BIOCHEMISTRY
including biophysical chemistry & molecular biology

Rapid Report

pubs.acs.org/biochemistry

Specific Labeling of Threonine Methyl Groups for NMR Studies of Protein−Nucleic Acid Complexes

Kaustubh Sinha,[†] Linda Jen-Jacobson,[†] and Gordon S. Rule*[,‡]

[†]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, United States

[‡]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

Supporting Information

ABSTRACT: Specific ^{13}C labeling of Thr methyl groups has been accomplished via the growth of a standard laboratory strain of *Escherichia coli* on [2-^{13}C]glycerol in the presence of deuterated isoketovalerate, Ile, and Ala. Diversion of the label from the Thr biosynthetic pathway is suppressed by including Lys, Met, and Ile in the growth medium. This method complements the repertoire of methyl labeling schemes for NMR structural and dynamic studies of proteins and is particularly useful for the study of nucleic acid binding proteins because of the high propensity of Thr residues at protein−DNA and −RNA interfaces.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for probing protein dynamics over a range of time scales.[1] The combination of progress in optimized pulse sequences, better hardware, and better isotopic labeling schemes has allowed researchers to study high-molecular weight proteins.[2,3] Historically, dynamics studies using NMR spectroscopy have focused on the protein backbone (amide groups). However, the NH group may not reflect side-chain motions and hence the need to study side-chain dynamics. Among all the groups available for studying side chains, methyls are especially suited for the purpose as they are well-distributed through the protein structure. Furthermore, the methyl HMQC (heteronuclear multiple-quantum coherence) experiments can exploit the TROSY effect resulting from cancellation of intra-methyl dipolar relaxation interactions to give better resolved and more sensitive spectra.[4] This has allowed the use of methyl groups in studying the dynamics of high-molecular weight proteins and their complexes.

Prerequisites for ^{13}C-based methyl dynamics studies in large proteins are the specific ^{13}C labeling of the methyl carbons (adjacent to ^{12}C) and their protonation in a perdeuterated background. Initially, methyl labeling utilized [^{1}H,^{13}C]pyruvate (Pyr) as the sole source of carbon in deuterated medium.[5] More recent strategies involve the use of selectively labeled α-keto acids to specifically incorporate protonated ^{13}C-labeled methyls into proteins.[6,7] Commercially available precursors allow incorporation of $^{13}CH_3$, $^{13}CH_2D$, or $^{13}CHD_2$ groups at the Ile, Leu, and Val methyl sites in the protein. The CH_3 isotopomer allows characterization of millisecond to microsecond motions using relaxation dispersion experiments.[8,9] The CH_2D and CHD_2 isotopomers are well suited for deuterium relaxation studies,[10] while the CHD_2 isotopomer is ideal for carbon relaxation.[11] More recently, a procedure for specifically labeling the Ala methyls has been published.[12] This protocol involves using perdeuterated α-ketoisovalerate, succinate, and isoleucine to prevent the scrambling of the ^{13}C label. A more general labeling approach uses [1-^{13}C]- or [2-^{13}C]glucose[13] or [1,3-^{13}C]- or [2-^{13}C]glycerol[14] as the sole carbon source to express proteins that are labeled at the methyl carbons. A disadvantage of using glucose instead of glycerol as the carbon source is that the conversion of glucose to Pyr dilutes the ^{13}C label by 50%.

The methyl labeling techniques developed to date have focused on Ala, Ile, Leu, and Val side chains, all of which lack hydrogen bonding capability. In contrast, Thr can participate in hydrogen bonding as well as nonpolar interactions. Because of the different properties of methyl-containing amino acids, the distribution of these residues differs considerably for protein−protein, protein−RNA, and protein−DNA interfaces (Figure 1). In the case of DNA−protein interfaces, Leu (−0.69), Ile (−0.25), Ala (−0.16), and Val (−0.17) are all under-represented, showing negative propensities (given after the residue name), while Ser (0.25), Thr (0.27), Phe (0.35), Trp (0.41), Tyr (0.46), and Arg (0.57) are over-represented at protein−DNA interfaces. Thr residues are found with nearly equal propensity in both the core and rim of protein−DNA interfaces. Consequently, Thr residues potentially are better

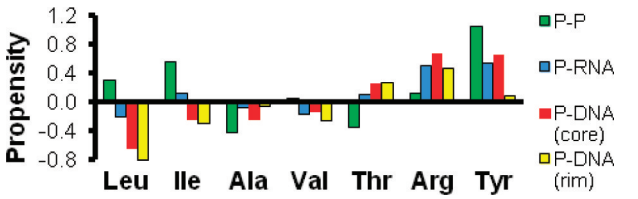

Figure 1. Propensity for selected residues found within protein−protein (P−P), protein−RNA (P−RNA), and protein−DNA (P−DNA) interfacial regions. The P−DNA interface is divided into a core region (solvent inaccessible) and a rim region (partially solvent accessible). Propensity is defined as $\ln(f_i/f_i^\circ)$, where f_i is the area of residue i in the interfacial region and f_i° is the area occupied by the residue on the remaining surface of the protein. Total P−DNA propensities for each residue are given in the text. Data from refs 15−17.

Received: September 25, 2011
Revised: October 29, 2011
Published: October 31, 2011

probes of DNA− and RNA−protein interfaces than any other methyl-containing amino acid currently utilized for NMR studies. Even though Thr methyls have a narrow chemical shift range, their proximity to nucleic acids is expected to result in a broader dispersion of shifts.

Presented here is a technique for uniquely labeling the methyl group of threonine residues of a protein expressed using a standard *Escherichia coli* T7 expression host[18] without genetic modification. The technique works in H_2O as well as ∼100% D_2O media. We show that there is no $^{13}C-^{13}C$ coupling to the adjacent carbon and that $^{13}C-H$ cross-peaks in the deuterated sample arise from the $^{13}CHD_2$ isotopomer.

We use [2-^{13}C]glycerol as the primary carbon source in the bacterial growth medium. The metabolism of glycerol in *E. coli* is well understood (see Figure S1 of the Supporting Information). Glycerol is converted to Pyr in the glycolytic pathway and then enters the TCA cycle as acetyl-CoA by condensation with oxaloacetate (OAA) to form citrate. Following the C2 atom from glycerol through the TCA cycle, the labeled carbon becomes either C1 or C4 of OAA. The OAA can then either be converted to Asp or re-enter the TCA cycle by reacting with another molecule of acetyl-CoA. Upon re-entry into the TCA cycle, both C1 and C4 of OAA will be lost as CO_2. Therefore, the ^{13}C label on glycerol either goes to Asp and subsequently to Thr or is lost as CO_2. The side-chain carboxyl group of Asp ultimately forms the methyl group on Thr, with the methyl protons arising from the media as well as from the carbon source. In 100% D_2O-containing media, it has been shown that 45% of the Thr methyls are labeled as $^{13}CHD_2$ while 81% of the Thr C_β atoms are deuterated.[19] The high level of deuteration at C_β leads to favorable relaxation properties in larger proteins.

A number of biosynthetic pathways can lead to a dilution of the ^{13}C label (see Figure S1 of the Supporting Information). Pyruvate can be converted to Ala, Leu, Val, and Ile, while Asp can be converted to Met and Lys. Thr itself can be converted to Ile. Therefore, inhibiting these pathways is important for the prevention of dilution of the label. This was accomplished by adding deuterated Ala, α-ketoisovalerate, Ile, and protonated Met and Lys to the medium before inducing protein expression (see the Supporting Information for more details). Because the carbon lost from the TCA cycle due to biosynthetic needs is obtained from CO_2, $NaH^{13}CO_3$ was added to the medium to enhance labeling (see below). The labeling scheme was tested on the 130-residue RNA binding domain of the transcriptional termination factor rho (rho130).[20] To estimate the level of incorporation of ^{13}C at the methyl group of Thr, Met was labeled by adding ^{13}C methyl-labeled Met to the medium. The Thr labeling efficiency was tested using the following sources of ^{13}C: (a) [2-^{13}C]glycerol, (b) [2-^{13}C]glycerol and $NaH^{13}CO_3$ (2 g/L), (c) [2-^{13}C]glycerol and $NaH^{13}CO_3$ (10 g/L), and (d) $NaH^{13}CO_3$ (10 g/L). Media containing 50 and 100% D_2O were used to test the contribution of the solvent to protonation of the Thr methyl.

The methyl region of the C−H correlated spectra of uniformly ^{13}C labeled rho130 and the specifically labeled sample (only the Thr and Met methyls labeled) is shown in Figure 2A,B. The Thr labeling protocol is extremely specific and reduces the intensity of the other methyl peaks by ∼95%, significantly reducing the extent of crowding in this region of the spectrum. Figure 2B also shows the one-dimensional slice along the carbon dimension of a threonine peak. The lack of any splitting shows that the carbon adjacent to the methyl ^{13}C carbon is not labeled. Figure 2C shows the spectrum of the deuterated specifically labeled sample acquired using a refocused INEPT experiment that selects for $^{13}CHD_2$ signals.[11] A C−H HMQC experiment showed that that $^{13}CHD_2$ isotopomer is the predominant protonated species (approximately 90% $^{13}CHD_2$ and 10% $^{13}CH_2D$) when the cells are grown in 100% D_2O. In the case of expression in 50% D_2O, the relative $^{13}CH_3$:$^{13}CH_2D$:$^{13}CHD_2$ ratio is approximately 40:40:20; thus, some fraction of the hydrogens on the Thr methyl group are supplied directly from the solvent as well as from the protonated carbon source.

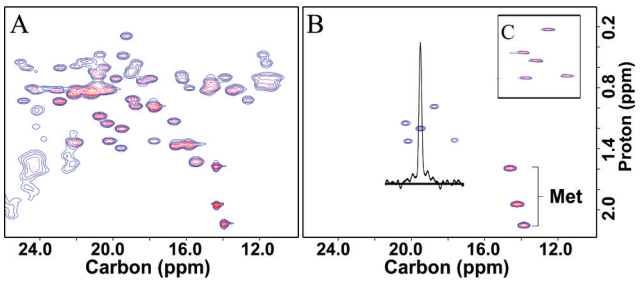

Figure 2. Two-dimensional $^1H-^{13}C$ spectra of rho130. (A) HSQC spectrum of uniformly ^{13}C labeled protein. (B) HSQC spectrum of rho130 specifically labeled at Thr and Met methyl groups. (C) Refocused INEPT experiment with the deuterated rho130 sample showing the Thr CHD_2 peaks.

The efficiency of ^{13}C labeling of the Thr methyls using glycerol alone is 32% (Figure 3). If all of the ^{13}C from C2 of glycerol was used for Thr synthesis, then the γ-methyl would be labeled to a level of 50% (see Figure S1 of the Supporting Information). Numerical simulations show that with 32% labeling, approximately 40% of the carbon flux in the TCA cycle is diverted to biosynthetic needs. Because this carbon is replenished by carboxylation of phosphoenolpyruvate with $^{12}CO_2$ to produce OAA, the observed labeling efficiency is less than 50%. Because the incorporated CO_2 becomes the C4 position in OAA, the methyl labeling efficiency can be increased by adding $NaH^{13}CO_3$ to the medium prior to induction. Figure 3 shows that an 80% increase in signal intensity, relative to that obtained with [2-^{13}C]glycerol alone, was obtained upon addition of 10 g/L $NaH^{13}CO_3$ to the medium. The use of $NaH^{13}CO_3$ (10 g/L) alone is as efficient as the use of [2-^{13}C]glycerol.

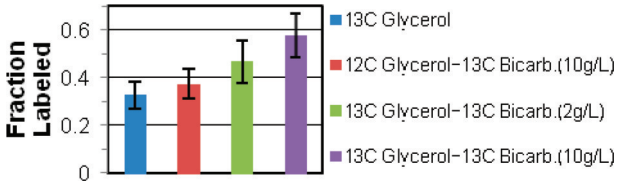

Figure 3. Mean normalized peak intensities when using different ^{13}C sources. Individual peaks were first normalized to the corresponding peaks in a uniform ^{13}C-labeled sample to account for intensity differences due to polarization transfers. Each set was then normalized to the most intense Met resonance. The bars indicate the range of values observed for the five Thr residues.

When [2-^{13}C]glycerol is used as the carbon source, the carbonyls of Arg, Asp, Asn, Glu, Gln, Leu, Pro, and Thr residues will be selectively labeled. The ^{13}C carbonyl can be used to filter the two-dimensional $^1H-^{15}N$ spectrum (Figure S2

of the Supporting Information), showing only peaks from NH groups preceding the labeled carbonyl. The intensity of the peaks in the filtered spectrum is consistent with the expected labeling pattern reported in previous studies[13,14,21] (Figure S3 of the Supporting Information). Because carbonyl editing allows identification of the preceding residue type for NH, this information can be used in the main-chain resonance assignment process.[22,23] Note that the residues that are labeled at the carbonyl position with [2-^{13}C]glycerol are complementary to the set of residues (I, L, Y, V, A, and P) identified as being useful for assignment purposes using individual ^{13}C carbonyl-labeled amino acids.[23]

In addition to selective labeling at the carbonyl position, the use of [2-^{13}C]glycerol leads to selective labeling of the C_α position.[13,14,21] An analysis of peak intensities in the C_α filtered ^{1}H−^{15}N spectra of rho130 (Figure S4 of the Supporting Information) shows that this labeling pattern is preserved with the Thr labeling scheme. The simplification of the C−H spectrum facilitates C_α relaxation studies, an important contribution to the characterization of backbone dynamics.

In summary, the Thr labeling protocol can be easily incorporated into standard protein expression protocols. The labeled methyl carbon is not coupled to the adjacent carbon, and when using 100% deuterated media, $^{13}CHD_2$ is the predominant species. Also, because of the use of deuterated α-ketoisovalerate, Ala, Met, and Ile, even proteins expressed in H_2O-based media would have a significantly deuterated methyl background. These characteristics make this protocol ideal for relaxation studies. We note that although the labeling efficiency of the Thr methyl with respect to ^{13}C exceeds 50%, the fraction of $^{13}CHD_2$ methyls is estimated to be ∼25% when 100% D_2O is used, reducing the overall efficiency relative to those of other methyl labeling schemes. However, the fraction of the $^{13}CHD_2$ isotopomer can be increased by increasing the H_2O content of the media. The protocol also selectively enriches the labeling of the carbonyl carbons, providing useful information for resonance assignments. This protocol complements the set of methyl labeling schemes available to the NMR community and should be particularly useful in the study of protein−DNA interactions.

■ ASSOCIATED CONTENT

Supporting Information

Summary of relevant metabolic pathways and additional details regarding the labeling procedure, two-dimensional ^{15}N−^{1}H HSQC spectra of rho130, without editing and edited by N−C_α and N−CO coupling, and peak intensities in the edited spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (412) 268-1839. E-mail: rule@andrew.cmu.edu.

Funding

This work was supported by the Mellon College of Science at Carnegie Mellon University and National Institutes of Health Merit Grant 5R37-GM029207 to L.J.-J.

■ ACKNOWLEDGMENTS

We thank Virgil Simplaceanu and Roberto Gil for support of the NMR spectrometers and for help in data collection.

■ REFERENCES

(1) Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006) *Chem. Rev. 106*, 3055−3079.
(2) Kay, L. E. (2005) *J. Magn. Reson. 173*, 193−207.
(3) Tzakos, A. G., Grace, C. R., Lukavsky, P. J., and Riek, R. (2006) *Annu. Rev. Biophys. Biomol. Struct. 35*, 319−342.
(4) Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E., and Kay, L. E. (2003) *J. Am. Chem. Soc. 125*, 10420−10428.
(5) Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T., and Kay, L. E. (1996) *J. Mol. Biol. 263*, 627−636.
(6) Tugarinov, V., Kanelis, V., and Kay, L. E. (2006) *Nat. Protoc. 1*, 749−754.
(7) Ruschak, A. M., Velyvis, A., and Kay, L. E. (2010) *J. Biomol. NMR 48*, 129−135.
(8) Tugarinov, V., and Kay, L. E. (2004) *J. Am. Chem. Soc. 126*, 9827−9836.
(9) Korzhnev, D. M., Kloiber, K., Kanelis, V., Tugarinov, V., and Kay, L. E. (2004) *J. Am. Chem. Soc. 126*, 3964−3973.
(10) Tugarinov, V., Ollerenshaw, J. E., and Kay, L. E. (2005) *J. Am. Chem. Soc. 127*, 8214−8225.
(11) Ishima, R., Louis, J. M., and Torchia, D. A. (1999) *J. Am. Chem. Soc. 121*, 11589−11590.
(12) Ayala, I., Sounier, R., Use, N., Gans, P., and Boisbouvier, J. (2009) *J. Biomol. NMR 43*, 111−119.
(13) Lundstrom, P., Teilum, K., Carstensen, T., Bezsonova, I., Wiesner, S., Hansen, D. F., Religa, T. L., Akke, M., and Kay, L. E. (2007) *J. Biomol. NMR 38*, 199−212.
(14) Higman, V. A., Flinders, J., Hiller, M., Jehle, S., Markovic, S., Fiedler, S., van Rossum, B. J., and Oschkinat, H. (2009) *J. Biomol. NMR 44*, 245−260.
(15) Lo Conte, L., Chothia, C., and Janin, J. (1999) *J. Mol. Biol. 285*, 2177−2198.
(16) Bahadur, R. P., Zacharias, M., and Janin, J. (2008) *Nucleic Acids Res. 36*, 2705−2716.
(17) Biswas, S., Guharoy, M., and Chakrabarti, P. (2009) *Proteins 74*, 643−654.
(18) Studier, F. W., and Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113−130.
(19) Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J., and Mulder, F. A. (2010) *J. Am. Chem. Soc. 132*, 2952−2960.
(20) Briercheck, D. M., Wood, T. C., Allison, T. J., Richardson, J. P., and Rule, G. S. (1998) *Nat. Struct. Biol. 5*, 393−399.
(21) LeMaster, D. M., and Kushlan, D. M. (1996) *J. Am. Chem. Soc. 118*, 9255−9264.
(22) McCallum, S. A., Hitchens, T. K., and Rule, G. S. (1999) *J. Mol. Biol. 285*, 2119−2132.
(23) Takeuchi, K., Ng, E., Malia, T. J., and Wagner, G. (2007) *J. Biomol. NMR 38*, 89−98.