J Biomol NMR (2013) 55:179–187
DOI 10.1007/s10858-012-9700-z

ARTICLE

MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins

Yingqi Xu · Stephen Matthews

Received: 2 October 2012 / Accepted: 20 December 2012 / Published online: 5 January 2013

Abstract NMR studies of large proteins have gathered much interest in recent years, especially after methyl-transverse relaxation optimized spectroscopy was successfully applied to systems as large as ~1 MDa in molecular weight. However, to fully take advantage of these spectra, there is a need for convenient and robust methods for making resonance assignments rapidly. Here, we present an improved version of our program MAP-XS (methyl assignment prediction from X-ray structure) for the automatic assignment of methyl peaks, based on nuclear Overhauser effects (NOE) correlations and chemical shifts together with available structures. No manual analysis of the NOE data is needed in this new version, which helps to further accelerate the assignment process. A refined algorithm as well as more efficient sampling produces results from single runs of MAP-XSII using unanalyzed NOE data are comparable to those achieved by the old version using manually curated data with every NOE peak correctly attributed to the two related methyl peaks; in addition, checking the results from multiple parallel runs against each other provides an effective mechanism for getting rid of the wrong assignments while keeping the correct ones, which significantly improves the reliability of final assignments. The new program is tested against three different proteins and delivers ~95 % correct assignments; positive results are also achieved for tests using different cut-off distances for NOEs, structures of lower resolutions, and ambiguous residue types.

Keywords Methyl TROSY · Large proteins · NOE · PRE · Chemical shift · Assignment

Abbreviations

TROSY	Transverse relaxation optimized spectroscopy
NOE	Nuclear overhauser effect
PRE	Paramagnetic relaxation enhancement
MBP	Maltose binding protein
EIN	N-terminal domain of *E. coli* Enzyme I
RDC	Residual dipolar coupling
PCS	Pseudocontact shift

Introduction

NMR studies of biological molecules have long been restrained to relatively small proteins and nucleic acids due to fast signal decay and severe peak overlap in large systems. This situation has been greatly improved in recent years thanks to new methods for sample labeling and NMR experiments. In particular, transverse-relaxation optimized spectroscopy (TROSY) (Pervushin et al. 1997) has pushed the size limit amendable for NMR structure determination to ~100 kDa (Salzmann et al. 2000; Tugarinov et al. 2005) with perdeuterated samples, and it has become feasible for intermolecular interaction and dynamics studies on even larger systems, as in the cases of the 900 kDa GroEL·GroES complex (Fiaux et al. 2002) and the 670 kDa 20S proteasome

Y. Xu · S. Matthews (✉)
Division of Molecular Biosciences, Imperial College London, South Kensington, London SW7 2AZ, UK
e-mail: s.j.matthews@imperial.ac.uk

Y. Xu · S. Matthews
Cross Faculty NMR Centre, Imperial College London, South Kensington, London SW7 2AZ, UK

Y. Xu · S. Matthews
Centre for Structural Biology, Imperial College London, South Kensington, London SW7 2AZ, UK

core particle (Sprangers and Kay 2007b). Many of these high molecular weight proteins and protein complexes play important roles in biological processes. NMR provides unrivalled insight into the dynamic features of such systems and their importance in function. In this way NMR complements information available from both X-ray crystallography and electron microscopy, which provide the static structures.

Compared with amide-TROSY, methyl-TROSY (Tugarinov et al. 2003) combined with methyl-specific labeling (Ruschak and Kay 2009) generates high quality spectra due to the more favorable relaxation properties of methyl groups. On the other hand, these spectra can only be fully interpreted after the assignments of methyl chemical shifts have been made, typically through methyl-detected, 'out-and-back' experiments (Tugarinov and Kay 2003; Sheppard et al. 2009) when the backbone resonances have been assigned. However, for very large systems such as proteasome, the backbone resonance assignments are not available, so alternative ways have to be sought, such as the chemical shift comparison between the intact protein and its smaller fragments (Sprangers and Kay 2007b; Gelis et al. 2007), and spectral changes caused by site-specific mutations (Amero et al. 2011). Although both methods are straightforward and can generate reliable assignments, they require multiple samples and may leave many methyl groups unassigned due to complications from structural changes. These remaining methyl resonances can be subsequently assigned using structural restraints like NOE, PRE (paramagnetic relaxation enhancement) (Venditti et al. 2011) or PCS (pseudocontact shift) (Velyvis et al. 2009), with reference to available crystallographic structures. We have been investigating a more straightforward solution and have found that in favorable cases it is possible to assign almost all methyl resonances correctly by directly matching the structure-based prediction of NOE correlations and chemical shifts with experimental data, without resorting to fragmentation or mutation (Xu et al. 2009).

Our original MAP-XS program (Xu et al. 2009) attracted interest from the NMR community (Venditti et al. 2011; Chao et al. 2012), but also raised few issues. First, the program relied on NOEs being unambiguously attributed to specific methyl peaks, which was often difficult due to chemical shift degeneracy. Using 4D-NOESY combined with non-uniform sampling and newly-developed processing methods helps to generate high-resolution spectra (Wen et al. 2012) for which most peaks can be unambiguously assigned, but it is still necessary to deal properly with multiple assignment possibilities. Second, although most of the predictions are correct, a few wrong assignments are predicted, especially when the input data is not of sufficient quality or contains errors. Therefore, a better method for distinguishing correct from incorrect assignment is needed. Previously, we ranked methyl assignments according to how well the corresponding NOE and chemical shifts match, and in general the highly ranked ones were more reliable. Here we propose a new evaluation of results by cross-checking assignments between multiple, parallel runs of the program. We find that assignments consistent across all runs are the most reliable and in many other cases correct assignments can also be readily identified. Third, while the program is aimed at obtaining all the possible assignments from a single set of NOE data, sometimes the quality of the NOESY spectra are not good enough. In these cases, it would be sensible to include as much available data as possible, including confirmed assignments from traditional fragmentation and mutation approaches together with PRE and RDC measurements. We may also wish to exploit multiple sets of NOESY spectra, such as those collected with different mixing times, from differential labelling patterns, or different recording methods (i.e. 3D ^{13}C-resolved or ^{1}H-resolved, 4D). The improved version of MAP-XS can now accommodate much flexibility in the input data and supports all possible combinations. Finally, a suitable structure is critical for generating correct assignments, we will discuss our findings from assignment tests using different structures for the same protein system.

Materials and methods

Algorithm for the program

MAP-XSII has being recoded in java and is different from its prototype in a few aspects. First of all, the score used to drive the simulation is changed, in particular the part regarding NOE (Xu et al. 2009) has to be changed to better adapt for the ambiguity. All NOE peaks are subjected to an automatic analysis aiming to find out the methyl peaks to which they are originated from. Due to chemical shift degeneracy, there are inevitably many NOE peaks with multiple choices of methyl peaks, especially when there is only one chemical shift to be checked (3D NOESY) or the resolution is limited (traditional 4D NOESY). Later all the possible combinations of methyl peaks contributing to an ambiguous NOE peaks are checked to see whether any of them is compatible with the structure according to the running methyl assignments. If such a match does occur, then a score for the methyl pair is calculated according to the following equation:

$$S_{NOE} = s \cdot \log \frac{D}{d} \cdot e^{-0.5 \cdot \sum_i \left(\frac{\delta_i^{NOE} - \delta_i^{p}}{\sigma_i} \right)^2} \quad (1)$$

where s is 1 if the symmetrical NOE is also observed or 0.4 otherwise, D is set to 5.0 Å and d is the distance between the two methyl groups in structure, while δ^{NOE} and δ^{p} denotes

the chemical shifts of NOE peak and corresponding methyl peaks respectively, σ denotes the chemical shift error for NOE peaks, and i runs through the dimensions of the NOESY. So the NOE score favors on those assignments with symmetrical peaks, corresponding to short distances, and matches better in chemical shifts. When multiple pairs of methyl peaks can contribute to the NOE, the sum of corresponding scores is calculated, but subjects to a ceiling constant, which is 2 by default; on the other hand, if none of the possible assignments can satisfy the distance requirement, a punishment score would be applied.

Chemical shifts are used in the same way as before. In addition, all other relevant measurements, such as PRE, PCS, RDC, can all be included into the score similarly for assignment prediction. Plenty of programs already exist for predictions of chemical shifts (Han et al. 2011), PRE (Venditti et al. 2011), PCS (Velyvis et al. 2009), and RDC (Sprangers and Kay 2007a), and outputs from them are used to match with the experimental values in MAP-XSII. The global score is summarized as:

$$S = -w_{cs}\sum_i\left(\frac{\delta_i^{\exp}-\delta_i^{calc}}{\sigma_i}\right)^2+\sum_p S_{NOE} - w_{PRE}\sum_N\left(\Gamma_2^{\exp}(N)-\Gamma_2^{calc}(N)\right)^2 \quad (2)$$

where items for chemical shifts, NOE, and PRE are included, while unlisted items for PCS and RDC are very similar to PRE. w_{CS} and w_{PRE} are adjustable weighting factors; δ, Γ_2, and σ represents chemical shift, transverse PRE rate, and chemical shift prediction error, respectively; i, p, and N runs through available chemical shifts, NOE peaks, and PRE measurements respectively.

The ambiguous nature in NOE scores also necessitates a more rigorous sampling method. A Metropolis Monte Carlo optimization procedure is applied to randomly generated initial assignments to achieve the best possible match of experimental and predicted NOE correlations and chemical shifts as well as other available NMR data. Many more swapping steps and corresponding score evaluation calculations are needed in the current version than the original MAP-XS, for which a downhill only sampling method has been able to get close enough to the best solution starting from initial assignments with about 50 % correct. To speed up the program, the global score is distributed across the methyl groups so that only a small proportion of local scores need to be evaluated after each swap. The scores related to chemical shift, PRE, and RDC data are easily attributed to corresponding methyl groups. For NOEs the score relating to each peak is equally apportioned to the two methyls involved. The methyl-specific scores can also help assess the assignment reliability, in a similar way previously used to rank the assignments (Xu et al. 2009). This strategy is especially useful when using 4D NOESY spectra, where ambiguity is reduced; for example, our program can produce results within 1 min for the test data of MBP. As multiple runs are often necessary to estimate cut-off distances as well as to generate more reliable assignments through cross-checking, this speed improvement makes significant difference to performance and usability.

Cross-checking of the assignments

We typically run 20 tests at a specific cut-off distance and sort the results according to the total score. The 10 best results (the number can be slightly different, depending on where the gap in the scores is located) are chosen, then for each peak the assignments from these tests are compared. The assignments consistent in all the tests are deemed more reliable (see Results) and kept. Some peaks from leucine or valine are assigned to the geminal methyl groups from the same residue, for which either one should be fine, and we usually keep the one assigned in most tests. In addition, when an assignment occurs in more than 70 % of tests and it also represents one of the highest scores, we also keep it. These simple choices are automatically made by the program, but further decisions can be made confident through manual inspection of the output.

NMR data

The program is tested on three systems with experimental data: the 360 kDa $\alpha_7\alpha_7$ proteasome (Sprangers and Kay 2007b), the 27 kDa N-terminal domain of *E. coli* Enzyme I (EIN) (Venditti et al. 2011), and the 42 kDa *E. coli* maltose binding protein(Wen et al. 2012). The NMR data for proteasome, EIN, and MBP were acquired from the groups of Lewis Kay, Marius Clore, and Jihui Wu respectively. For proteasome, a 3D ^{13}C-resolved NOESY on ILV-labelled sample is used; for EIN, a 4D NOESY on AILV-labelled sample is used, and PRE measurements for 5 different samples are also available; for MBP, a sparsely-sampled 4D NOESY with diagonal peaks suppressed on ILV-labelled sample is used. Details regarding the samples and the NMR experiments can be found in previous publications (Sprangers and Kay 2007b; Wen et al. 2012; Venditti et al. 2011).

We also simulated 4D NOE data for malate synthase G (MSG) using the experimental chemical shifts (Tugarinov and Kay 2003) and crystal structure (1D8C). A cutoff distance of 7.0 Å was used; probability of NOE peaks getting missing was linearly increased from 0 to 0.6 according to the corresponding distance in the range of 4.0–7.0 Å, and the probability of the symmetrical peaks being observed was also similarly simulated; positions of

all NOE peaks were allowed to be randomly shifted in a small range (0.02 and 0.2 ppm for ^{1}H and ^{13}C respectively) centered at the corresponding chemical shifts.

Results

Using unassigned NOE peaks to speed up the process

Although the existence of ambiguous NOE assignments and poor initial methyl assignments (e.g. randomly generated, which are often 100 % wrong) necessitate an extended swapping process for better sampling, we find that the best results generated are always comparable to those generated by the old MAP-XS (Xu et al. 2009) using manually assigned NOE peaks. For example, the results of 20 trial runs for the proteasome are listed in Fig. 1, which have been sorted according to the global scores. As expected, the trials with higher scores make more correct assignments in general, although for more similar scores this correlation is not strictly observed, which reflects the zig-zag nature of the score map and the limited space being sampled. The upper half of these trials generate >90 % correct assignments (93 peaks in total).

Similar results are also achieved for EIN and MBP (Table 1), for which ~90 and ~80 % assignments are correct, respectively. The difference in the percentage of correct assignment can be attributed to the extent of NOE coverage and completeness, where a more intensive network leads to more reliable result. Venditti et al. (2011) mentioned that the old MAP-XS didn't work well with EIN, probably due to the NOE ambiguity and limitations in chemical shift prediction; the new MAP-XS, however, can correctly assign the vast majority of the methyl peaks, without using any PRE data.

Using cross-checking to generate more reliable assignments

The likely rough surface of the global score and the random nature of the swapping steps would imply that the results from single runs could be very different. Therefore, we often run the program many times, so that optimal results are selected and compared with each other. In Fig. 1 the methyl peaks colored green are consistently assigned to the same methyl group in all selected trials, whereas yellow and red peaks have been assigned to different methyl groups among these trials, with the yellow peaks having a dominant assignment with an occurrence >70 %. Manual inspections of these peaks with assignment discrepancies suggest ways to make choices. In Fig. 2a the peak is assigned to HD2 of L38 in 9 trials and to HD1 of L38 in another trial. As these assignments belong to the same residue and we are not aiming to make stereo-specific assignments, either assignment is satisfactory. In practice we simply choose the dominant one or the top score if neither is dominant. In Fig. 2b, the peak is assigned to I70 in 9 trials and to I144 in the

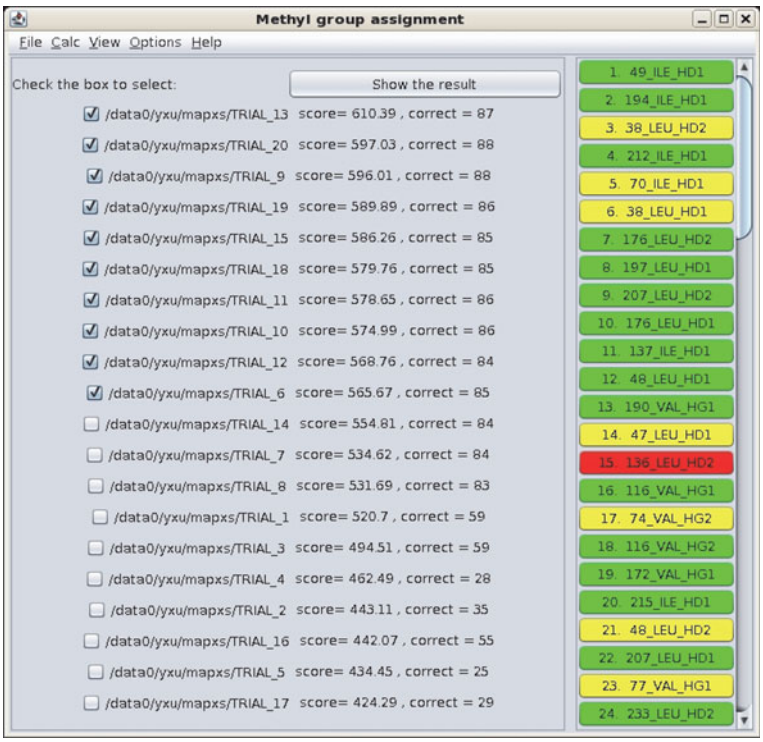

Fig. 1 Screenshot showing the sorted results of methyl assignment trial runs for proteasome, at the cut-off distance at 9.0 Å. For each trial, the file saving the assignment result, the global score, and the number of correct assignments are shown. Users can select those trials with higher scores, and the consistency of assignments for each peak among the selected trials can be viewed on the right column (see the text for further explanations)

Table 1 Summary of the assignment results for the 3 test proteins

	EIN	MBP	Proteasome
Number of methyl peaks	139 (131 + 8)[a]	121	93
Number of NOE peaks	716	319	684
Cut-off distance used (Å)	7.5	6.5	9.0
Expected NOE number	1044	526	854
PDB ID for structure	1ZYM	1DMB	1YAU
Number of selected trials	8/20	14/20	10/20
Scores for selected trials	561.5–567.7	271.9–290.3	565.7–610.4
Number of correct assignments for selected trials	121–124	86–100	84–88
Number of correct/wrong assignments for revised assignments	120/2	80/6	82/1
Average CPU time consumed	~2 min	<1 min	~20 min

[a] 8 of the 139 peaks have no confirmed assignments

remaining one; I70 should be the choice as it is dominant and has more favorable scores. These selections are made by the program and together with the consistent assignments form the revised assignment list, which possess similar numbers of correct assignments as single runs, but many fewer wrong assignments (Table 1). Manually, further selections can be made, such as in Fig. 2c, where the peak should be assigned to I67 rather than I12, as the former has much higher scores.

The program automatically makes 122, 86, and 83 assignments for EIN, MBP, and proteasome respectively, with only 2, 6, and 1 errors respectively. For both EIN and proteasome, where an intense NOE network is available (>5 NOE connections per methyl peak on average), the assignment reliability can be higher than 98 %; for MBP, where ~2.5 NOE connections are available per methyl peak, the assignment reliability is also higher than 93 %. Most of the wrong assignments are characterized with very few NOE connections; when they do have many NOE connections, the wrongly assigned methyl groups often locate closely to the real ones in the structure.

Using different distance cut-off for NOE

Whether an NOE correlation can be observed depends on many factors besides the corresponding distance, and in practice it is difficult to specify a single distance cut-off for a NOESY spectrum. Usually more peaks are absent with increased distances, although a few NOE correlations can be observed among protons beyond the cut-off for the majority of the resonances, due to favorable relaxation, internal motions and/or spin diffusion. For simplicity, we still predict NOE correlations by calculating distances according to the structure and applying a cut-off. To see how the program performs with different cut-off distances, we have tested at a range of cut-off distances for all three proteins.

In Fig. 3, the scores, numbers of correct assignments for single trials and numbers of correct and wrong assignments after cross-checking are plotted against the cut-off distances. In general, higher scores are obtained with larger cut-off distances, which is expected as the NOE correlations

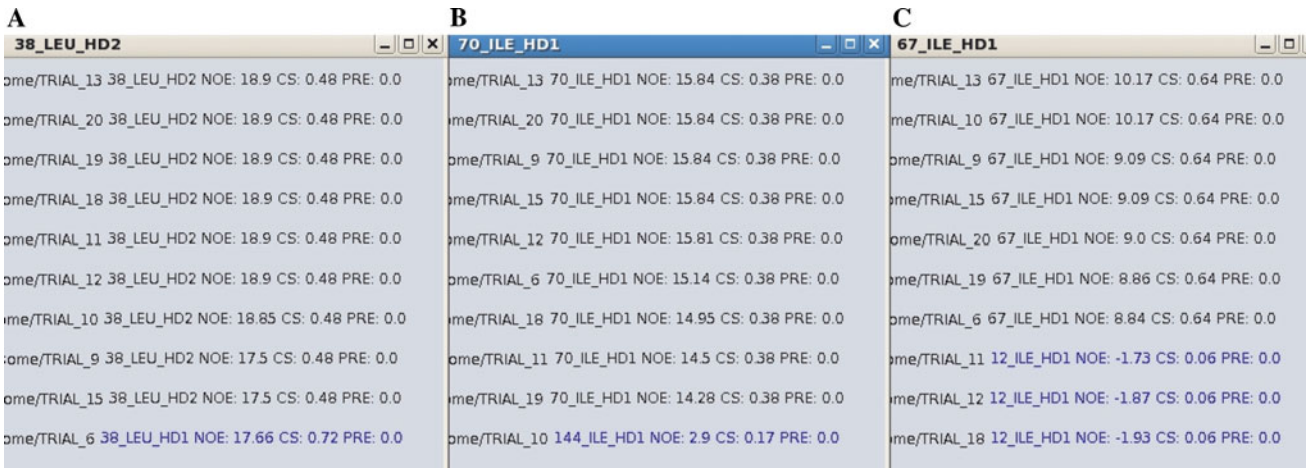

Fig. 2 Screenshots showing some example methyl peaks assigned differently in multiple runs. Checking the scores for these assignments helps to identify the right ones

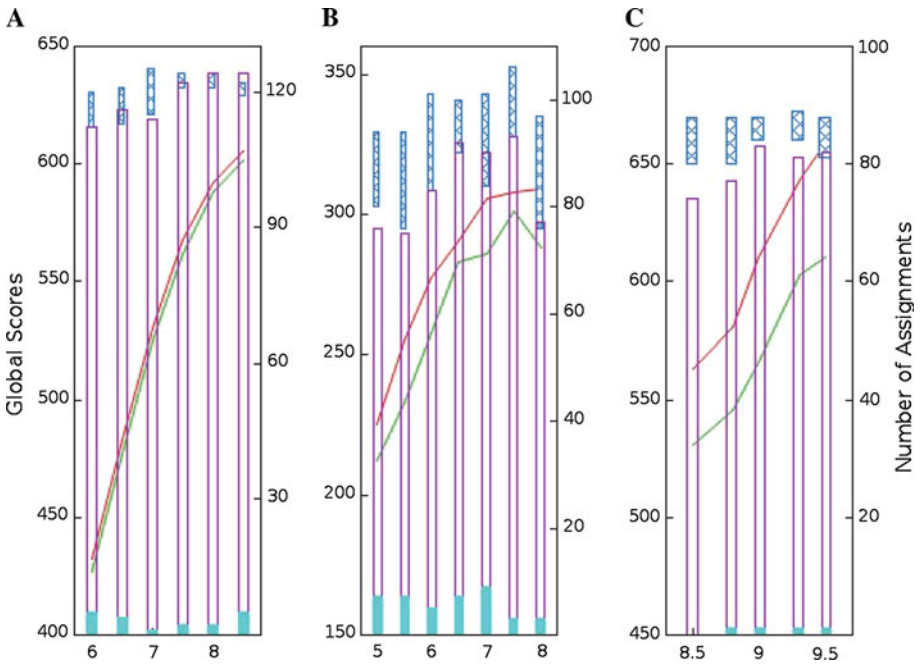

Fig. 3 Summary of assignment results for EIN (**a**), MBP (**b**) and proteasome (**c**) at different cut-off distances. The final scores and correct assignment numbers for single trials are *plotted* as lines (*red* for highest scores, *green* for lowest scores) and *error bars*, while the numbers of correct/wrong assignments after cross-checking are *plotted* as *empty/crossed boxes*

rising from far apart proton pairs can only be correctly matched when the cut-off distances are long enough. On the other hand, increased cut-off also leads to increased ambiguity of the NOE assignment, especially for 3D NOESY data, which slows the calculation down and decreases the convergence. Despite this, our program tolerates a range of the cut-off distances, with very similar assignment correctness for most of the trials. In addition, >90 % assignments after cross checking are correct in all these cases. In practice, the run with the most assignments made after cross checking coincides with the highest correct percentage of these assignments.

Using different structures

As a structure-based prediction method, the output results rely heavily on the reliability of the structure used. We have tested these case protein systems with different structures, and interestingly, there appear to be a correlation between the assignment reliability and the scores generated by our program (Fig. 4). We can therefore run tests with all the different structures available and take the ones with highest scores. Another observation is that results from the worse structure are still reliable, with <10 % errors for the cross-checked assignments, but the number of assigned peaks drops significantly, therefore the coverage of usable probes will be affected. We reason that only when the structure fits well with the experimental NOESY can consistent (and correct) assignments be achieved; otherwise there will be different solutions matching the NOE with the structure similarly well (but not the best possible), such assignment will most likely be removed during cross-checking.

Using ambiguous residue type assignments

Knowledge of the residue type for every methyl peak can greatly reduce the number of possible assignment solutions, and this was a requirement as an input for our previous version of MAP-XS. The residue types for methyl peaks are often identified through their characteristic chemical shifts or specific labeling patterns. In practice, methyl peaks from isoleucine, alanine, and methionine can all be readily distinguished, while methyl peaks from valine and leucine are more difficult to discern due to overlapped chemical shift distributions and shared precursors for isotope labeling. Although peaks from valine and leucine can be unambiguously identified through experiments correlating them with other side chain ^{13}C resonances that are distinct for these two amino acids (Hu et al. 2012), we have introduced an ambiguous residue type VL in this new version, which represents the possible type of either Leu or Val.

To test how the ambiguous typing affects the performance, we have run the program with all Leu and Val methyls typed to VL. Although the general performance is not as good as that achieved with unambiguous residue type assignments, it can be seen from Fig. 5 that the revised assignments are still highly reliable. The performance deterioration can be attributed to worse convergence due to greater ambiguity, as evidenced by the lower and/or broader distribution of final scores in Fig. 5; and employing longer simulations results in improved scores. In practice, we still

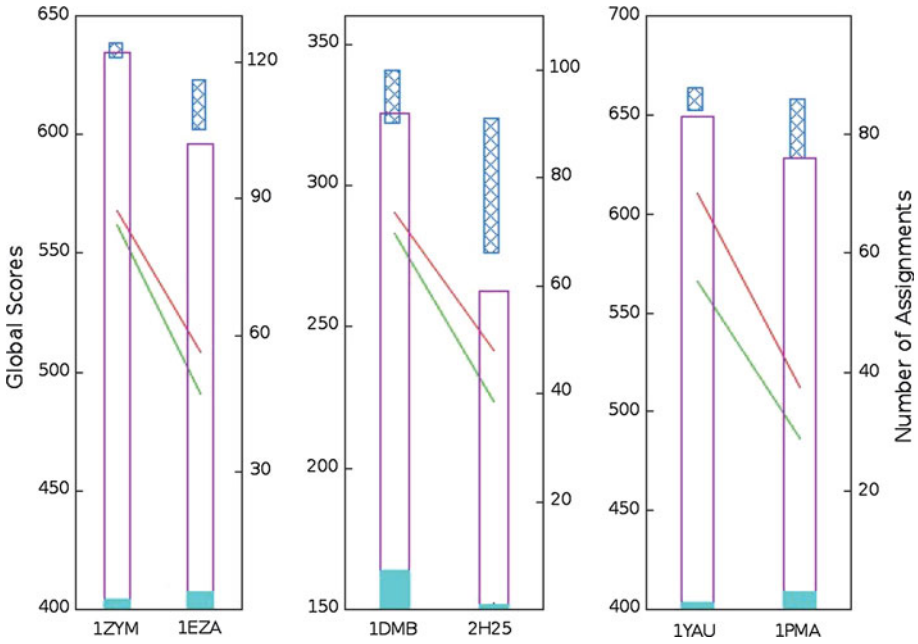

Fig. 4 Comparison of assignment results when different structures are used. 1ZYM, 1DMB, and 1YAU are the structures we mainly used in this paper, while 1EZA, 2H25, and 1PMA are alternate structures for EIN, MBP, and proteasome respectively. Among them, 1EZA and 2H25 are NMR-determined structures, and 1PMA is a crystallographic structure with a relatively low resolution (3.4 Å). When superimposed, the heavy atoms from these alternate structures are ~2 Å away from their counterparts in average. The plots show the same parameters as in Fig. 3

expect that many leucine and valine peaks can be unambiguously characterized, so that optimal performance will be retained.

Discussion and conclusions

Although the NOESY spectra for very large protein complexes can suffer from low sensitivity as well as low resolution, the examples described here suggest that they are very informative and work well for proteins with good solubility. New methods for recording 4D NOESY spectra with high resolution are also available, greatly reducing the errors or ambiguity of NOE assignments (Wen et al. 2012). Although a complete NOE network is expected to generate reliable assignments, in the case of MBP, for which only ~2.5 NOE connections per methyl peak are observed, our program generates assignments for more than 2/3 methyl peaks with >90 % correct. Although the NOE cut-off distances for all three examples are quite large, MAP-XS worked well for a wide range of cut-off settings, suggesting that it would be advisable to use longer mixing times so that more NOE connections can be used.

The assignment problem becomes more challenging with growing complexity from longer sequences with more

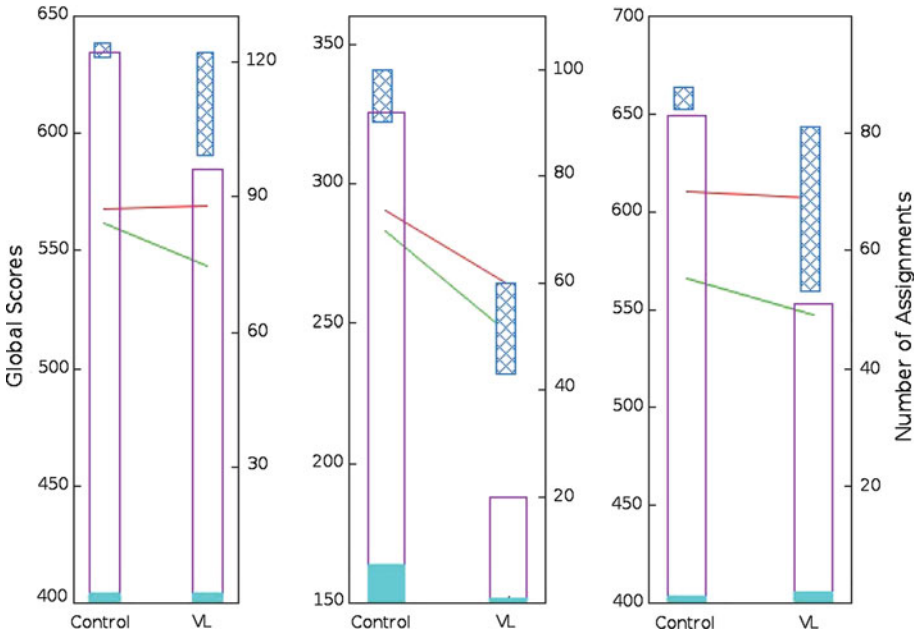

Fig. 5 Comparison of assignment results when the methyl peaks from Val and Leu residues are not manually distinguished (VL). From *left* to *right* represents EIN, MBP, and proteasome respectively

methyl peaks, due to the inevitable peak overlapping. To see how well MAP-XS can cope with this, we tested the program with simulated NOE data for malate synthase G (MSG); 1101 NOE peaks were generated among the 270 methyl peaks with assigned chemical shifts (Tugarinov and Kay 2003); considering symmetrical peaks, the number of simulated NOE peaks is only slightly more than the number of reported distance restraints among methyl groups (Tugarinov et al. 2005). Tests show that 170–210 peaks can often be correctly assigned with single trials, while cross-checking helps to generate 143 assignments automatically, with only 5 errors, which originates from closely located methyl groups and those with no NOE or missing in the crystal structure.

New types of structural restraints such as RDC, PCS, and PRE have been playing an increasingly important role in NMR structure determination, especially for large proteins (Ruschak et al. 2010; Gelis et al. 2007). These restraints have also been used to make resonance assignments when the structure is already known. Compared with NOEs, they provide information about structural order on much larger scales, which can be particularly advantageous for the assignment of isolated methyl groups where NOEs are absent. Therefore, in addition to improving the algorithm of NOE-based assignment, we also endeavor to provide a platform where all available data can be integrated so that reliable assignments can be made as quickly and easily as possible. In addition to unassigned NOEs as input data, our program can accommodate PRE, PCS and RDC data, and unambiguous assignments made through other methods, such as mutation, can also be fixed.

As an example we have used the PRE data for EIN to examine whether improved results can be achieved. For simplicity, we took the predicted values for PREs in the last round refinement from Venditti et al. (2011) to be compared with the experimental values. The inclusion of PRE restraints improves greatly the convergence of the simulation, but no further improvement in the overall assignment reliability is obtained. This is probably due to the high quality NOE and chemical shift data that is available. As noted previously, some PRE restraints are not fully consistent with the monomer structure, likely due to transient interactions in the solution; this explains why simple addition of PRE into the score fails to deliver results as good as the ones manually validated in Venditti et al. (2011). When the NOE connections are sparse, we expect that additional data such as PRE measurements will help greatly with the assignment, especially when there is no conflict between PRE and NOE.

Usage of PRE, PCS, and RDC represents more work, however. In addition to extra samples needed for the experimental measurements, predictions of PRE, PCS, and RDC values are also more complicated than NOE predictions. For PRE predictions, the positions of spin labels often need to refined; for PCS and RDC, prior knowledge of the magnetic susceptibility tensor and the molecular alignment tensor are needed respectively. Fortunately, with our improved MAP-XS, it's expected that at least some assignments can be made reliably with only NOE and chemical shifts, and these assignments may then be used to help refine the spin-labeled structure, decide the magnetic susceptibility tensor or the molecular alignment tensor. It can also be anticipated that initial results from NOE-based assignments will help make the follow on decisions such as how many and what extra samples would be needed to fully validate or complement the assignments.

In conclusion the new version of MAP-XS has removed the previous requirements of manual NOE assignment and unambiguous residue type assignment, found a better way of picking out reliable assignments, and provided a simple platform for users to integrate all kinds of available data. With greatly improved usability, we hope that the program can bring more ease for people working on large systems.

Acknowledgments The research is supported by grants from the Wellcome Trust and BBSRC. We are grateful to Dr. Sprangers and Prof. Kay, Dr. Venditti and Prof. Clore, Dr. Wen and Prof. Wu for sharing their data on proteasome, EIN, and MBP respectively. The program can be downloaded from nmr.bc.ic.ac.uk/map-xs/.

References

Amero C, Dura MA, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin MJ, Vernet T, Franzetti B, Boisbouvier J (2011) A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J Biomol NMR 50(3): 229–236. doi:10.1007/s10858-011-9513-5

Chao FA, Shi L, Masterson LR, Veglia G (2012) FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. J Magn Reson 214:103–110. doi:10.1016/j.jmr.2011.10.008

Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K (2002) NMR analysis of a 900 K GroEL GroES complex. Nature 418(6894):207–211

Gelis I, Bonvin A, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131(4):756–769. doi:10.1016/j.cell.2007.09.039

Han B, Liu YF, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50(1):43–57. doi:10.1007/s10858-011-9478-4

Hu WD, Namanja AT, Wong S, Chen Y (2012) Selective editing of Val and Leu methyl groups in high molecular weight protein NMR. J Biomol NMR 53(2):113–124. doi:10.1007/s10858-012-9629-2

Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94(23):12366–12371

Ruschak AM, Kay LE (2009) Methyl groups as probes of supramolecular structure, dynamics and function. J Biomol NMR 46(1):75–87. doi:10.1007/s10858-009-9376-1

Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE (2010) The proteasome antechamber maintains substrates in an unfolded state. Nature 467(7317):868–871. doi:10.1038/nature09444

Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K (2000) NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. J Am Chem Soc 122(31):7543–7548

Sheppard D, Guo CY, Tugarinov V (2009) 4D H-1-C-13 NMR Spectroscopy for assignments of alanine methyls in large and complex protein structures. J Am Chem Soc 131(4):1364. doi:10.1021/ja808202q

Sprangers R, Kay LE (2007a) Probing supramolecular structure from measurement of methyl (1)H-(13)C residual dipolar couplings. J Am Chem Soc 129(42):12668–12669. doi:10.1021/ja075846i

Sprangers R, Kay LE (2007b) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445(7128):618–622. doi:10.1038/nature05512

Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc 125(45):13868–13878. doi:10.1021/ja030345s

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125(34):10420–10428. doi:10.1021/ja030153x

Tugarinov V, Choy WY, Orekhov VY, Kay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci U S A 102(3):622–627. doi:10.1073/pnas.0407792102

Velyvis A, Schachman HK, Kay LE (2009) Assignment of Ile, Leu, and Val Methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase. J Am Chem Soc 131(45):16534–16543. doi:10.1021/ja906978r

Venditti V, Fawzi NL, Clore GM (2011) Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl–methyl nuclear overhauser enhancement spectroscopy. J Biomol NMR 51(3):319–328. doi:10.1007/s10858-011-9559-4

Wen J, Zhou P, Wu JH (2012) Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl–methyl NOESY for large proteins. J Magn Reson 218:128–132. doi:10.1016/j.jmr.2012.02.021

Xu YQ, Liu MH, Simpson PJ, Isaacson R, Cota E, Marchant J, Yang DW, Zhang XD, Freemont P, Matthews S (2009) Automated assignment in selectively methyl-labeled proteins. J Am Chem Soc 131(27):9480. doi:10.1021/ja9020233