations with NikA, we found that the β-strands become gradually disordered with increasing percentages of unassigned interface ^{1}H chemical shifts, whereas the four-helix bundle, which is also involved in interface interactions, does not change significantly. This is in part due to the input of (intra-monomeric) hydrogen bond restraints for the helical regions in NikA. Without hydrogen bond restraints in the helical regions, calculations showed no change in the tolerance for unassigned ^{1}H chemical shifts, but disorder in the α-helices and the RMSD to the reference structure increased, which indicates that input of hydrogen bond restraints in helical regions is helpful, although the tolerance against unassigned chemical shifts in the determination of the correct overall dimer structure will not be improved.

5. Conclusions

Our results reveal that successful structure determinations of homodimeric proteins depend on having a high completeness, around 80–90%, of ^{1}H chemical shift assignments. The test calculations in this paper show that the percentage of residues involved in the inter-monomeric interface has an impact on the tolerance against unassigned ^{1}H interface chemical shifts. A protein with a higher percentage of interface residues (NikA) tolerates a higher percentage of unassigned ^{1}H chemical shifts than a protein with a smaller interface region (REI). In addition, methyl and aromatic proton chemical shifts are more important than other proton chemical shifts. The calculations showed that even a lack of, for instance, 5 "essential" aromatic chemical shifts for NikA or 4 "essential" methyl chemical shifts in WW can cause incorrect structures. For homodimeric proteins with β-strands involved in the interface between the two monomers, backbone chemical shifts play an important role in the structure determination. Structures can tolerate more unassigned side-chain proton chemical shifts if the backbone chemical shifts are completely assigned. For homodimeric proteins with α-helices involved in the interface between two monomers, the input of hydrogen bond restraints in the α-helical regions is helpful for the α-helix formation and for a slight reduction of the RMSD deviation. Finally, particular attention should be paid to the evaluation of structural quality and reliability. Some of our calculations, in which the percentage of unassigned interface proton chemical shifts was close to the limit of tolerance, yielded erroneous structure ensembles but with low RMSD and target function values.

Acknowledgments

We gratefully acknowledge financial support by the National Science Council of Taiwan and the Lichtenberg program of the Volkswagen Foundation.

References

[1] M. Williamson, How Proteins Work, Garland, New York, 2012.
[2] P. Guerry, T. Herrmann, Advances in automated NMR protein structure determination, Q. Rev. Biophys. 44 (2011) 257–309.
[3] P. Güntert, Automated structure determination from NMR spectra, Eur. Biophys. J. 38 (2009) 129–143.
[4] A. Rosato, J.M. Aramini, C. Arrowsmith, A. Bagaria, D. Baker, A. Cavalli, J.F. Doreleijers, A. Eletsky, A. Giachetti, P. Guerry, A. Gutmanas, P. Güntert, Y. He, T. Herrmann, Y.J. Huang, V. Jaravine, H.R.A. Jonker, M.A. Kennedy, O.F. Lange, G. Liu, T.E. Malliavin, R. Mani, B. Mao, G.T. Montelione, M. Nilges, P. Rossi, G. van der Schot, H. Schwalbe, T. Szyperski, M. Vendruscolo, R. Vernon, W.F. Vranken, S. de Vries, G.W. Vuister, B. Wu, Y. Yang, A.M.J.J. Bonvin, Blind testing of routine, fully automated determination of protein structures from NMR data, Structure 8 (2012) 227–236.
[5] M. Nilges, A calculation strategy for the structure determination of symmetrical dimers by ^{1}H-NMR, Proteins 17 (1993) 297–309.
[6] S.I. O'Donoghue, M. Nilges, Calculation of symmetric oligomer structures from NMR data, in: N.R. Krishna, L.J. Berliner (Eds.), Structure Computation and Dynamics in Protein NMR, Kluwer Academic/Plenum Publishers, New York, 1999, pp. 131–161.
[7] B. Bardiaux, A. Bernard, W. Rieping, M. Habeck, T.E. Malliavin, M. Nilges, Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA, Proteins 75 (2009) 569–585.
[8] C. Mumenthaler, P. Güntert, W. Braun, K. Wüthrich, Automated combined assignment of NOESY spectra and three-dimensional protein structure determination, J. Biomol. NMR 10 (1997) 351–362.
[9] C. Mumenthaler, W. Braun, Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry, J. Mol. Biol. 254 (1995) 465–480.
[10] M. Nilges, M.J. Macias, S.I. ODonoghue, H. Oschkinat, Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin, J. Mol. Biol. 269 (1997) 408–422.
[11] J.P. Linge, S.I. O'Donoghue, M. Nilges, Automated assignment of ambiguous nuclear overhauser effects with ARIA, Methods Enzymol. 339 (2001) 71–90.
[12] T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, J. Mol. Biol. 319 (2002) 209–227.
[13] J. Kuszewski, C.D. Schwieters, D.S. Garrett, R.A. Byrd, N. Tjandra, G.M. Clore, Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments, J. Am. Chem. Soc. 126 (2004) 6258–6273.
[14] Y.J. Huang, R. Tejero, R. Powers, G.T. Montelione, A topology-constrained distance network algorithm for protein structure determination from NOESY data, Proteins 62 (2006) 587–603.
[15] N.J. Greenfield, Y.J. Huang, T. Palm, G.V.T. Swapna, D. Monleon, G.T. Montelione, S.E. Hitchcock-DeGregori, Solution NMR structure and folding dynamics of the N terminus of a rat non-muscle α-tropomyosin in an engineered chimeric protein, J. Mol. Biol. 312 (2001) 833–847.
[16] W. Gronwald, S. Moussa, R. Elsner, A. Jung, B. Ganslmeier, J. Trenner, W. Kremer, K.P. Neidig, H.R. Kalbitzer, Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE), J. Biomol. NMR 23 (2002) 271–287.
[17] P. Savarin, S. Zinn-Justin, B. Gilquin, Variability in automated assignment of NOESY spectra and three-dimensional structure determination: A test case on three small disulfide-bonded proteins, J. Biomol. NMR 19 (2001) 49–62.
[18] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, J. Mol. Biol. 273 (1997) 283–298.
[19] H. Yoshida, N. Furuya, Y.J. Lin, P. Güntert, T. Komano, M. Kainosho, Structural basis of the role of the NikA ribbon-helix-helix domain in initiating bacterial conjugation, J. Mol. Biol. 384 (2008) 690–701.
[20] D.S. Wishart, B.D. Sykes, F.M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, Biochemistry 31 (1992) 1647–1651.
[21] N. Furuya, T. Komano, Specific binding of the NikA protein to one arm of 17-base-pair inverted repeat sequences within the oriT region of plasmid R64, J. Bacteriol. 177 (1995) 46–51.
[22] S. Ohnishi, P. Güntert, S. Koshiba, T. Tomizawa, R. Akasaka, N. Tochio, M. Sato, M. Inoue, T. Harada, S. Watanabe, A. Tanaka, M. Shirouzu, T. Kigawa, S. Yokoyama, Solution structure of an atypical WW domain in a novel β-clam-like dimeric form, FEBS Lett. 581 (2007) 462–468.
[23] O. Epp, E.E. Lattman, M. Schiffer, R. Huber, W. Palm, Molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution, Biochemistry 14 (1975) 4943–4952.
[24] E.L. Ulrich, H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R.K. Wenger, H.Y. Yao, J.L. Markley, BioMagResBank, Nucl. Acids Res. 36 (2008) D402–D408.