Journal of the Korean Magnetic Resonance Society **2014**, *18*, 36-40
DOI 10.6564/JKMRS.2014.18.1.036

Enhanced Chemical Shift Analysis for Secondary Structure prediction of protein

Won-Je Kim[2ǂ], Jin-Kyu Rhee[3ǂ], Jong-Jae Yi[1], Bong-Jin Lee[2], and Woo Sung Son[1*]

[1]College of pharmacy, CHA University, 120 Haeyong-ro, Pocheon-si, Gyeonggi-do, 487-010, Republic of Korea
[2]College of pharmacy, Seoul National University, San 56-1, Shillim-dong, Gwanak-gu, 151-742, Republic of Korea
[3]Korea Basic Science Institute Western Seoul Center, University-Industry Cooperation Building, 150 Bugahyeon-ro, Seodamun-gu, 120-140, Republic of Korea

Received May 10, 2014; Revised May 26, 2014; Accepted June 10, 2014

Abstract Predicting secondary structure of protein through assigned backbone chemical shifts has been used widely because of its convenience and flexibility. In spite of its usefulness, chemical shift based analysis has some defects including isotopic shifts and solvent interaction. Here, it is shown that corrected chemical shift analysis for secondary structure of protein. It is included chemical shift correction through consideration of deuterium isotopic effect and calculate chemical shift index using probability-based methods. Enhanced method was applied successfully to one of the proteins from Mycobacterium tuberculosis. It is suggested that correction of chemical shift analysis could increase accuracy of secondary structure prediction of protein and small molecule in solution.

Keywords Chemical Shift, Deuterium Isotope Effect, PSSI, Secondary Structure, NMR, Pyrazinamidase

Introduction

Elucidation of three-dimensional structure of protein using nuclear magnetic resonance (NMR) spectroscopy is much laborious and time consuming work. In this regard, secondary structure analysis based on chemical shifts data could be very useful to obtain structural information of proteins and organic molecules. Usually, chemical shifts are good indicator of molecular structure [1-4], and can give discrete information from chemical functional groups in different local and global environments. Among the application of these chemical shifts data, chemical shift index (CSI) has been widely used to identify the type of secondary structure such as helices, beta-strands, and random coil regions [5-6].

In protein NMR, isotope labeling technique with multi-dimensional pulse techniques is most often used to get high-resolution spectra including ^{13}C, ^{15}N, and ^{2}H. Especially, deuterium labeling is very useful to overcome low signal-to-noise ratio of larger proteins (MW > 20 kDa) [7]. In larger protein, the relaxation rates increase which lead to lower signal-to-noise because the overall tumbling is very slow in solution state. Through deuteration, the decreased protein density can eliminate some relaxation pathways, and the signal-to-ratio of NMR spectra from the deuterated protein is dramatically improved.

Despite advantage of deuterium labeling, the factors

* Address correspondence to: **Woo Sung Son,** College of Pharmacy, CHA University, 120 Haeryong-ro, Pocheon-si, Gyeonggi-do, 487-010, Republic of Korea, Tel: +82-31-850-9398; Fax: +82-31-850-9398; E-mail: wson@cha.ac.kr
ǂ These authors contributed equally.

affecting chemical shift should be considered because change of chemical shift can give effect on predicting secondary structure of protein using CSI [8]. To increase accuracy of secondary structure prediction by CSI, deuterium isotope effect should be taken into account during prediction. Also, advanced computational approach could make improvement in the accuracy and robustness of the results of secondary structure prediction. Here, it is showed that enhanced method could make improvement on secondary structure predicting by CSI through correcting chemical shifts deviation from deuterium isotope effect and harboring joint probability-based protocol. These methods were applied to pyrazinamidase (PncA) protein from *Mycobacterium tuberculosis* [9].

Experimental Methods

Protein Preparation- The pncA gene from *Mycobacterium tuberculosis* (http://tbgenomics.org, ORF No. Rv2043c) encoding residues 1-186 was cloned into a pET-22b(+) expression vector (EMD Bioscience, Inc., San Diego, CA, USA) to produce a final construct encoding the PncA protein with a C-terminal His-tag. The plasmid was transformed into an *E. coli* strain BL21-CodonPlus (DE3)-RIL (Stratagene, Inc., La Jolla, CA, USA) for protein expression. The final protein product contains 186 residues and additional LEHHHHHH polypeptide sequences at its C-termini.

The cell cultures were grown at 37°C in M9 minimal media and protein expression was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich, Inc., St. Louis, MO, USA) at an optical density (600 nm) of 0.5, followed by 48 h of additional growth at 15°C.

Global labelling of the protein ([U-^{15}N, [U-^{13}C, U-^{15}N]) with ^{13}C and ^{15}N was accomplished by including 99% $^{13}C_6$-Dglucose (2 g/liter) and 98% $^{15}NH_4Cl$ (1 g/liter) as the sole carbon and nitrogen sources (Cambridge Isotope Laboratory, Inc., Andover, MA, USA), respectively. 70% random fractionally deuterated sample was prepared with 99.9% D_2O (Cambridge Isotope Laboratory, Inc.).

PncA protein was purified using immobilized metal affinity chromatography on His-bind resin (Chelating Sepharose^{TM} Fast Flow, GE Healthcare Bio-Sciences, Corp., Piscataway, NJ, USA), anion exchange chromatography (HiTrap DEAE Sepharose^{TM} Fast Flow, GE Healthcare Bio-Sciences) followed by size-exclusion chromatography (Sephadex 75TM, GE Healthcare Bio-Sciences). All samples were buffer exchanged into 20 mM sodium phosphate, pH 6.7, with 10 mM dithiothreitol (Sigma-Aldrich, Inc.), 5% (v/v) D_2O (99.9% Deuterium Oxide, Sigma-Aldrich, Inc.) and 1% (v/v) sodium Azide (Sigma-Aldrich, Inc.). Samples contained from 0.3 to 0.6 mM PncA in volumes ranging from 300 to 380 μl.

Nuclear magnetic resonance (NMR)- NMR experiments were conducted at 313K using Bruker DRX 500, AVANCE 600 (with cryoprobe) and AVANCE 800 (with cryoprobe, Bruker BioSpin GmbH, Karlsruhe, Germany) MHz spectrometers. The following experiments were recorded for backbone assignment: 2D [^{15}N, ^{1}H]-HSQC, 2D [^{15}N, ^{1}H]-TROSY, 3D TROSY-HNCA (H^N_i, N_i, C^α_i, $\{C^\alpha_{i-1}\}$), 3D TROSY-HN(CO)CA (H^N_i, N_i, C^α_{i-1}), 3D TROSY-HNCACB (H^N_i, N_i, C^α_i, $\{C^\alpha_{i-1}\}$, C^β_i, $\{C^\beta_{i-1}\}$), 3D TROSY-HN(CO)CACB (H^N_i, N_i, C^α_{i-1}, C^β_{i-1}), 3D TROSY-HN(CA)CO (H^N_i, N_i, C'_i, $\{C'_{i-1}\}$) and 3D TROSY-HNCO (H^N_i, N_i, C'_{i-1}). ^{1}H chemical shifts were referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and ^{15}N and ^{13}C chemical shifts were referenced indirectly by multiplying the ^{1}H carrier frequency by 0.101329118 and 0.25144953, respectively. Spectral data were processed using NMRPipe [10] and analyzed in NMRView [11]. Sequential backbone assignments were performed with multi-dimensional NMR spectra listed above and 15N and 1H chemical shifts were referenced to corresponding peaks on NOESY spectra during assignment.

Chemical Shift Index (CSI)- Secondary structure analysis of PncA was performed using the $^{13}C^\alpha$ chemical shift index (CSI) method [12-13]. The random coil $^{13}C^\alpha$ values used in this method were modified to account for deuterium isotope effects on ^{13}C nuclei [8]. For C^α carbons, upfield isotope shifts of 0.50 ppm are expected, based on several comparisons of $^{13}C^\alpha$ shifts in protonated and deuterated forms of the same protein [8, 14-17].

TROSY effect was not compensated and for more accurate secondary structure prediction, neighboring residue effects [18] and secondary structural effects were considered [19] and probability based protein secondary structure identification (PSSI) method was used [20].

TALOS prediction- To identify and confirm the secondary structure of protein, fitting backbone

dihedral angle was performed by TALOS algorithm in NMRPipe package [2].

Results and Discussion

Table 1. Total deuterium isotope shifts for C^{α} and C^{β} nuclei. Statistical analysis indicates that the total deuterium isotope effect can be accurately predicted for most C^{α} and C^{β} nuclei in a perdeuterated protein.

$^{1}\Delta C(D)$ (one bond) = −0.29 ± 0.05
$^{2}\Delta C(D)$ (two bond) = −0.13 ± 0.02
$^{3}\Delta C(D)$ (three bond) = −0.07 ± 0.02
$^{1}\Delta C_{gly}(D)$ = −0.39 ± 0.04

	ΔC^{α}	ΔC^{β}	70% ΔC^{α}	70% ΔC^{β}
Asn, Asp, Ser, His, Phe, Trp, Tyr, Cys	−0.55	−0.71	−0.385	−0.497
Lys, Arg, Pro	−0.69	−1.11	−0.483	−0.777
Gln, Glu, Met	−0.69	−0.97	−0.483	−0.679
Ala	−0.68	−1.00	−0.476	−0.700
Ile	−0.77	−1.28	−0.539	−0.896
Leu	−0.62	−1.26	−0.434	−0.882
Thr	−0.63	−0.81	−0.441	−0.567
Val	−0.84	−1.20	−0.588	−0.840

Backbone chemical shifts are predictably affected by the secondary structure of the protein. The one protocol uses only C^{α} and C^{β} chemical shifts measured either directly from a protonated protein followed by a correction for the total deuterium isotope effect. Deuterium isotope shifts are significant for ^{15}N and ^{13}C nuclei located within at least three bonds of the site of deuteration [8]. The second commonly utilized method for secondary structure determination from chemical-shift analysis is the CSI protocol. Except deuterium isotope effect, considered other deviated factors (TROSY effect, neighboring effect), validated CSI protocol was used for determining secondary structure.

TALOS is a database system for empirical prediction of phi and psi backbone torsion angles using a combination of five kinds (HA, CA, CB, CO, N) of chemical shift assignments for a given protein sequence. The TALOS approach is an extension of the well-known observation that many kinds of secondary chemical shifts (i.e. differences between chemical shifts and their corresponding random coil values) are highly correlated with aspects of protein secondary structure. The goal of TALOS is to use secondary shift and sequence information in order to make quantitative predictions for the protein backbone angles phi and psi, and to provide a measure of the uncertainties in these predictions. For deuterium isotope effect correction, modified chemical shift was used for TALOS angel prediction. 70% value of total deuterium isotope shifts for C^{α} and C^{β} nuclei was added since 70% random fractionally deuterated samples were used for backbone resonance assignment [16].

The relationship has been well established between NMR chemical shifts and secondary structure in proteins. The C^{α} and carbonyl carbon atoms experience an upfield shift in extended structures, e.g. β–strand, and a downfield shift in helical structures. Both the C^{β} carbon and the H^{α} proton chemical shifts exhibit the opposite correlation, which means a downfield shift in β–strands and an upfield shift in helices.

These shifts are sufficiently consistent to permit the prediction of secondary structural elements for number of proteins. Using these chemical shifts, the regions of helical and β-strand secondary structure in

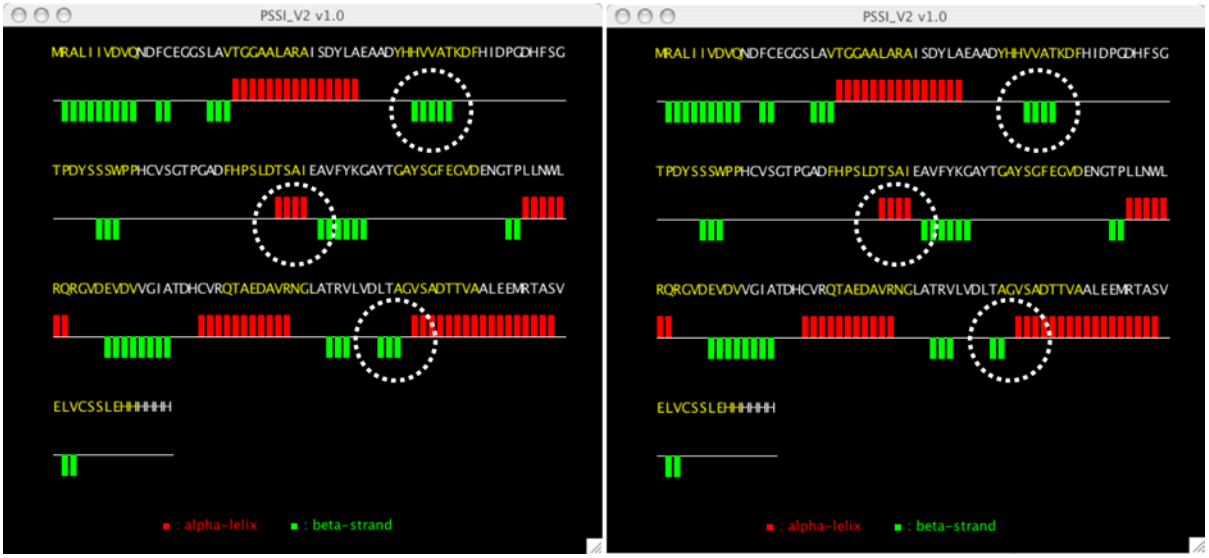

Figure 1. Secondary structure of PncA from PSSI-CSI. Calibrated CSI (left) and uncalibrated CSI (right) were plotted with and without PSSI. The change of CSI by PSSI was marked with dotted circle at each panel.

PncA.

Deuterated protein samples adopt substantially upfield shifted ^{13}C chemical shifts in NMR measurements. To convert $^{13}C(D)$ shifts to approximate $^{13}C(H)$ shifts, the total ^{2}H isotope effect, $\Delta C(D)$, must be considered for each carbon atom in a perdeuterated protein. Because about 70% random fractionally deuterated PncA sample was used for backbone resonance assignments, total deuterium isotope shifts for C^{α} and C^{β} nuclei was applied with 0.7 times values of total deuterium isotope shifts in perdeuterated protein which was previously reported (Table 1)[16]. For predicting secondary structure of PncA, chemical shift difference method between measured values and random-coil values using C^{α}, C^{β}, and ($\Delta C^{\alpha} - \Delta C^{\beta}$)smoothed which was smoothed with a three-point binomial function [21-22], and PSSI based CSI protocol was used (Figure 1).

Backbone dihedral angles (ψ) predicted using TALOS methods from chemical shifts and extracted from modeled three-dimensional structure were also applied for verifying prediction results of secondary structure (Figure 2). Sequential and inter-strand backbone NOE connectivity was similar to secondary structure pattern from chemical shift. From the PSSI-CSI prediction, changes of secondary structure prediction were found at several residues (Figure 1).

In conclusion, it is shown that correction of chemical shift deviation from deuterium isotope effect could make effect on prediction secondary structure of protein using CSI and PSSI protocol could increase the accuracy of prediction. Also, it is suggested that these enhanced methods could be used to increase overall accuracy of the three-dimensional structure calculated with chemical shift data.

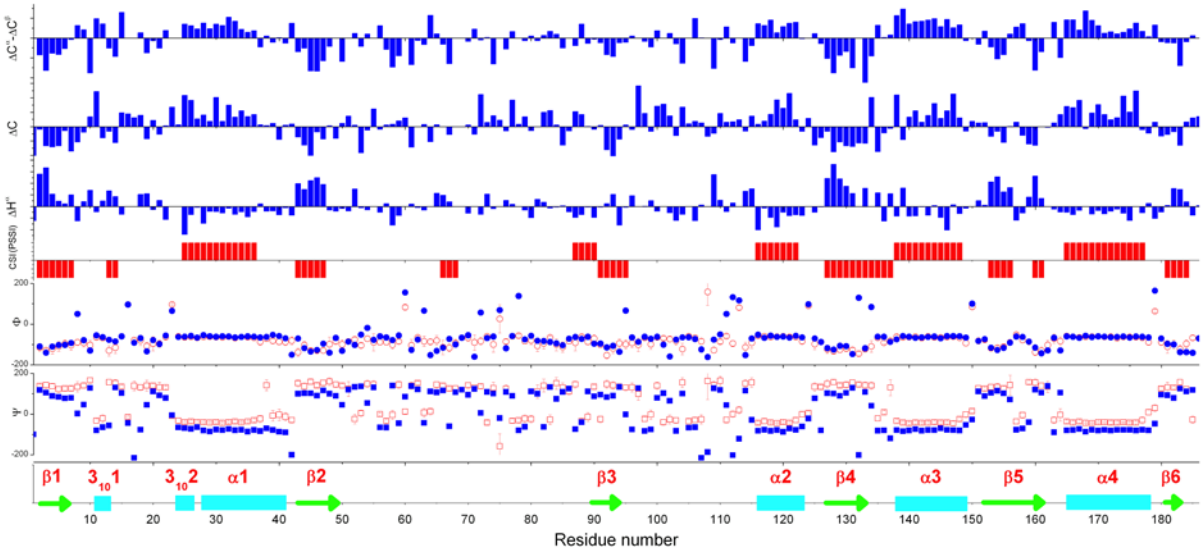

Figure 2. Summary of backbone resonance assignment and secondary structure of PncA. Delta values of backbone carbon and proton to random coil chemical shifts were plotted. Backbone dihedral angles (phi, psi) were calculated using TALOS (open circles and rectangles) and compared with those extracted from *P. horikoshii* crystal structure. Secondary structure from homology modeling was shown in bottom.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A1043288). This work was also supported by a grant of Korea Healthcare Technology R&D Project (A092006) & National Research Foundation of Korea (R1A2A1A01003569).

References

1. Wüthrich, K. NMR of Proteins and Nucleic Acids (Wiley, New York) (1986).
2. Cornilescu, G., Delaglio F., Bax A. J. Biomol. NMR **13**, 289-302. (1999).
3. Wishart, D. S., Case, D. A. Methods Enzymol. **338**, 3-34. (2001).
4. Sanders, J. K. M., Hunter, B. K. Modern NMR Spectroscopy (Oxford Univ. Press, Oxford, UK) (1993).
5. Wishart, D. S., Sykes, B. D., Richards, F. M. Biochemistry **31**, 1647-1651. (1992).
6. Lee, K., Kang, S., Bae, Y., Lee, K., Kim, J., Lee, I., Lee, B. J. Kor. Mag. Reson. Soc. **17**, 105-110. (2014).
7. Sattler, M., Fesik, S. W. Structure **4**, 1245-1249. (1996).
8. Hansen, P. E. Prog. Nucl. Mag. Res. Sp. **20**, 207-255. (1988).
9. Yi, J., Yoo, J. K., Kim, J. K., Son, W. S. J. Kor. Magn. Reson. Soc. **17**, 30-39. (2013).
10. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. J. Biomol. NMR **6**, 277-293. (1995).
11. Johnson, B. A., Blevins, R. A. J. Biomol. NMR **4**, 603-614. (1994).
12. Wishart, D. S., Sykes, B. D. J. Biomol. NMR **4**, 171-180. (1994).
13. Wishart, D. S., Sykes, B. D. Methods Enzymol. **239**, 363-392. (1994).
14. Gardner, K. H., Rosen, M. K., Kay, L. E. Biochemistry **36**, 1389-1401. (1997).
15. Shan, X. J., Gardner, K. H., Muhandiram, D. R., Kay, L. E., Arrowsmith, C. H. J. Biomol. NMR **11**, 307-318. (1998).
16. Venters, R. A., Farmer, B. T., Fierke, C. A., Spicer, L. D. J. Mol. Biol. **264**, 1101-1116. (1996).
17. Gardner, K. H., Kay, L. E. J. Am. Chem. Soc. **119**, 7599-7600. (1997).
18. Wang, Y., Jardetzky, O. J. Am. Chem. Soc. **124**, 14075-14084. (2002).
19. Wang, Y. J. Biomol. NMR **30**, 233-244. (2004).
20. Wang, Y., Jardetzky, O. Protein Sci. **11**, 852-861. (2002).
21. Metzler, W. J., Constantine, K. L., Friedrichs, M. S., Bell, A. J., Ernst, E. G., Lavoie, T. B., Mueller, L. Biochemistry **32,** 13818-13829. (1993).
22. Metzler, W. J., Leiting, B., Pryor, K., Mueller, L., Farmer, B. T. Biochemistry **35**, 6201-6211. (1996).