Published in final edited form as:

J Magn Reson. 2015 May ; 254: 93–97. doi:10.1016/j.jmr.2015.02.013.

Reliable Resonance Assignments of Selected Residues of Proteins with Known Structure Based on Empirical NMR Chemical Shift Prediction

Da-Wei Li^{1,3}, Dan Meng³, and Rafael Brüschweiler^{1,2,3,*}

¹Campus Chemical Instrumentation Center, The Ohio State University, Columbus, OH 43210

²Department of Chemistry & Biochemistry, The Ohio State University, Columbus, OH 43210

³Chemical Sciences Laboratory, Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32306

Abstract

A robust NMR resonance assignment method is introduced for proteins whose 3D structure has previously been determined by X-ray crystallography. The goal of the method is to obtain a subset of correct assignments from a parsimonious set of 3D NMR experiments of ¹⁵N, ¹³C labeled proteins. Chemical shifts of sequential residue pairs are predicted from static protein structures using PPM_One, which are then compared with the corresponding experimental shifts. Globally optimized weighted matching identifies the assignments that are robust with respect to small changes in NMR cross-peak positions. The method, termed PASSPORT, is demonstrated for 4 proteins with 100 – 250 amino acids using 3D NHCA and a 3D CBCA(CO)NH experiments as input producing correct assignments with high reliability for 22% of the residues. The method, which works best for Gly, Ala, Ser, and Thr residues, provides assignments that serve as anchor points for additional assignments by both manual and semi-automated methods or they can be directly used for further studies, e.g. on ligand binding, protein dynamics, or post-translational modification, such as phosphorylation.

Keywords

Protein structure based NMR resonance assignments; chemical shift prediction

© 2015 Published by Elsevier Inc.

Corresponding Author: bruschweiler.1@osu.edu Tel. +1 (614) 688-2083.

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.

Software and web server availability

The automatic assignment web server and an executable program for download are available at <http://spin.ccic.ohio-state.edu/index.php/assign>.

Introduction

Protein resonance assignment is a prerequisite for the biophysical investigation of proteins by NMR spectroscopy, such as studies of protein-protein and ligand-binding interactions, dynamics, or post-translational chemical modifications. Despite of many years of progress in the development of resonance assignment protocols, it often is still a time-consuming step, which typically requires the manual analysis of multiple 2D and 3D NMR experiments. For many proteins of interest their 3D structures have already been determined by X-ray crystallography and deposited in the Protein Databank (PDB).[1] For these protein systems, studies of their biophysical and biological properties by NMR can provide useful information about their function. In many such cases, the reliable assignment of a subset of suitably positioned residues can prove valuable. It is then desirable to identify such assignments by keeping the number of multidimensional NMR assignment experiments at a minimum allowing one to shift available resources toward applications.

The most common NMR assignment strategies of proteins are based on the sequential assignment method.[2, 3] This strategy uses the primary sequence information of a protein to derive assignments without requiring any 3D structural information. Automated versions of this assignment strategy have been implemented in a number of software packages.[4–9] This strategy is the method of choice when the objective of the study is the determination of the 3D protein structure itself. During the structure determination process it is common to assign more and more proton resonances whose distance constraints, in turn, help further improve the 3D structure[8, 10] or, if available, use 3D structural information from homologous proteins.[11]

For proteins whose 3D structure is already known from X-ray crystallography, the use of high-resolution 3D structural information can considerably facilitate the resonance assignment process. A general strategy proposed for this task used residual dipolar couplings (RDCs) in a single alignment medium along with database-derived average chemical shifts of the C α and C β nuclei of the different amino acid types.[12] The combinatorial assignment problem, whose complexity grows with N! where N is the number of residues, was represented in terms of a complete bipartite graph and the optimal global solution was obtained by using the Hungarian optimization method, which is efficient also for larger proteins. Other structure-based assignment strategies were developed specifically for paramagnetic proteins[13] or by primarily using ^1H - ^1H NOESY-derived distance information.[14–19] Because protons of sequential residues are in close spatial vicinity, NOESY experiments provide local connectivity information, in addition to long-range distance constraints. Recent work demonstrated for small to medium-large proteins (<160 amino acids), the power of a single 3D ^{13}C - or ^{15}N -edited NOESY experiment for simultaneous resonance assignment and structure determination. [8, 20]

Here, we present a method that produces reliable assignments for a subset of residues if the protein structure is known. The method finds the optimal global assignment based on predicted and experimental chemical shifts and it identifies residues whose assignments are most robust. The continuous growth of both the protein data bank (PDB)[1] and the BMRB chemical shift database[21] has led to the recent development of methods for the

increasingly accurate empirical prediction of chemical shifts of backbone nuclei of proteins with known 3D structures.[22–28] This assumes that the available X-ray crystallographic structure is representative for the NMR conditions, which can be in solution or in the solid state. The method presented here uses two complementary heteronuclear 3D experiments of a uniformly ^{15}N , ^{13}C -labeled protein that connect backbone resonances of sequential residues as input and it produces assignments based on global matching between the chemical shifts determined experimentally and ones predicted from the 3D structure. We demonstrate the method, which is termed PASSPORT for "Protein assignment strategy using chemical shift predictions for proteins with known structure", using just two triple-resonance 3D experiments as input.

Methods and Materials

Because the currently available chemical shift predictors have on average an accuracy that is lower than the chemical shift differences between NMR cross-peaks that are next to each other in crowded regions of a protein NMR spectrum, such as a 2D ^{15}N - ^1H HSQC spectrum, the reliable assignment based on a single spectrum is not feasible without the use of additional information. Here, we use local sequential connectivity information obtained from 3D CBCA(CO)NH[29] and 3D HNCA experiments[29–31] to generate clusters of resonances that belong to stretches of neighboring amino acids of variable lengths. These chemical shift clusters have a higher specificity than individual chemical shifts and thereby allow the better discrimination between correct and incorrect assignments.

In the 3D CBCA(CO)NH experiment a strip along the carbon dimension corresponding to the cross-section at the $^1\text{H}^{\text{N}}$ and ^{15}N chemical shifts of a given residue $i+1$ contains two carbon cross-peaks, which belong to the $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ nuclei of the preceding residue i . For the 3D HNCA experiment the analogous NH($i+1$) strip contains two peaks along the carbon dimension with one corresponding to the $^{13}\text{C}\alpha(i+1)$ resonance of residue $i+1$ and the other belonging to the $^{13}\text{C}\alpha(i)$ resonance of the preceding residue. In principle, superposition of the two spectra permits identification of the $\text{C}\alpha$ chemical shift of a NH group and the $\text{C}\alpha$, $\text{C}\beta$ chemical shifts of the preceding residue. The resulting five chemical shifts $\text{N}(i+1)$, $\text{HN}(i+1)$, $\text{C}\alpha(i+1)$, $\text{C}\alpha(i)$, $\text{C}\beta(i)$ form a group, which is called "sequential resonance cluster". From a known 3D structure, e.g. taken from the PDB, the chemical shifts of all sequential resonance clusters can be predicted and the root-mean-square difference (RMSD) of the chemical shifts (in units of ppm) between each predicted and each experimental sequential resonance cluster can be determined as follows:

$$C_{ij} = \sqrt{\frac{w_C(\delta_{C\alpha,i}^{\text{pred}} - \delta_{C\alpha,j}^{\text{exp}})^2 + w_C(\delta_{C\beta,i}^{\text{pred}} - \delta_{C\beta,j}^{\text{exp}})^2 + w_C(\delta_{C\alpha,i+1}^{\text{pred}} - \delta_{C\alpha,j+1}^{\text{exp}})^2 + w_N(\delta_{N,i+1}^{\text{pred}} - \delta_{N,j+1}^{\text{exp}})^2 + w_H(\delta_{H,i+1}^{\text{pred}} - \delta_{H,j+1}^{\text{exp}})^2}{5}} \quad \text{Eq. (1)}$$

Where δ^{pred} and δ^{exp} are the predicted and experimental chemical shifts, respectively. C_{ij} is a measure of the match between an experimental resonance cluster i and a putative cluster j . The lower C_{ij} the better, and thus more likely, is the match. Different types of nuclei are given different relative weights w_i in order to account for differential intrinsic chemical shift variations, whereby w_C , w_N and w_H were set to 1.0, 0.16, and 4.0 respectively. The assignment results are insensitive to small changes of the weights. They were initially

defined by the average prediction error for each type of nucleus and they were subsequently adjusted to optimize the assignment results. For most resonance clusters, all five terms are included. In those cases where some of the terms are not present, C_{ij} is rescaled correspondingly (by dividing in Eq. (1) by the actual number of terms). For example, because there is no $\delta_{C\beta}$ term for GLY residues, C_{ij} is rescaled by a factor 1.25. In case there is an obvious inconsistency between experimental and predicted chemical shifts, such as when GLY is compared with non-GLY residues, C_{ij} will be set to a very large value reflecting a very poor match and making this match prohibitively expensive, i.e. effectively forbidden, during weighted matching.

For well-resolved resonances, the HNCA experiment can provide sequential connectivity information for stretches of more than 2 residues. Whenever possible, we use sequential residue pairs that contain up to 2 additional residues in each direction and hence to the RMSD goes over resonance clusters that contain resonances from up to 5 residues. This further increases the specificity and accuracy of assignments. Since $C\alpha$ resonances tend to be crowded, there are often multiple possibilities to select the preceding and the following amino acid from the HNCA spectrum. In such cases, the combination is selected that gives the minimal RMSD with respect to a predicted resonance cluster. This implies that the combination selected for a given experimental cluster depends on the predicted cluster it is compared with.

Once the cost matrix C has been determined according to Eq. (1), the selection of the best assignment is based on weighted matching, which is a global optimization algorithm.[32] It identifies a unique one-to-one map between computed and experimental resonance clusters that minimizes the sum of all resonance cluster RMSDs. The global minimum is found efficiently by means of the Hungarian algorithm.[33] All chemical shifts were predicted using the PPM_One, Sparta+, [23] and Shiftx2 [27] chemical shift prediction programs. For protein parts that are missing in the X-ray crystal structure, we use the corresponding random coil chemical shifts, because these protein parts often adopt random coil conformations in solution. The random coil chemical shifts used are the chemical shift averages taken from the BMRB database.[21]

Results of our method show that it produces for proteins with 100 – 200 residues 50% correct assignments. Clearly, the proposed assignments are only of limited use, unless false positive assignments can be identified with high confidence. To identify and eliminate false positive assignments, we add random Gaussian distributed noise with a standard deviation of 0.6 ppm to all elements of the cost matrix of Eq. (1) before solving the weighted matching problem. This procedure is repeated 30 times and the resulting assignments are compared with each other. A confidence level is determined for each assignment according to the number of identical assignments obtained during the 30 iterations. For example, if a resonance cluster is assigned to a particular protein position 24 times out of the 30 times of tries, the confidence level of this assignment is 80%.

We tested the PASSPORT method for four proteins, which had been previously assigned in our lab using standard sequential assignment methods using half a dozen or more 2D and 3D NMR experiments (Table 1). Cross-peaks were picked using the standard restrained peak

picking facility of the Sparky software using the HN peak position obtained from a 2D HSQC experiment as input. Resulting peak lists from CBCA(CO)NH and HNCA data sets, the protein sequence and the chemical shift predictions from the corresponding PDB files serve as input for PASSPORT.

Results and Discussion

The PASSPORT method is demonstrated for the oncoprotein MDM2,[34] a 29.5 kDa construct consisting of the N- and P-domains of Copper ATPase, and the separate cytosolic calcium-binding domains 1 and 2 (CBD1 and CBD2) of the sodium-calcium exchanger NCX, which are listed in Table 1 along with some additional information.[35, 36] The performance of PASSPORT for the four proteins is summarized in Figure 1. Chemical shifts were predicted using PPM_One. The number of correct assignments (blue symbols) and number of false assignments (red symbols) are shown as a function of confidence level cutoff, i.e. any assignment whose confidence level is below the cutoff has been rejected. If the confidence level cutoff is at 70%, on average 22% of all amino acids can be assigned with good reliability for the four proteins. Of these assignments, 10% (or 2% of all amino acids) were wrongly assigned. At a confidence level cutoff of 90%, there is only one false positive for one protein while 10% of the residues were assigned. Application of an even stricter criterion (confidence level cutoff 100%) reduces the number of assignments to 5% of all residues with no false positives for all four proteins. Ribbon models of the four proteins are depicted in Figure 2 with the correctly assigned residues colored in red at a 70% confidence level cutoff.

While the assignments obtained in this manner can serve as anchor points for further resonance assignments using additional 2D or 3D NMR experiments, the subset of assignments can also be directly used for certain NMR applications. For many proteins whose structures have been determined by X-ray crystallography, protein-protein and protein-ligand binding studies performed by NMR can provide useful information about binding sites, binding modes, and affinities. In such cases, the assignment of a subset of strategically positioned residues can be sufficient to provide useful information. As an example, some of the assigned residues in the ATP-binding site of Copper ATPase show chemical shift changes during the titration of the ATP-analog AMPPNP. The titration trajectory from residue A531 (residue numbering according to PDB 2B8E) that is in fast exchange is shown in Figure 3. This residue was assigned correctly assigned by PASSPORT in this 29.5 kDa protein illustrating the use of the assignment strategy for the targeting of specific residues.

The performance of the PASSPORT method depends critically on the accuracy of the chemical shift prediction. Inspection of resonance clusters that could not be reliably assigned shows that there is usually one or several other sequential residue pairs along the protein sequence that have comparable chemical shift RMSDs. For the four proteins studied here, PPM_One, Sparta+ and Shiftx2 show comparable performance for PASSPORT. They also have the tendency to yield similar sets of correctly and wrongly assigned resonance sets.

Other factors that limit the performance of PASSPORT are missing cross-peaks, additional peaks from side chains, and false peaks introduced by noise, which all reduce the specificity and hence increase the uncertainty in the chemical shift RMSD between a resonance cluster and a protein segment according to Eq. (1). Another important challenge is cross-peak overlap. When two resonance clusters have almost identical N and H^N chemical shifts, there are generally 4 peaks (2 $C\alpha$ and 2 $C\beta$ peaks) along the carbon dimension, which renders the pairing of the $C\alpha$ with the $C\beta$ peaks ambiguous. Moreover, in such cases the correct identification of the intraresidual $C\alpha(i)$ and sequential $C\alpha(i-1)$ and $C\beta(i-1)$ chemical shifts by the superposition of the CBCA(CO)NH and HNCA spectra is challenging too. The identification of sequential connectivity via the HNCA experiment relies on the matching of two $C\alpha$ chemical shifts, which becomes ambiguous when $C\alpha$ chemical shifts are degenerate and has a negative impact on the assignment performance. In fact, if these ambiguities could be all resolved so that only correct “sequential resonance clusters” can be provided to the global matching algorithm, nearly 100% coverage and 100% true positive rate is achieved for many small to medium sized proteins.

Very recently, a closely related strategy, termed RASP, for the assignment for proteins with known structure was introduced.[19] The RASP strategy also relies on the matching of experimental and predicted resonance clusters or spin systems using a slightly different combination of NMR experiments and a different optimization algorithm. The latter relies on minimizing the difference between experimental and predicted chemical shifts by working with stretches of three sequential residues. A very true positive rate was reported when the algorithm was provided the correct spin-system information based on the known assignment.[19] The spin-system information is critical for the performance of RASP, but hard to obtain from a small set of multidimensional NMR spectra with finite resolution. Therefore, RASP requires the input of an experienced spectroscopist who identifies spin systems in the presence of spectral overlap. This process is repeated iteratively to achieve the best assignment result.[19] By contrast, our approach directly accepts raw peak lists, such as the ones obtained using Sparky, rather than complete spin systems without any human intervention and it is very tolerant to false peaks, missing peaks, and peak overlap. On the down side, the total number of assignments drops significantly.

Among the 20 types of amino acids, Gly, Ala, Ser, and Thr (GAST) have the highest probability to be assigned correctly. In fact, about half of these residues can be assigned correctly using a cutoff at 70% confidence level for the four proteins studied here. In Figure 4, the mean and standard deviation of the carbon chemical shifts of all 20 amino acids, extracted from BMRB database, are depicted. For Gly, $C\alpha$ chemical shifts are around 45 ppm, which populates a region that generally is less crowded. For Ala, $C\beta$ chemical shifts are around 20 ppm, which is significantly below the $C\alpha$ and $C\beta$ chemical shifts of all other residues. For Ser and Thr, $C\beta$ chemical shifts usually exceed their corresponding $C\alpha$ chemical shifts. Therefore, all 4 residue types are unique, populating some of the less crowded regions in chemical shift space, which makes the matching between experiment and prediction less ambiguous. Therefore, the PASSPORT method is most directly useful, if an amino-acid residue of interest is of the GAST type. The unique chemical shift properties of the GAST residues have been recognized previously [37]. For example, based on these properties pulse sequences have been designed that identify specifically these residues in

direct ^{13}C detected experiments (Chakraborty et al. 2012). By contrast, the current approach relies on standard 3D NMR pulse sequences that are part of most assignment datasets. Therefore, at a fraction ($\sim 1/3$) of the regular experimental time, reliable assignments of these and other residues can be obtained. Additional 2D and 3D NMR experiments can be collected if a more complete set of assignments is required for subsequent applications. In this way, PASSPORT can also be combined with existing (semi-)automated assignment protocols and an increasing number of input spectra, which helps reduce the ambiguity in resonance clusters and increases the number of assignments.

In summary, we presented the globally optimized PASSPORT backbone assignment strategy for proteins with known 3D structures. It can efficiently identify assignments that are correct with very high probability. For the 4 proteins used in this study, which range between 12.6 and 29.5 kDa, the method is able to robustly assign 20% of the residues with high confidence using raw cross-peak lists of two complementary 3D NHCA and a 3D CBCA(CO)NH spectra. These assignments can be used as anchor points for additional assignments or they can directly serve as sources of information on protein parts that are of biological interest.

Acknowledgement

This work was supported by the National Institutes of Health (grant R01 GM 066041) and the National Science Foundation (grant NSF MCB 1330150). R.B. is an Ohio Research Scholar.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids. Res.* 2000; 28:235–242. [PubMed: 10592235]
2. Wu, K. *thrich, NMR of proteins and nucleic acids.* New York: Wiley; 1986.
3. Cavanagh, J. *Protein NMR spectroscopy : principles and practice.* 2nd ed.,. Amsterdam ; Boston: Academic Press; 2007.
4. Zimmerman DE, Kulikowski CA, Huang YP, Feng WQ, Tashiro M, Shimotakahara S, Chien CY, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. *Mol. Biol. J.* 1997; 269:592–610.
5. Jung YS, Zweckstetter M. Backbone assignment of proteins with known structure using residual dipolar couplings. *J. Biomol. NMR.* 2004; 30:25–35. [PubMed: 15452432]
6. Jung YS, Zweckstetter M. Mars - robust automatic backbone assignment of proteins. *J. Biomol. NMR.* 2004; 30:11–23. [PubMed: 15452431]
7. Bahrami A, Assadi AH, Markley JL, Eghbalnia HR. Probabilistic Interaction Network of Evidence Algorithm and its Application to Complete Labeling of Peak Lists from Protein NMR Spectroscopy. *PLoS Comput. Biol.* 2009; 5
8. Schmidt E, Güntert P. A New Algorithm for Reliable and General NMR Resonance Assignment. *J. Am. Chem. Soc.* 2012; 134:12817–12829. [PubMed: 22794163]
9. 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]
10. Guerry P, Herrmann T. Advances in automated NMR protein structure determination. *Q. Rev. Biophys.* 2011; 44:257–309. [PubMed: 21411039]
11. Bartels C, Billeter M, Güntert P, Wuthrich K. Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *J. Biomol. NMR.* 1996; 7:207–213. [PubMed: 22911044]

12. Hus JC, Prompers JJ, Brüschweiler R. Assignment strategy for proteins with known structure. *J. Magn. Reson.* 2002; 157:119–123. [PubMed: 12202140]
13. Pintacuda G, Keniry MA, Huber T, Park AY, Dixon NE, Otting G. Fast structure-based assignment of ^{15}N HSQC spectra of selectively N-15-labeled paramagnetic proteins. *J. Am. Chem. Soc.* 2004; 126:2963–2970. [PubMed: 14995214]
14. Xiong F, Bailey-Kellogg C, Yang JY. A hierarchical grow-and-match algorithm for backbone resonance assignments given 3D structure. *Ieee, New York.* 2007
15. Apaydin MS, Conitzer V, Donald BR. Structure-based protein NMR assignments using native structural ensembles. *J. Biomol. NMR.* 2008; 40:263–276. [PubMed: 18365752]
16. Xiong F, Pandurangan G, Bailey-Kellogg C. Contact replacement for NMR resonance assignment. *Bioinformatics.* 2008; 24:I205–I213. [PubMed: 18586716]
17. Stratmann D, Guittet E, van Heijenoort C. Robust structure-based resonance assignment for functional protein studies by NMR. *J. Biomol. NMR.* 2010; 46:157–173. [PubMed: 20024602]
18. Jang R, Gao X, Li M. Towards Fully Automated Structure-Based NMR Resonance Assignment of ^{15}N -Labeled Proteins From Automatically Picked Peaks. *J. Comput. Biol.* 2011; 18:347–363. [PubMed: 21385039]
19. MacRaid CA, Norton RS. RASP: rapid and robust backbone chemical shift assignments from protein structure. *J. Biomol. NMR.* 2014; 58:155–163. [PubMed: 24445369]
20. Schmidt E, Güntert P. Reliability of exclusively NOESY-based automated resonance assignment and structure determination of proteins. *J. Biomol. NMR.* 2013; 57:193–204. [PubMed: 24036635]
21. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL. BioMagResBank. *Nucleic Acids Res.* 2008; 36:D402–D408. [PubMed: 17984079]
22. Shen Y, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR.* 2007; 38:289–302. [PubMed: 17610132]
23. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR.* 2010; 48:13–22. [PubMed: 20628786]
24. Xu XP, Case DA. Automated prediction of ^{15}N , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and $^{13}\text{C}'$ chemical shifts in proteins using a density functional database. *J Biomol. NMR.* 2001; 21:321–333. [PubMed: 11824752]
25. Xu XP, Case DA. Probing multiple effects on ^{15}N , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and $^{13}\text{C}'$ chemical shifts in peptides using density functional theory. *Biopolymers.* 2002; 65:408–423. [PubMed: 12434429]
26. Neal S, Nip AM, Zhang HY, Wishart DS. Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. *J Biomol NMR.* 2003; 26:215–240. [PubMed: 12766419]
27. Han B, Liu YF, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR.* 2011; 50:43–57. [PubMed: 21448735]
28. Li DW, Brüschweiler R. PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *J.Biomol. NMR.* 2012; 54:257–265. [PubMed: 22972619]
29. Grzesiek S, Bax A. Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. *J.Am. Chem. Soc.* 1992; 114:6291–6293.
30. Kay LE, Ikura M, Tschudin R, Bax A. 3-Dimensional Triple-Resonance Nmr- Spectroscopy of Isotopically Enriched Proteins. *J Magn Reson.* 1990; 89:496–514.
31. Farmer BT, Venters RA, Spicer LD, Wittekind MG, Muller L. A Refocused and Optimized Hnca - Increased Sensitivity and Resolution in Large Macromolecules. *J. Biomol. NMR.* 1992; 2:195–202. [PubMed: 1422152]
32. Horst, R.; Hoang, T. Global optimization : deterministic approaches. 3rd ed.. Berlin ; New York: Springer; 1996.
33. Munkres J. Algorithms for the Assignment and Transportation Problems. *J Soc Ind Appl Math.* 1957; 5:32–38.
34. Showalter SA, Bruschiweiler-Li L, Johnson E, Zhang F, Brüschweiler R. Quantitative lid dynamics of MDM2 reveals differential ligand binding modes of the p53- binding cleft. *J. Am. Chem. Soc.* 2008; 130:6472–6478. [PubMed: 18435534]

35. Johnson E, Bruschweiler-Li L, Showalter SA, Vuister GW, Zhang F, Bruschweiler R. Structure and dynamics of Ca^{2+} -binding domain 1 of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger in the presence and in the absence of Ca^{2+} . *J. Mol. Biol.* 2008; 377:945–955. [PubMed: 18280495]
36. Salinas RK, Bruschweiler-Li L, Johnson E, Bruschweiler R. Ca^{2+} Binding Alters the Interdomain Flexibility between the Two Cytoplasmic Calcium-binding Domains in the $\text{Na}^+/\text{Ca}^{2+}$ Exchanger. *J. Biol. Chem.* 2011; 286:32123–32131. [PubMed: 21778234]
37. Grzesiek S, Bax A. Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly $^{13}\text{C}/^{15}\text{N}$ -Enriched Proteins. *J. Biomol. NMR.* 1993; 3:185–204. [PubMed: 8477186]

Highlights

- Structure-based resonance assignment using protein NMR chemical shift predictor speeds up assignment process of key residues
- Requires minimal set of 2D and 3D NMR datasets for rapid assignment of key residues

Figure 1.

Assignment performance of PASSPORT as a function of confidence level cutoff. (A) Two-domain protein consisting of the N- and P-domains of copper ATPase, (B) oncoprotein MDM2, and cytosolic calcium-binding domains (C) CBD1 and (D) CBD2 of the sodium-calcium exchanger NCX. The blue, red, and gray areas represent the number of correctly assigned, misassigned, and unassigned residues, respectively.

Figure 2.

Ribbon representations of (A) construct existing of the N- and P-domains of copper ATPase, (B) the oncoprotein MDM2, the cytosolic calcium-binding domains (C) CBD1 and (D) CBD2 of the sodium-calcium exchanger NCX. Correctly assigned residues are colored in red. Residue A531 of Cu-ATPase, which shows chemical shift perturbation upon binding of AMPPNP is colored green.

Figure 3. Chemical shift trajectory of residue A531 in HSQC-TROSY of Cu-ATPase upon titration with ATP-analogue AMPPNP, which was correctly assigned using the PASSPORT algorithm.

Figure 4. Carbon chemical shift statistics (mean and standard deviation) of the 20 amino acid types from BMRB database protein chemical shifts.

Table 1

Proteins used in this study to test the PASSPORT assignment method.

Protein Name	PDB code ¹	Protein Length ²	Molecular weight (kDa)	Manual assignment percentage
Copper ATPase	2B8E	281 (246)	29.5	88%
MDM2	1YCR	111 (98)	12.6	100%
CBD1	3US9	139 (124)	14.9	87%
CBD2	3US9	156 (131)	17.6	81%

¹ PDB codes of X-ray crystal structures.

² Values in parentheses indicate the length of the protein part that has observable electron density in the crystal structure.