Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202403148
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

Expanding the Toolset of Biomolecular NMR with Efficient and Cost-Effective ^{17}O-Labeling via Bacterial Expression

Ziya Tian,[a] Erica Truong,[a, b] Wenhao Hu,[a, b] Jiaxing Fan,[a, b] Riqiang Fu,[b] Timothy A. Cross,[a, b] Xinsong Lin,[a] Rongfu Zhang,*[a, b] and Yan-Yan Hu*[a, b]

Oxygen plays a central role in biomolecular structures and functions, with ^{17}O NMR emerging as a powerful tool for elucidating biomolecular properties. However, the low natural abundance of the NMR-active isotope, ^{17}O (0.0373 %), presents a significant hurdle to its widespread application. Here, we introduce a rapid and cost-effective approach for amino acid-specific ^{17}O-labeling of recombinant proteins. Using a common bacterial expression system and with a 30-minute rapid synthesis protocol of ^{17}O-labeled amino acids via mechanochemical saponification, we have generated Leu- and Phe-specific ^{17}O-labeled recombinant proteins derived from diverse organisms, including CrgA and FtsQ from *Mycobacterium tuberculosis* and E protein from SARS-CoV-2 virus, demonstrating the applicability of our technique for amino acids known to be isotopically labeled without scrambling. We have acquired magic-angle-spinning ^{17}O NMR of these proteins to confirm the successful ^{17}O labeling and illustrate the sensitivity of ^{17}O NMR to the protein's local structural environments. Our work significantly broadens the accessibility of ^{17}O-NMR, empowering researchers to delve deeper into protein biophysics and biochemistry. This approach opens new avenues for understanding cellular processes at the molecular level by providing an effective tool for investigating oxygen-related interactions and chemistry within biomolecules.

Main

Nuclear magnetic resonance (NMR) spectroscopy is a foundational technique for determining the structures and dynamics of proteins.[1] Despite its effectiveness, NMR, especially solid-state NMR, generally struggles to detect signals from poorly abundant isotopes. Therefore, isotope enrichment of proteins, typically ^{13}C or ^{15}N, is often necessary for biomolecular NMR investigations. Oxygen is a vital structural component and plays a central role in the functions of proteins.[2] In addition, its inherent large chemical shifts and quadrupolar couplings are remarkably sensitive to local environments. Given its significance, ^{17}O NMR emerges as a highly advantageous tool for investigating the structures, properties, and functions of proteins. However, its use in biomolecular NMR is severely limited, partly attributed to the low natural abundance of its only NMR-active isotope, ^{17}O (0.0373 %), and the quadrupolar nature of ^{17}O (spin-5/2) that introduces NMR interactions on the order of MHz, leading to signal broadening. The recent developments in ultra-high magnetic fields[3] have greatly improved the resolution of ^{17}O NMR.[4] Despite the enhancements in the strengths of magnetic fields, routine application of ^{17}O NMR for protein structural characterization still presents significant challenges. In contrast to ^{13}C and ^{15}N-labeled amino acids, only two ^{17}O-labeled amino acids, L-glutamic acid and L-tyrosine, are commercially available at exorbitant prices (Table S1). Consequently, cost-effective, high-yield synthesis of ^{17}O-labeled amino acids becomes a necessity. However, conventional synthesis of ^{17}O-labeled amino acids typically requires a substantial excess of costly ^{17}O-enriched water (Table S1: currently priced at approximately $3,000 per 1 g) and entails intricate processing at elevated temperatures, rendering it impractical for routine implementation. Additionally, there is a dearth of standardized protocols for integrating ^{17}O amino acids into protein sequences. The traditional practice of using a bacterial expression system for incorporating ^{13}C and ^{15}N labels into proteins through controlled minimal media is not feasible for routine research due to the substantial quantity of ^{17}O-labeled amino acids required during culture growth. While one documented study employed this approach to express amino-acid-specific ^{17}O-labeled yeast ubiquitin in an auxotrophic *E. coli* strain,[5] this method's practical viability is limited due to its high cost, low yield, and protracted enrichment process.

The key to enabling the broader use of biomolecular ^{17}O NMR to reveal insights beyond what traditional ^{13}C and ^{15}N NMR can offer lies in making ^{17}O labeling of proteins convenient and cost-effective. Successful ^{17}O-labeling of biomolecules has been previously limited to specific high-yield model proteins, such as ubiquitin.[5] There is a growing interest in developing a widely applicable and efficient method for ^{17}O-labeling of proteins that can be easily adapted to commonly used cell lines in most structural biology labs.

In this study, we combine rapid and cost-effective ^{17}O labeling of amino acids with the commonly utilized bacterial strain BL21(DE3) to establish a general and economical protocol for ^{17}O labeling of proteins. We demonstrate the successful ^{17}O labeling of the *Mycobacterium tuberculosis* (*Mtb*) cell-division

[a] Z. Tian, E. Truong, W. Hu, J. Fan, T. A. Cross, X. Lin, R. Zhang, Y.-Y. Hu
Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA
[b] E. Truong, W. Hu, J. Fan, R. Fu, T. A. Cross, R. Zhang, Y.-Y. Hu
National High Magnetic Field Laboratory, Tallahassee, FL 32310, USA
E-mail: rzhang3@fsu.edu
yhu@fsu.edu

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202403148

protein CrgA through heterogeneous expression in the BL21(DE3)RP *E. coli* strain, utilizing ^{17}O-labeled leucine synthesized via mechanochemical saponification.[6] CrgA, a 93-residue transmembrane protein, plays a pivotal role in the *Mtb* divisome through interactions with multiple other cell division proteins. To show the versatility of this ^{17}O-labeling protocol, we also labeled *Mtb* FtsQ and SARS-CoV-2 E proteins using the same method. We further tested the method using ^{17}O-labeled Phenylalanine to label CrgA.

As shown in Figure 1, the ^{17}O-labeling of amino acids via mechanochemical reactions was carried out in a ball milling jar for about 30 min, and the product was directly transferred into M9 minimal media for protein expression. This protocol significantly reduces the time required to obtain ^{17}O-labeled amino acids by eliminating the purification process and preventing product loss during amino acid purification. The conversion of L-Leucine methyl ester hydrochloride precursor to L-Leucine was close to 100%, as indicated by the complete shift of the 1736 cm^{-1} peak to 1577 cm^{-1} in the infrared analysis (Figure S1). [6] To examine possible changes in the chirality of amino acids synthesized with this method, we performed circular dichroism spectroscopy on our lab-synthesized Leu and commercial L-Leu. The results (Figure S2) show matched CD line shapes, suggesting the conservation of chirality. Laurencin's report[6] also exhibited no loss of chirality in their synthesized Fmoc-amino acids. The successful ^{17}O-labeling of the final L-Leucine product was verified by liquid chromatography-mass spectrometry (LC–MS) (Figure S3 and Figure S4 for L-Phe). Based on the LC–MS spectra, the level of ^{17}O enrichment for L–Leu and L-Phe products was found to be ~60% and ~72%, respectively. We further tested the chirality of synthesized LIt is worth mentioning that the small quantities of side products from the mechanochemical reaction showed no influence on the growth rate and protein yield of CrgA expression (Figure S5). For 250 mL of M9 culture, we obtained ~25 mg of purified CrgA protein, comparable to our CrgA yield of ^{13}C and ^{15}N labeling in a similar culture volume. For 500 mL of M9 culture, ~4.8 mg of FtsQ58-126 (Figure S6) and ~4 mg of E (Figure S7) were obtained, respectively. Detailed procedures are provided in the method section.

It should be noted that amino acid type-specific isotopic labeling cannot be uniformly applied to all 20 amino acids. Not every amino acid is incorporated directly into the target protein. For example, Glu and Asp residues are difficult to label, as they tend to scramble into other amino acids. Over the past few decades, significant research has focused on identifying amino acids that can be used to isotopically enrich recombinant proteins overexpressed in *E. coli*, aiming for minimal or no isotopic scrambling.[7] It has been demonstrated that Ile, Ala, Leu, Phe, Val, Met, Trp, and Thr can be labeled with minimal or no scrambling if labeling conditions are properly controlled, with varied additions of unlabeled amino acids and short expression times.[1c,8]

In Figure 2A, the ^{17}O MAS NMR spectrum of ^{17}O-Leu-labeled CrgA in liposomes reveals a resonance at approximately 277 ppm, corresponding to the Leu carbonyl ^{17}O sites. In addition, the ^{17}O NMR spectra of ^{17}O-Leu-labeled SARS-CoV-2 E protein (Figure 2B) and *Mtb* FtsQ$_{58\text{-}126}$ (Figure 2C) exhibit peaks at approximately 256 ppm and 287 ppm, respectively. A notable 20 ppm chemical shift difference for these Leu carbonyl ^{17}O signals indicates the excellent sensitivity of ^{17}O as a probe for distinct local environments resulting from different protein characteristics. The differences in ^{17}O chemical shifts likely result from variations in protein dynamics, making ^{17}O a unique probe for protein motions. However, the spectral resolution is insufficient to resolve the subtle differences in individual ^{17}O-labeled Leu carbonyls within the same protein. Since ^{17}O is quadrupolar (spin 5/2), the ^{17}O resonances are significantly

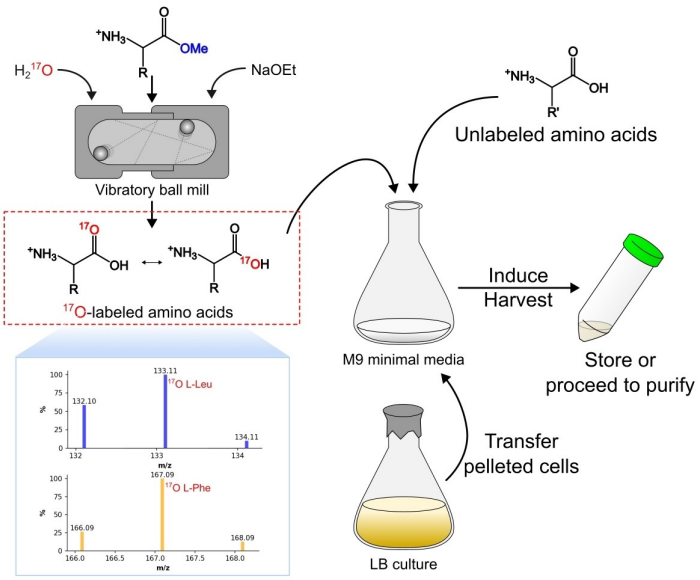

Figure 1. Bacterial expression flowchart using M9 minimal media supplemented with ^{17}O-labeled amino acids prepared with a cost-effective mechanochemical approach.

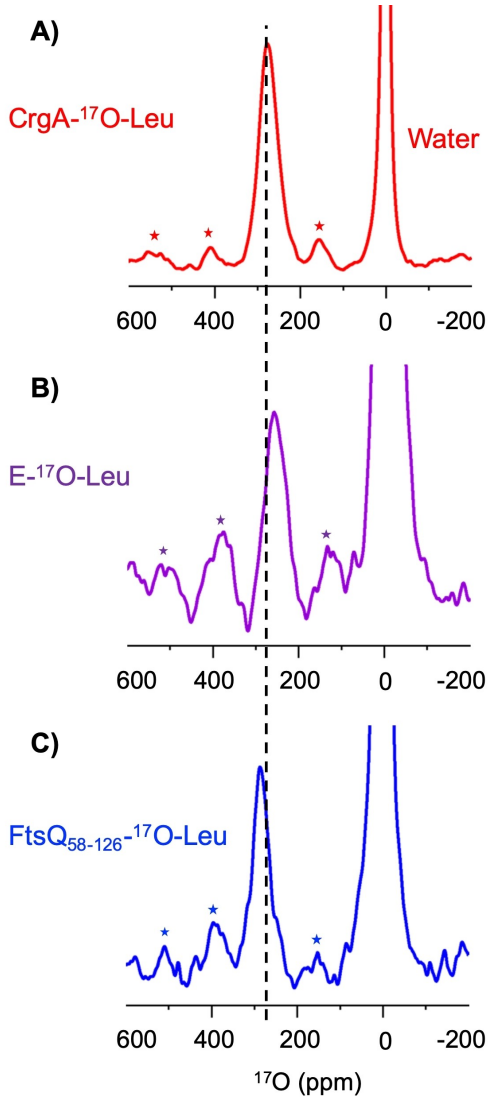

Figure 2. ^{17}O MAS NMR spectra of ^{17}O-Leu labeled CrgA, E, and $FtsQ_{58-126}$ in POPC/POPG liposomes. Leu carbonyl ^{17}O signal and water ^{17}O signal are assigned. Spinning side bands are indicated by stars.

broadened by quadrupolar interactions on the order of MHz. The single ^{17}O NMR peak results from overlapping resonances of all labeled carbonyls (11 Leu for CrgA, 14 Leu for E, and 7 Leu for FtsQ58-126). A higher magnetic field, fast magic-angle-spinning, and advanced NMR methods will be implemented to resolve the subtle differences in the local environments of the labeled amino acids within the same protein. For the ^{17}O-Phe-labeled CrgA, a high-sensitivity ^{17}O NMR signal was observed at ~278 ppm within the same range as noted above for Leu labeling (Figure S8), confirming the carbonyl labeling and suggesting the high adaptability of this method for ^{17}O-labeling and structural sensitivity. The amino acid sequences of CrgA, E and FtsQ58-126 with Leucine ^{17}O labeling sites highlighted are shown in Figure S9.

In summary, we have established a practical and cost-effective method for ^{17}O-labeling of proteins, comparable to standard ^{13}C and ^{15}N labeling and even more economical. This isotopic labeling strategy introduces a viable avenue for routine ^{17}O NMR investigations into protein structures and dynamics, particularly useful for probing the pivotal roles played by oxygen atoms within the active sites of proteins in catalyzing biochemical reactions. Previously inaccessible, these functional sites can now be effectively investigated using NMR spectroscopy enabled by our method.

Supporting Information

The authors have cited additional references within the Supporting Information.[1]

Author Contributions

Y.-Y. H., R. Z., and Z. T. designed the research; Z. T., R. Z., E. T., W. H., J. F., Y.-Y. H., T. A. C., X. L., and R. F. performed the research and analyzed the data; Z. T., R. Z., and Y.-Y. H. wrote and edited the manuscript.

Acknowledgements

We would like to thank Dr. J. S. Raaj Vellore Winfred of the Materials Characterization Lab (MACLab) at Florida State University for his assistance with acquiring FT-IR data, and Dr. Peter Randolph of the Physical Biochemistry Facility at Florida State University for assisting with the acquisition of CD data. Additionally, we thank Dr. Austin Peach from Schurko's group and Yudan Chen from Hu's group at Florida State University for valuable discussions.

This work was supported by National Institutes of Health Grants AI119187 and GM122698, as well as funding from Florida State University. All NMR experiments were carried out at the National High Magnetic Field Laboratory supported by the NSF Cooperative Agreements DMR1644779 and DMR2128556 and by the State of Florida.

Conflict of Interests

The authors declare no competing interests.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

[1] a) R. Zhang, H. Qin, R. Prasad, R. Fu, H.-X. Zhou, T. A. Cross, *Communications Biology* **2023**, *6*, 1109; b) F. Wang, S. K. Ramakrishna, P. Sun, R. Fu, *Journal of Magnetic Resonance* **2021**, *332*, 107067; c) N. Das, J. Dai, I. Hung, M. Rajagopalan, H.-X. Zhou, T. A. Cross, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E119–E126
[2] a) T. Oozeki, T. Nakai, K. Kozakai, K. Okamoto, S. i Kuroda, K. Kobayashi, K. Tanizawa, T. Okajima, *Nat. Commun.* **2021**, *12*, 933; b) L. Crocetti, M. T.

Quinn, A. Cilibrizzi, M. P. Giovannoni, *Front. Chem.* **2022**, *10:980152*; c) A. Shimada, T. Tsukihara, S. Yoshikawa, *Front. Chem.* **2023**, *11:1108190*; d) S. Soussou, A. Jablaoui, V. Mariaule, A. Kriaa, H. Boudaya, M. Wysocka, A. Amouri, A. Gargouri, A. Lesner, E. Maguin, M. Rhimi, *Sci. Rep.* **2023**, *13*, 17571.

[3] Z. Gan, I. Hung, X. Wang, J. Paulino, G. Wu, I. M. Litvak, P. L. Gor'kov, W. W. Brey, P. Lendi, J. L. Schiano, M. D. Bird, I. R. Dixon, J. Toth, G. S. Boebinger, T. A. Cross, *J. Magn. Reson.* **2017**, *284*, 125–136.

[4] a) J. Paulino, M. Yi, I. Hung, Z. Gan, X. Wang, E. Y. Chekmenev, H. X. Zhou, T. A. Cross, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11908–11915; b) R. Zhang, T. A. Cross, X. Peng, R. Fu, *J. Am. Chem. Soc.* **2022**, *144*, 7881–7888; c) E. G. Keeler, V. K. Michaelis, M. T. Colvin, I. Hung, P. L. Gor'kov, T. A. Cross, Z. Gan, R. G. Griffin, *J. Am. Chem. Soc.* **2017**, *139*, 17953–17963; d) V. Martins, J. Xu, X. Wang, K. Chen, I. Hung, Z. Gan, C. Gervais, C. Bonhomme, S. Jiang, A. Zheng, B. E. G. Lucier, Y. Huang, *J. Am. Chem. Soc.* **2020**, *142*, 14877–14889.

[5] B. Lin, I. Hung, Z. Gan, P. H. Chien, H. L. Spencer, S. P. Smith, G. Wu, *ChemBioChem* **2021**, *22*, 826–829.

[6] J. Špačková, I. Goldberga, R. Yadav, G. Cazals, A. Lebrun, P. Verdié, T.-X. Métro, D. Laurencin, *Chem. Eur. J.* **2023**, *29*, e202203014.

[7] a) D. Lacabanne, B. H. Meier, A. Böckmann, *J. Biomol. NMR* **2018**, *71*, 141–150; b) J. Schörghuber, L. Geist, G. Platzer, M. Feichtinger, M. Bisaccia, L. Scheibelberger, F. Weber, R. Konrat, R. J. Lichtenecker, *J. Biomol. NMR* **2018**, *71*, 129–140.

[8] a) Y. Miao, T. A. Cross, R. Fu, *J. Biomol. NMR* **2013**, *56*, 265–273; b) N. Das, D. T. Murray, T. A. Cross, *Nature protocols* **2013**, *8*, 2256–2270; c) M. Sharma, M. Yi, H. Dong, H. Qin, E. Peterson, D. D. Busath, H. X. Zhou, T. A. Cross, *Science* **2010**, *330*, 509–512.

Manuscript received: August 21, 2024
Accepted manuscript online: December 1, 2024
Version of record online: December 5, 2024