



Published in final edited form as:

*J Magn Reson.* 2015 April ; 253: 166–172. doi:10.1016/j.jmr.2015.02.006.

## On the problem of resonance assignments in solid state NMR of uniformly $^{15}\text{N}$ , $^{13}\text{C}$ -labeled proteins

**Robert Tycko**

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

### Abstract

Determination of accurate resonance assignments from multidimensional chemical shift correlation spectra is one of the major problems in biomolecular solid state NMR, particularly for relative large proteins with less-than-ideal NMR linewidths. This article investigates the difficulty of resonance assignment, using a computational Monte Carlo/simulated annealing (MCSA) algorithm to search for assignments from artificial three-dimensional spectra that are constructed from the reported isotropic  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shifts of two proteins whose structures have been determined by solution NMR methods. The results demonstrate how assignment simulations can provide new insights into factors that affect the assignment process, which can then help guide the design of experimental strategies. Specifically, simulations are performed for the catalytic domain of SrtC (147 residues, primarily  $\beta$ -sheet secondary structure) and the N-terminal domain of MLKL (166 residues, primarily  $\alpha$ -helical secondary structure). Assuming unambiguous residue-type assignments and four ideal three-dimensional data sets (NCACX, NCOCX, CONCA, and CANCA), uncertainties in chemical shifts must be less than 0.4 ppm for all MCSA runs to yield fully correct assignments for SrtC, and less than 0.2 ppm for MLKL. Eliminating CANCA data has no significant effect, but additionally eliminating CONCA data leads to more stringent requirements for chemical shift precision. Introducing moderate ambiguities in residue-type assignments does not have a significant effect.

### Introduction

Site-specific resonance assignments are usually a prerequisite for the extraction of useful structural or dynamical information from NMR spectra of proteins. When proteins are labeled with  $^{15}\text{N}$  and  $^{13}\text{C}$  only at specific sites or in a small number of amino acids [1, 2], assignments are trivial. However, when proteins are uniformly or extensively  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled, as is now a standard practice in biomolecular solid state NMR, assignments must be obtained by an analysis of multiple multidimensional chemical shift correlation spectra [3–

© 2015 Published by Elsevier Inc.

contact information: Dr. Robert Tycko, National Institutes of Health, Building 5, Room 112, Bethesda, MD 20892-0520 USA. phone 1-301-402-8272. robertty@mail.nih.gov.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

18]. Although in principle this analysis may seem straightforward, in practice it can be quite difficult, tedious, and error-prone. The limited signal-to-noise ratios and relatively broad lines of most solid state NMR spectra of proteins, compared with their solution NMR counterparts, makes the task of obtaining site-specific resonance assignments from multidimensional solid state NMR spectra especially problematic.

In an infamous paper published in 1996, when solid state NMR spectroscopy of uniformly labeled proteins under magic-angle spinning was in its infancy, I attempted to quantify how the difficulty of obtaining unique assignments would depend on the solid state NMR linewidths [19]. Using isotropic  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shifts from solution NMR studies of the 76-residue protein ubiquitin, I constructed a set of artificial three-dimensional (3D) spectra for protein segments of various lengths. I then used an elaborate computational algorithm to search for resonance assignments in a sequential manner, assuming various linewidths (*i.e.*, various uncertainties in the precise values of the chemical shifts when comparing crosspeak positions from different 3D spectra). By running the same algorithm many times, I determined a lower limit on the number of distinct resonance assignments that were consistent with all 3D spectra, as a function of the assumed linewidth and the length of the protein segment. The main result was that the linewidths would have to be under 0.5 ppm for the assignments to be unique if the protein segment was more than 30 residues in length [19].

Although this old result is not terribly inaccurate, the 3D spectra considered in these old simulations were not the same as 3D spectra that are now used in real experiments on real proteins. Most importantly, sidechain chemical shifts were not included in the artificial 3D spectra. Moreover, the well-known correlations of  $^{13}\text{C}$  chemical shifts with residue type, which in practice are very important in the assignment process, were not used in the assignment algorithm.

More recently, as my research group began to study uniformly labeled proteins in noncrystalline states, such as amyloid fibrils and viral capsids [20–22], we realized that the “traditional” manual and sequential approach to resonance assignment, which can be very effective when applied to proteins with very sharp solid state NMR lines, can easily fail when applied to non-ideal (but scientifically interesting) systems. The main problem is that, although one may be able to find a plausible set of resonance assignments that seems fully consistent with one’s multidimensional spectra, there is no guarantee that this set of assignments is unique. We were therefore led to develop a more general computational approach for analyzing multidimensional data, based on a relatively simple and flexible Monte Carlo/simulated annealing (MCSA) algorithm [23, 24]. The MCSA algorithm allows us to treat nearly any combination of multidimensional spectra, to include partial residue-type assignments derived from chemical shift values (or from other information), and to repeat the assignment process many times in order to identify protein segments that have unique resonance assignments, segments that have non-unique assignments, and segments that are “invisible”. The same algorithm can also be applied to nucleic acids or other heteropolymers, and to solution NMR data. The MCSA algorithm, contained in a Fortran95 program called massign2b (available upon request from robertty@mail.nih.gov), has been used in several of our recent publications [20–22, 25].

The resonance assignment problem is a complicated one. The uniqueness of assignments from a given set of multidimensional data obviously depends on the linewidths and on the length of the protein sequence. It may also depend on the protein's secondary structure, the diversity of the amino acid composition, the presence or absence of repetition in the sequence, and the extent of "invisible" segments (*i.e.*, protein segments that do not contribute to the solid state NMR data, due to static or dynamic disorder). Since every protein has its own characteristics, both in terms of its biochemical/biophysical properties and its NMR properties, it is not possible to formulate definite rules that govern the uniqueness of assignments in all cases. Nonetheless, given the importance of the assignment problem, the time-consuming nature of sample preparations and solid state NMR measurements, the availability of the MCSA algorithm, and the relative ease of computer simulations, it makes sense to use computational simulations to investigate how the uniqueness of assignments depends on various factors. Results from assignment simulations can help guide one's choice of samples and measurements in real experiments.

## MCSA algorithm

Figure 1 shows a simplified flow chart for the MCSA algorithm [24]. One first prepares input files, including one file that contains the protein sequence, at least one signal table for each multidimensional spectrum that contains rows of correlated chemical shift values, linewidths (*i.e.*, uncertainties in the chemical shifts), and residue-type assignments, and one file that specifies the "connections" between signal tables. Signal tables are created manually, by careful inspection of the experimental data to identify crosspeaks and sets of crosspeaks that correlate with one another. The number of chemical shift columns in each table can vary, but is normally at least equal to the dimension of the corresponding NMR spectrum. Residue-type assignments for each signal row need not be unambiguous, as discussed below. The connection table specifies which pairs of chemical shift columns from different signal tables should agree after signals (*i.e.*, rows from signal tables) are assigned to specific residues.

One then runs the `massign2b` program, providing as input the minimum and maximum values of coefficients  $w_G$ ,  $w_B$ ,  $w_E$ , and  $w_U$  in the score function, the number of Monte Carlo attempts in each independent MCSA run, and the number of runs. The score for a given assignment of signals to residues in the protein sequence is defined to be  $S(n_G, n_B, n_E, n_U) = w_G n_G - w_B n_B - w_E n_E + w_U n_U$ , where  $n_G$  is the number of good connections (*i.e.*, agreements between chemical shifts from different signal tables),  $n_B$  is the number of bad connections (*i.e.*, disagreements between chemical shifts from different signal tables),  $n_E$  is the number of assignment "edges" (*i.e.*, connections that can not be tested due to gaps between the residues to which signals have been assigned), and  $n_U$  is the total number of rows from all signal tables that have been "used" (*i.e.*, assigned to specific residues). In evaluating the connections, two chemical shifts are considered to agree when the square of their difference is less than the sum of their squared uncertainties.

The `massign2b` program seeks to maximize the score, which means maximizing the numbers of good connections and used signals and minimizing the numbers of bad connections and edges. Typically, the initial assignments at each residue are "null", meaning

that no signals are assigned to any residue. In each Monte Carlo attempt, the assignment from one randomly chosen signal table at one randomly chosen residue is changed (or replaced by a null assignment) and the concomitant change in score  $S$  is evaluated. The attempt is then accepted or rejected according to the standard Metropolis condition [26]. That is, the attempt is accepted only if  $\exp(-\Delta S)$  is greater than a random number between 0.0 and 1.0. As the run proceeds, the score function coefficients are increased from their minimum to their maximum values, making the acceptance probability decrease toward zero.

Compared with the original `massign2` program described by Hu *et al.* [24], our current `massign2b` program includes two additional features: (1) Rather than starting with null assignments, one can start with a set of tentative or partial assignments, determined from an earlier MCSA run or from other information; (2) Rather than increasing the score coefficients linearly from their minimum to their maximum values, the score coefficients can remain constant once the acceptance probability falls to a specified value, then increase to their maximum values toward the end of a run. These features were added to facilitate applications to larger proteins, as described by Bayro *et al.* [22]

It should be noted that many other research groups have developed and applied automated or computer-aided methods for obtaining resonance assignments from multidimensional NMR spectra [27–43], including methods that are conceptually related to the MCSA algorithm [35, 40]. The point of this article is not to demonstrate that a particular algorithm or program is the best choice for resonance assignments, but rather to illustrate how simulations can provide useful insights into various factors and requirements related to the uniqueness of assignments. Other algorithms or programs can certainly be used to accomplish the same ends.

## Simulated assignments for two proteins with different secondary structures

### Choice of proteins and construction of 3D spectra

For the simulations described below, two proteins with similar sizes and with nearly complete  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shift assignments from solution NMR were chosen: (1) The 147-residue catalytic domain of the *Bacillus anthracis* sortase protein SrtC, whose structure was reported by Robson *et al.* [44] (Protein Data Bank code 2LN7, BioMagResBank code 18152); (2) The 166-residue N-terminal domain of the mixed-lineage kinase domain-like (MLKL) protein, whose structure was reported by Su *et al.* [45] (Protein Data Bank code 2MSV, BioMagResBank code 25135). As shown in Figure 2, these two proteins have quite different secondary structure compositions, with SrtC being primarily  $\beta$ -strands and MLKL being primarily  $\alpha$ -helices.

For each protein, four signal tables representing artificial 3D solid state NMR spectra were constructed from the reported solution NMR chemical shifts. Rows in the first table, representing an NCACX spectrum, contain chemical shifts of  $N_i$ ,  $CO_i$ ,  $Ca_i$ , and  $C\beta_i$ , *i.e.*, backbone amide nitrogen, backbone carbonyl carbon,  $\alpha$ -carbon, and  $\beta$ -carbon of residue  $i$ . Rows in the second table, representing an NCOCX spectrum, contain chemical shifts of  $N_{i+1}$ ,  $CO_i$ ,  $Ca_i$ , and  $C\beta_i$ . Rows in the third table, representing a CONCA spectrum, contain

chemical shifts of  $\text{CO}_{i-1}$ ,  $\text{N}_i$ , and  $\text{C}\alpha_i$ . Rows in the fourth table, representing a CANCA spectrum, contain chemical shifts of  $\text{C}\alpha_i$ ,  $\text{N}_i$ , and  $\text{C}\alpha_{i-1}$ . For SrtC, the NCACX, NCOCX, CONCA, and CANCA tables contained 119, 116, 110, and 116 signal rows, respectively. For MLKL, these tables contained 148, 148, 145, and 147 signal rows, respectively. The numbers of signal rows do not equal the numbers of residues in the protein sequences because certain chemical shift values are not available, especially for residues 1–6, 22, 38–41, 54, 65–74, 115, 122–124, 132, and 146 of SrtC and for residues 1–12, 52, 53, 65, 66, 93, 104, 106–108, 134, 135, 137, and 138 of MLKL.

Sidechain chemical shifts other than those of  $\beta$ -carbons are not included explicitly in the NCACX and NCOCX signal tables. In our experience, crosspeaks to sidechain carbons beyond  $\beta$ -carbons are often weak in experimental spectra, especially in the case of 3D NCOCX spectra, and do not contribute much information about connections between different 3D spectra. Sidechain chemical shifts beyond  $\text{C}\beta$  are more important for establishing residue-type assignments, which are included explicitly in the signal tables.

Chemical shift assignments from the BioMagResBank and signal tables for massign2b are contained in Supporting Information.

### Dependence on chemical shift uncertainties

First, the dependence of MCSA results on the assumed uncertainties in chemical shift values was examined. Uncertainties of 0.1, 0.2, 0.3, 0.4, 0.6, and 0.8 ppm were tested. For each condition, 10 independent MCSA runs were performed, with  $2 \times 10^8$  attempts in each run. Score coefficients  $w_G$ ,  $w_B$ ,  $w_E$ , and  $w_U$  had minimum values of 0.0, 2.0, 0.0, and 0.0, and maximum values of 10.0, 50.0, 2.0, and 2.0. Each run was executed in approximately 200 s, using one core of an Intel Xeon E5640 processor. After each run, the final values of  $n_G$ ,  $n_B$ ,  $n_E$ , and  $n_U$  were stored and the final assignments of NCACX signals were compared with the correct NCACX assignments.

Results for SrtC and MLKL are summarized in Tables 1 and 2, which report the number of successful runs and the maximum number of incorrect assignments of NCACX signals (*i.e.*, the largest number of NCACX signal rows that were assigned to the wrong residue) in the successful runs for each condition. Here, a successful run is defined as one for which the final value of  $n_B$  was zero and the final value of  $n_U$  equaled the total number of signal rows in all signal tables, *i.e.*, a run that produced assignments that were fully consistent with information in the input files (but were not necessarily correct). Runs that produced assignments with bad connections or did not use all available signals were considered unsuccessful. Table 1 shows that, for SrtC, all successful runs with uncertainties of 0.3 ppm or less produced the correct assignments. However, larger uncertainties produced assignments that were fully consistent with the input files, but nonetheless were not entirely correct. Table 2 shows that, for MLKL, uncertainties above 0.1 ppm produced successful runs with incorrect assignments.

These results highlight the complexity of the resonance assignment problem. Even with ideal data from four 3D spectra, including unambiguous residue-type assignments of all signals, it is possible to find assignments that include all available signals and are internally

consistent, but still contain errors, unless the chemical shift uncertainties are quite small. Moreover, proteins with different sequences and secondary structures can present significantly different assignment difficulties.

### Dependence on the number of 3D spectra

Next, the dependence of MCSA results on the number of available 3D spectra was examined, using either three (NCACX, NCOCX, CONCA) or two (NCACX, NCOCX) signal tables. As shown in Table 1, eliminating the CANCA data did not degrade the quality of the final NCACX assignments for SrtC, as no incorrect assignments from successful runs were observed with chemical shift uncertainties of 0.3 ppm or less. Additionally eliminating the CONCA data produced incorrect assignments with 0.3 ppm uncertainty. As shown in Table 2, eliminating the CANCA data also did not degrade the quality of final NCACX assignments for MLKL, but additionally eliminating the CONCA data led to incorrect assignments even with 0.1 ppm uncertainty in chemical shifts. These results suggest that information about sequential correlations of C $\alpha$  chemical shifts may not be important in the assignment process, at least when unambiguous residue-type assignments are available, while sequential correlations of CO and C $\alpha$  chemical shifts may be more valuable.

### Effect of residue-type ambiguity

The assumption of entirely unambiguous residue-type assignments is rather unrealistic. Although certain residues, such as Ala, Ser, Thr, and Gly, can be reliably distinguished from others by their characteristic C $\alpha$  and C $\beta$  chemical shifts, most other residue types can not be identified unambiguously without observing at least several sidechain chemical shifts. Crosspeaks to sidechain sites beyond C $\beta$  are often weak, absent, or difficult to identify in 3D spectra that are used for resonance assignments. Therefore, the effect of residue-type ambiguity on resonance assignments was examined next. Specifically, the NCACX signal tables were modified by replacing E and Q residue-type assignments with EQ residue-type assignments (meaning that Glu and Gln signals could be assigned to either Glu or Gln), H and W with HW, R and K with RK, D and N with DN, and F and Y with FY. Residue-type assignments in other signal tables were made more ambiguous, since sidechain signals are generally weaker in experimental NCOCX spectra and are absent from CONCA and CANCA spectra. Specifically, residue-type assignments were grouped as EQ, MKRHW, NDL, and FY in NCOCX signal tables and as NDLKMRHQWFYC and VPI in CONCA and CANCA signal tables. Somewhat surprisingly, introduction of these residue-type ambiguities did not significantly affect the success rates of MCSA runs or the accuracy of NCACX assignments for either SrtC or MLKL, as shown in Tables 1 and 2. Greater ambiguities must eventually lead to inaccurate assignments, but this was not tested.

### Effect of “invisible” segments

Finally, the effect of “invisible” segments was examined by eliminating signals from the grey-colored segments in Figure 2 from the signal tables. Essentially, signals from long stretches of irregular secondary structure, which might conceivably be attenuated by dynamic disorder in the solid state, were eliminated. Signals from  $\beta$ -strands in SrtC and  $\alpha$ -helices in MLKL were retained, along with signals from relatively short loops. For SrtC, 65 NCACX signal rows, 63 NCOCX signal rows, 62 CONCA signal rows, and 64 CANCA



signal rows were retained. For MLKL, the corresponding numbers were 124, 124, 123, and 124.

Although one might expect the introduction of “invisible” segments to impair the accuracy of the final resonance assignments (since bad connections could be avoided by creating additional gaps between segments with assigned signals), in fact the results in Tables 1 and 2 show no such effect for either SrtC or MLKL.

## Discussion

Several aspects of these resonance assignment simulations merit further discussion. First, it is clear that narrow NMR lines are a prerequisite for accurate assignments for proteins comprised of ~150 residues, even when high-quality 3D spectra can be obtained. Strictly speaking, the chemical shift uncertainties discussed above and indicated in Tables 1 and 2 are not the same as NMR linewidths, since the uncertainty can be significantly less than the linewidth if the signal-to-noise ratio is high. For example, a simple exercise in which noise is added to an ideal Gaussian lineshape and the result is fit with a Gaussian function indicates that the uncertainty in the peak position is less than one third of the full-width-at-half-maximum when the signal-to-noise (*i.e.*, the ratio of the true peak height to the root-mean-squared noise) is 5:1. Interestingly, the requirement on chemical shift uncertainties is markedly different for the two proteins considered in these simulations, with smaller uncertainties being required in the case of MLKL. This difference is attributable to the lower degree of spectral dispersion for MLKL, as illustrated by the simulated 2D spectra in Figure 3, which is in turn attributable to the high  $\alpha$ helix content of the MLKL structure.

Second, the simulations suggest that 3D CANCA spectra do not contain important assignment information that is complementary to the information in 3D NCACX, NCOCX, and CONCA spectra. This observation may have practical implications, since acquisition of a 3D CANCA spectrum can be time-consuming. Additional simulations on proteins with different sequences and secondary structures are required before the generality of this observation can be established. Simulations of this type, in which the numbers and properties of data sets are varied, can be a useful tool for designing efficient assignment strategies.

Third, although assignment simulations for SrtC and MLKL suggest that the presence of “invisible” segments does not increase the difficulty of determining unique resonance assignments, the situation may be different for other proteins. In particular, proteins with repetitive sequences are problematic, as it becomes difficult to assign signals uniquely to repeated sequence motifs if some of the repeats are “invisible” [20].

Fourth, complete assignment of all NMR signals is not necessarily a prerequisite for structural or dynamical studies. However, it is important to identify accurately the signals that do not have definite site-specific assignments. This identification can be made by comparing results of multiple MCSA runs.

Although only  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shifts are considered in the simulations described above, multidimensional solid state NMR spectra of proteins that include  $^1\text{H}$  chemical shifts

are also obtained in some experiments [46–48]. Simulations that include  $^1\text{H}$  chemical shifts can be carried out with the mcassign2b program, without any modifications, and could be used to design assignment strategies for such experiments.

Finally, as shown in Tables 1 and 2, the number of successful MCSA runs decreased when the chemical shift uncertainties were large. Apparently, the “score landscape” (analogous to a potential energy landscape in protein folding simulations [49]) becomes rougher and the MCSA algorithm is more likely to become “trapped” in local score maxima. Similarly, a manual approach to sequential assignments is more likely to reach a “dead end” when the NMR linewidths are large. A larger number of attempts in each MCSA run generally increases the number of successful runs.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases and by the Intramural AIDS Targeted Antiviral Program of the National Institutes of Health.

## References

1. Paravastu AK, Leapman RD, Yau WM, Tycko R. Molecular structural basis for polymorphism in Alzheimer's  $\beta$ -amyloid fibrils. *Proc Natl Acad Sci U S A*. 2008; 105:18349–18354. [PubMed: 19015532]
2. Sharpe S, Kessler N, Anglister JA, Yau WM, Tycko R. Solid state NMR yields structural constraints on the V3 loop from HIV-1 gp120 bound to the 447-52D antibody Fv fragment. *J Am Chem Soc*. 2004; 126:4979–4990. [PubMed: 15080704]
3. Hong M. Resonance assignment of  $^{13}\text{C}/^{15}\text{N}$  labeled solid proteins by two- and three-dimensional magic-angle-spinning NMR. *J Biomol NMR*. 1999; 15:1–14. [PubMed: 10549131]
4. Castellani F, van Rossum BJ, Diehl A, Rehbein K, Oschkinat H. Determination of solid state NMR structures of proteins by means of three-dimensional  $^{15}\text{N}$ - $^{13}\text{C}$ - $^{13}\text{C}$  dipolar correlation spectroscopy and chemical shift analysis. *Biochemistry*. 2003; 42:11476–11483. [PubMed: 14516199]
5. Petkova AT, Baldus M, Belenky M, Hong M, Griffin RG, Herzfeld J. Backbone and side chain assignment strategies for multiply labeled membrane peptides and proteins in the solid state. *J Magn Reson*. 2003; 160:1–12. [PubMed: 12565042]
6. Igumenova TI, Wand AJ, McDermott AE. Assignment of the backbone resonances for microcrystalline ubiquitin. *J Am Chem Soc*. 2004; 126:5323–5331. [PubMed: 15099118]
7. Siemer AB, Ritter C, Steinmetz MO, Ernst M, Riek R, Meier BH.  $^{13}\text{C}$ ,  $^{15}\text{N}$  resonance assignment of parts of the HET-s prion protein in its amyloid form. *J Biomol NMR*. 2006; 34:75–87. [PubMed: 16518695]
8. Chen L, Kaiser JM, Polenova T, Yang J, Rienstra CM, Mueller LJ. Backbone assignments in solid state proteins using J-based 3D heteronuclear correlation spectroscopy. *J Am Chem Soc*. 2007; 129:10650–10651. [PubMed: 17691789]
9. Franks WT, Kloepper KD, Wylie BJ, Rienstra CM. Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins. *J Biomol NMR*. 2007; 39:107–131. [PubMed: 17687624]
10. Goldbourn A, Gross BJ, Day LA, McDermott AE. Filamentous phage studied by magic-angle spinning NMR: Resonance assignment and secondary structure of the coat protein in Pfl. *J Am Chem Soc*. 2007; 129:2338–2344. [PubMed: 17279748]



11. Pintacuda G, Giraud N, Pierattelli R, Bockmann A, Bertini I, Emsley L. Solid state NMR spectroscopy of a paramagnetic protein: Assignment and study of human dimeric oxidized Cu(II)-Zn(II) superoxide dismutase (SOD). *Angew Chem-Int Edit.* 2007; 46:1079–1082.
12. Helmus JJ, Surewicz K, Nadaud PS, Surewicz WK, Jaroniec CP. Molecular conformation and dynamics of the Y145Stop variant of human prion protein. *Proc Natl Acad Sci U S A.* 2008; 105:6284–6289. [PubMed: 18436646]
13. Li Y, Berthold DA, Gennis RB, Rienstra CM. Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy. *Protein Sci.* 2008; 17:199–204. [PubMed: 18227427]
14. Higman VA, Flinders J, Hiller M, Jehle S, Markovic S, Fiedler S, van Rossum BJ, Oschkinat H. Assigning large proteins in the solid state: A MAS NMR resonance assignment strategy using selectively and extensively <sup>13</sup>C-labeled proteins. *J Biomol NMR.* 2009; 44:245–260. [PubMed: 19609683]
15. Shi L, Ahmed MAM, Zhang W, Whited G, Brown LS, Ladizhansky V. Three-dimensional solid state NMR study of a seven-helical integral membrane proton pump: Structural insights. *J Mol Biol.* 2009; 386:1078–1093. [PubMed: 19244620]
16. Sperling LJ, Berthold DA, Sasser TL, Jeisy-Scott V, Rienstra CM. Assignment strategies for large proteins by magic-angle spinning NMR: The 21-kDa disulfide-bond-forming enzyme DsbA. *J Mol Biol.* 2010; 399:268–282. [PubMed: 20394752]
17. Habenstein B, Wasmer C, Bousset L, Sourigues Y, Schuetz A, Loquet A, Meier BH, Melki R, Boeckmann A. Extensive de novo solid state NMR assignments of the 33 kDa C-terminal domain of the Ure2 prion. *J Biomol NMR.* 2011; 51:235–243. [PubMed: 21805376]
18. Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Atomic model of the type III secretion system needle. *Nature.* 2012; 486:276–279. [PubMed: 22699623]
19. Tycko R. Prospects for resonance assignments in multidimensional solid state NMR spectra of uniformly labeled proteins. *J Biomol NMR.* 1996; 8:239–251. [PubMed: 8953215]
20. Hu K-N, McGlinchey RP, Wickner RB, Tycko R. Segmental polymorphism in a functional amyloid. *Biophys J.* 2011; 101:2242–2250. [PubMed: 22067164]
21. Lu J-X, Qiang W, Yau W-M, Schwieters CD, Meredith SC, Tycko R. Molecular structure of  $\beta$ -amyloid fibrils in Alzheimer's disease brain tissue. *Cell.* 2013; 154:1257–1268. [PubMed: 24034249]
22. Bayro MJ, Chen B, Yau W-M, Tycko R. Site-specific structural variations accompanying tubular assembly of the HIV-1 capsid protein. *J Mol Biol.* 2014; 426:1109–1127. [PubMed: 24370930]
23. Tycko R, Hu KN. A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning. *J Magn Reson.* 2010; 205:304–314. [PubMed: 20547467]
24. Hu K-N, Qiang W, Tycko R. A general Monte Carlo/simulated annealing algorithm for resonance assignment in NMR of uniformly labeled biopolymers. *J Biomol NMR.* 2011; 50:267–276. [PubMed: 21710190]
25. Tycko R, Savtchenko R, Ostapchenko VG, Makarava N, Baskakov IV. The  $\alpha$ -helical C-terminal domain of full-length recombinant PrP converts to an in-register parallel  $\beta$ -sheet structure in PrP fibrils: Evidence from solid state nuclear magnetic resonance. *Biochemistry.* 2010; 49:9488–9497. [PubMed: 20925423]
26. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys.* 1953; 21:1087–1092.
27. Nelson SJ, Schneider DM, Wand AJ. Implementation of the main chain directed assignment strategy: Computer-assisted approach. *Biophys J.* 1991; 59:1113–1122. [PubMed: 1868156]
28. Bartels C, Guntert P, Billeter M, Wuthrich K. GARANT: A general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *J Comput Chem.* 1997; 18:139–149.
29. Buchler NEG, Zuiderweg ERP, Wang H, Goldstein RA. Protein heteronuclear NMR assignments using mean-field simulated annealing. *J Magn Reson.* 1997; 125:34–42. [PubMed: 9245358]

30. Li KB, Sanctuary BC. Automated resonance assignment of proteins using heteronuclear 3D NMR. 2. Side chain and sequence-specific assignment. *J Chem Inf Comput Sci*. 1997; 37:467–477. [PubMed: 9177001]
31. Lukin JA, Gove AP, Talukdar SN, Ho C. Automated probabilistic method for assigning backbone resonances of <sup>13</sup>C,<sup>15</sup>N-labeled proteins. *J Biomol NMR*. 1997; 9:151–166. [PubMed: 9090130]
32. Leutner M, Gschwind RM, Liermann J, Schwarz C, Gemmecker G, Kessler H. Automated backbone assignment of labeled proteins using the threshold accepting algorithm. *J Biomol NMR*. 1998; 11:31–43. [PubMed: 9615996]
33. Bailey-Kellogg C, Widge A, Kelley JJ, Berardi MJ, Bushweller JH, Donald BR. The NOESY jigsaw: Automated protein secondary structure and main-chain assignment from sparse, unassigned NMR data. *J Comput Biol*. 2000; 7:537–558. [PubMed: 11108478]
34. Moseley HNB, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Methods Enzymol*. 2001; 339:91–108.
35. Hitchens TK, Lukin JA, Zhan YP, McCallum SA, Rule GS. Monte: An automated monte carlo based approach to nuclear magnetic resonance assignment of proteins. *J Biomol NMR*. 2003; 25:1–9. [PubMed: 12566995]
36. Lemak A, Steren CA, Arrowsmith CH, Llinas M. Sequence specific resonance assignment via multicanonical Monte Carlo search using an ABACUS approach. *J Biomol NMR*. 2008; 41:29–41. [PubMed: 18458824]
37. Fredriksson J, Bermel W, Staykova DK, Billeter M. Automated protein backbone assignment using the projection-decomposition approach. *J Biomol NMR*. 2012; 54:43–51. [PubMed: 22806129]
38. Schmidt E, Guentert P. A new algorithm for reliable and general NMR resonance assignment. *J Am Chem Soc*. 2012; 134:12817–12829. [PubMed: 22794163]
39. Schmidt E, Gath J, Habenstein B, Ravotti F, Szekeley K, Huber M, Buchner L, Boeckmann A, Meier BH, Guentert P. Automated solid state NMR resonance assignment of protein microcrystals and amyloids. *J Biomol NMR*. 2013; 56:243–254. [PubMed: 23689812]
40. Yang Y, Fritzsche KJ, Hong M. Resonance assignment of the NMR spectra of disordered proteins using a multi-objective non-dominated sorting genetic algorithm. *J Biomol NMR*. 2013; 57:281–296. [PubMed: 24132778]
41. Abbas A, Guo X, Jing B-Y, Gao X. An automated framework for NMR resonance assignment through simultaneous slice picking and spin system forming. *J Biomol NMR*. 2014; 59:75–86. [PubMed: 24748536]
42. Nielsen JT, Kulminkaya N, Bjerring M, Nielsen NC. Automated robust and accurate assignment of protein resonances for solid state NMR. *J Biomol NMR*. 2014; 59:119–134. [PubMed: 24817190]
43. Moseley HNB, Sperling LJ, Rienstra CM. Automated protein resonance assignments of magic angle spinning solid state NMR spectra of  $\beta$ 1 immunoglobulin binding domain of protein G (GB1). *J Biomol NMR*. 2010; 48:123–128. [PubMed: 20931264]
44. Robson SA, Jacobitz AW, Phillips ML, Clubb RT. Solution structure of the sortase required for efficient production of infectious *Bacillus anthracis* spores. *Biochemistry*. 2012; 51:7953–7963. [PubMed: 22974341]
45. Su L, Quade B, Wang H, Sun L, Wang X, Rizo J. A plug release mechanism for membrane permeation by MLKL. *Structure*. 2014; 22:1489–1500. [PubMed: 25220470]
46. Barbet-Massin E, Pell AJ, Retel JS, Andreas LB, Jaudzems K, Franks WT, Nieuwkoop AJ, Hiller M, Higman V, Guerry P, Bertarello A, Knight MJ, Felletti M, Le Marchand T, Kotelovica S, Akopjana I, Tars K, Stoppini M, Bellotti V, Bolognesi M, Ricagno S, Chou JJ, Griffin RG, Oschkinat H, Lesage A, Emsley L, Herrmann T, Pintacuda G. Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J Am Chem Soc*. 2014; 136:12489–12497. [PubMed: 25102442]
47. Chevelkov V, Habenstein B, Loquet A, Giller K, Becker S, Lange A. Proton-detected MAS NMR experiments based on dipolar transfers for backbone assignment of highly deuterated proteins. *J Magn Reson*. 2014; 242:180–188. [PubMed: 24667274]

48. Xiang S, Chevelkov V, Becker S, Lange A. Towards automatic protein backbone assignment using proton-detected 4D solid state NMR data. *J Biomol NMR*. 2014; 60:85–90. [PubMed: 25193427]
49. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: The energy landscape perspective. *Ann Rev Phys Chem*. 1997; 48:545–600. [PubMed: 9348663]

Author Manuscript

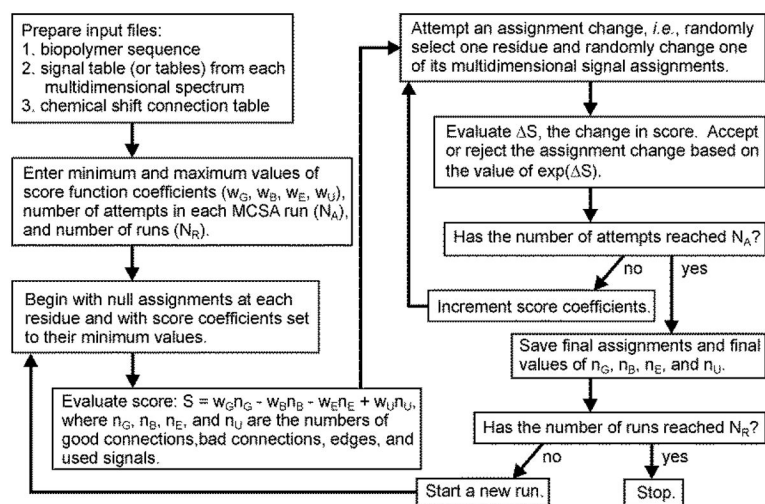
Author Manuscript

Author Manuscript

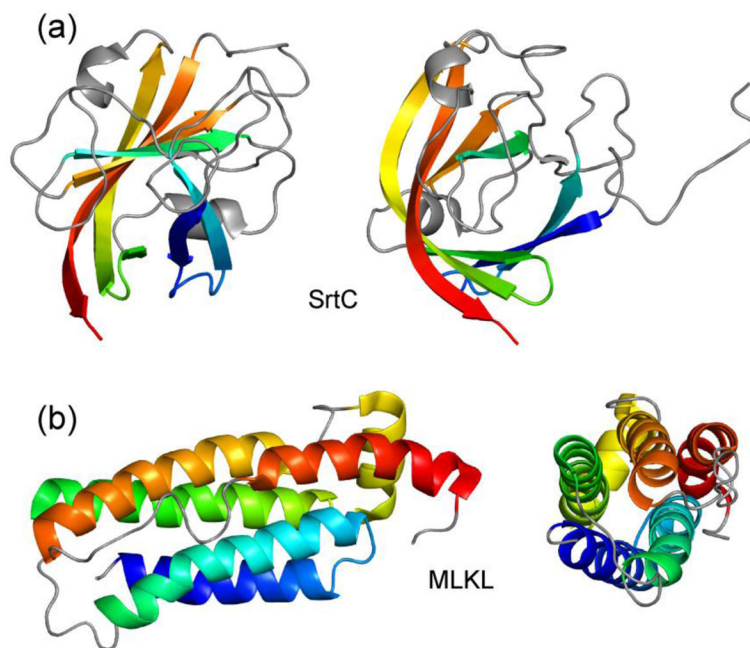
Author Manuscript

### Highlights

- Unique resonance assignments can be difficult to obtain
- The MCSA assignment algorithm can test for uniqueness
- Assignment simulations on artificial 3D data provide insights

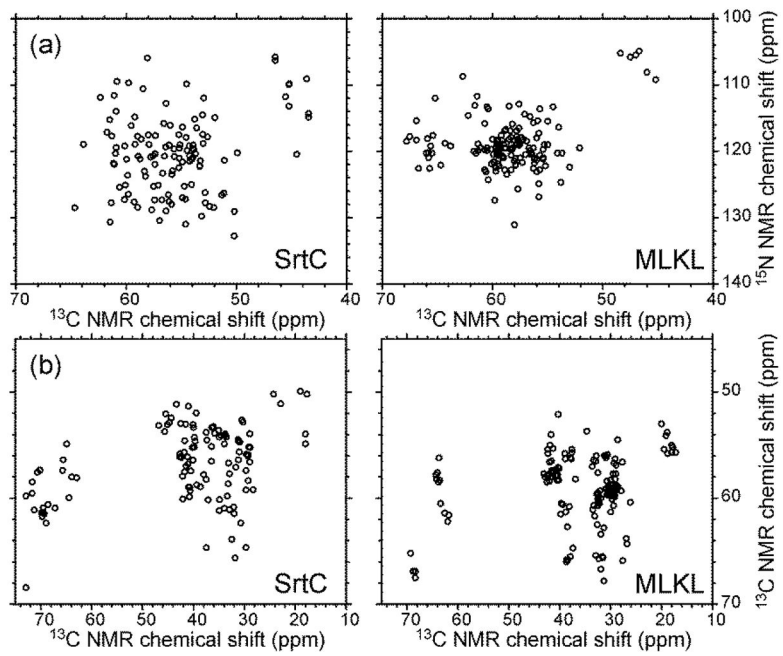


**Figure 1.** Flow chart for the Monte Carlo/simulated annealing algorithm for resonance assignment, as contained in the program mcassign2b.



**Figure 2.** Cartoon representations of the structures of *B. anthracis* SrtC catalytic domain [44] (a) and MLKL N-terminal domain [45] (b), as determined by solution NMR (PDB 2LN7 and PDB 2MSV, respectively). Chemical shift values for these two proteins (BMRB 18152 and BMRB 25135) were used to generate artificial 3D solid state NMR spectra for assignment simulations. Each protein is viewed from two directions, approximately 90° apart. Signals from segments in grey were taken to be “invisible” in certain simulations, as discussed in the text.





**Figure 3.** Simulated 2D N-C $\alpha$  (a) and C $\alpha$ -C $\beta$  (b) spectra of SrtC and MLKL.

**Table 1**  
 Summary of results from MCSA assignment simulations on SrtC catalytic domain (PDB 2LN7)

chemical shift uncertainties (ppm)	number of 3D spectra	ambiguous residue-type assignments?	“invisible” segments?	number of successful runs from 10 attempts	maximum number of incorrect NCA CX assignments in successful runs
0.1	4	no	no	10	0
0.2	4	no	no	10	0
0.3	4	no	no	10	0
0.4	4	no	no	7	2
0.6	4	no	no	9	6
0.8	4	no	no	3	10
0.1	3	no	no	10	0
0.2	3	no	no	10	0
0.3	3	no	no	10	0
0.4	3	no	no	10	2
0.6	3	no	no	10	8
0.8	3	no	no	9	13
0.1	2	no	no	10	0
0.2	2	no	no	10	0
0.3	2	no	no	10	2
0.4	2	no	no	10	4
0.6	2	no	no	10	12
0.8	2	no	no	10	28
0.1	4	yes	no	10	0
0.2	4	yes	no	10	0
0.3	4	yes	no	8	0
0.4	4	yes	no	7	2
0.6	4	yes	no	6	6
0.8	4	yes	no	7	12
0.1	4	no	yes	10	0
0.2	4	no	yes	10	0

chemical shift uncertainties (ppm)	number of 3D spectra	ambiguous residue-type assignments?	“invisible” segments?	number of successful runs from 10 attempts	maximum number of incorrect NCACX assignments in successful runs
0.3	4	no	yes	10	0
0.4	4	no	yes	10	0
0.6	4	no	yes	10	2
0.8	4	no	yes	8	4

**Table 2**  
Summary of results from MCSA assignment simulations on MLKL N-terminal domain (PDB 2MSV)

chemical shift uncertainties (ppm)	number of 3D spectra	ambiguous residue-type assignments?	“invisible” segments?	number of successful runs from 10 attempts	maximum number of incorrect NCACX assignments in successful runs
0.1	4	no	no	10	0
0.2	4	no	no	10	4
0.3	4	no	no	5	6
0.4	4	no	no	2	8
0.6	4	no	no	0	-
0.8	4	no	no	0	-
0.1	3	no	no	10	0
0.2	3	no	no	10	4
0.3	3	no	no	10	12
0.4	3	no	no	10	14
0.6	3	no	no	5	26
0.8	3	no	no	2	48
0.1	2	no	no	10	4
0.2	2	no	no	10	8
0.3	2	no	no	9	22
0.4	2	no	no	10	26
0.6	2	no	no	0	-
0.8	2	no	no	1	87
0.1	4	yes	no	10	0
0.2	4	yes	no	10	4
0.3	4	yes	no	7	12
0.4	4	yes	no	1	8
0.6	4	yes	no	0	-
0.8	4	yes	no	0	-
0.1	4	no	yes	10	0
0.2	4	no	yes	10	4

chemical shift uncertainties (ppm)	number of 3D spectra	ambiguous residue-type assignments?	“invisible” segments?	number of successful runs from 10 attempts	maximum number of incorrect NCACX assignments in successful runs
0.3	4	no	yes	5	12
0.4	4	no	yes	3	8
0.6	4	no	yes	0	-
0.8	4	no	yes	0	-