BIOCHEMISTRY
including biophysical chemistry & molecular biology

Rapid Report

pubs.acs.org/biochemistry

Quantitation of Recombinant Protein in Whole Cells and Cell Extracts via Solid-State NMR Spectroscopy

Erica P. Vogel and David P. Weliky*

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

Supporting Information

ABSTRACT: Recombinant proteins (RPs) are commonly expressed in bacteria followed by solubilization and chromatography. Purified RP yield can be diminished by losses at any step with very different changes in methods that can improve the yield. Time and labor can therefore be saved by first identifying the specific reason for the low yield. This study describes a new solid-state nuclear magnetic resonance approach to RP quantitation in whole cells or cell extracts without solubilization or purification. The method is straightforward and inexpensive and requires only ∼50 mL culture and a low-field spectrometer.

A common approach to producing recombinant protein (RP) begins with incorporation of recombinant DNA (rDNA) into bacteria followed by cell growth, expression and lysis, and finally chromatography to obtain pure RP. The assessment of RP quantity and purity after the expression, solubilization, and/or chromatography steps is typically done using sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) that separates proteins by molecular weight (MW). For several different RPs in our laboratory, the RP gel band was not clearly observed after expression or solubilization and the final RP purified yield was unacceptably low, e.g., 0.1 mg of RP/L of culture.[1] One hypothesis to explain this result is low RP expression followed by high-yield solubilization and chromatography. A second distinct hypothesis is high RP expression followed by poor solubilization and high-yield chromatography. A third hypothesis is high RP expression and solubilization followed by chromatographic loss of RP. Distinguishing among these hypotheses is important because (1) the corrective changes to the experimental protocol to improve RP yield are very different for each hypothesis and (2) implementing these changes is often time- and labor-intensive. For example, low protein expression might be improved by codon changes in the rDNA or by varying induction time, whereas low solubilization might be improved by comprehensive screening of lysis buffers that differ in terms of additives such as denaturants and detergents.

This study focuses on distinguishing between the first low expression and the second low solubilization hypotheses. The third chromatographic loss hypothesis is typically straightforwardly tested by comparing the relative RP gel band intensities of washes versus elutions from the chromatographic column. RP expression is typically examined by first boiling an aliquot of cells in buffer containing SDS buffer with subsequent SDS−PAGE of solubilized protein. The RP quantity is estimated by comparison of the intensity of the RP band to the intensities of bands of native bacterial proteins. There are a few reports of more accurate quantitation.[2] This approach relies on a RP MW that is fortuitously different from the MWs of any of the abundant bacterial proteins. Alternatively, the quantity of the solubilized RP could be much higher than the quantities of any of these native proteins, i.e., high RP expression and high solubility.

An assumption of the approach is that most of the RP is solubilized by boiling. However, the largest RP fraction in cells is typically solid inclusion body (IB) aggregates that can be difficult to solubilize. It is therefore important to develop alternative approaches for RP quantitation in either whole cells or cell extracts enriched with IB solids. One potential method is IR spectroscopy of IBs and is based on the hypothesis of an increased fraction of β-sheet for the RP in IBs relative to the native structure, perhaps because of partial amyloid structure in the IB.[3] However, the fractional increase in β-sheet structure is likely highly variable among RPs in IBs with one RP in IBs showing retention of a large fraction of native helical structure.[4]

This study describes an alternate solid-state nuclear magnetic resonance (SSNMR) approach to quantifying RP in whole bacterial cells and cell extracts enriched with IBs. The approach does not depend on the structure(s) of the RPs in IBs. We note that there have been earlier applications of SSNMR to whole bacterial cells and cell extracts with a typical goal of elucidation of details of atomic-resolution structure.[5−8] The new method has been tested with five different RPs whose amino acid sequences are given in the Supporting Information. The generality of the approach is supported by use of different plasmid and *Escherichia coli* strain types.

One RP is human proinsulin (HPI), which is the precursor to the hormone insulin.[9] Folded HPI is a monomer with an α-helical core.[10] Three RPs (Hairpin, Fgp41, and Fgp41+) are different ectodomain segments of the HIV gp41 protein.[1,11,12] gp41 is an integral HIV membrane protein, and the ∼175 N-terminal residues of gp41 make up the ectodomain that lies outside the virus. The ectodomain is subdivided into the ∼20 N-terminal fusion peptide (FP) residues that bind to membranes and the larger C-terminal region that folds as a helical hairpin with a 180° turn.[13] There is further assembly of three hairpins to form a molecular trimer with a six-helix bundle (SHB) structure that is hyperthermostable. Hairpin, Fgp41, and Fgp41+ likely all form SHB structure with sequence differences among constructs as well as a lack of FP and most of the loop in Hairpin. The fifth

Received: June 4, 2013
Published: June 6, 2013

dx.doi.org/10.1021/bi4007034 | *Biochemistry* 2013, 52, 4285−4287

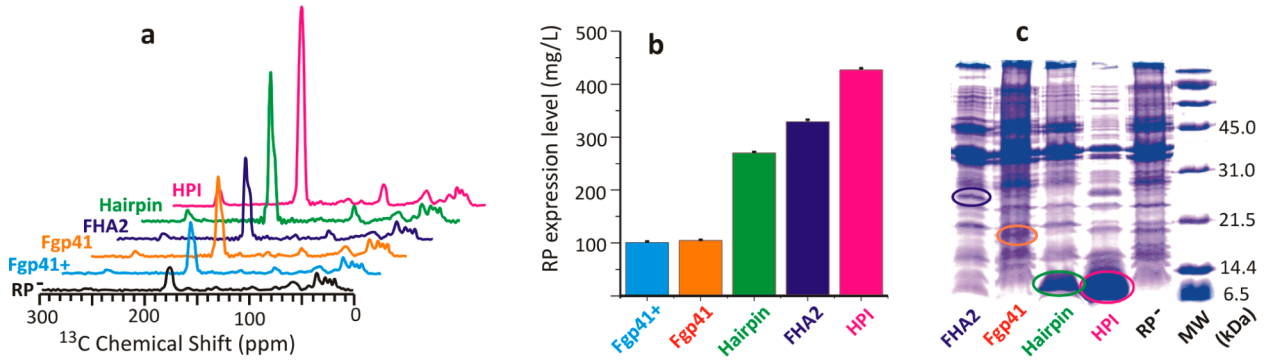

Figure 1. HC variant of RP quantitation. Panel a displays ^{13}C SSNMR spectra of [^{13}CO]Leu ICP samples. Panel b displays the RP expression levels calculated from the $RP^{+}_{lab} - RP^{-}_{lab}$ difference intensities of the ^{13}CO (~175 ppm) region. The differences between the ordering of spectral intensities in panel a and the ordering of expression levels in panel b are largely due to different N_{Leu}/N_{tot} ratios in the RP sequences. Panel c displays an SDS−PAGE gel of most of the samples, and the marked bands may correspond to the RPs.

RP, FHA2, is the full ectodomain of the HA2 subunit of the hemagglutinin protein of the influenza virus.[14−16] HA2 has a topology and membrane fusion function similar to those of gp41, but there is little sequence homology between HA2 and gp41.[17] Membrane-associated FHA2 has folded SHB structure. Previous efforts to solubilize each of these RPs from bacteria were consistent with a large fraction of RP in IBs.

We present "HC" and "HCN" SSNMR approaches to RP quantitation that require double-resonance ^{1}H, ^{13}C and triple-resonance ^{1}H, ^{13}C, ^{15}N SSNMR spectrometers and probes, respectively. For the HC approach, two samples are prepared that are denoted "RP^{+}_{lab}" and "RP^{-}_{lab}", and the bacteria have a plasmid with and without the RP rDNA insert, respectively. The preparation of either sample includes the addition of a ^{13}CO-labeled amino acid to the expression medium. For this study, this is [^{13}CO]Leu. The SSNMR ($RP^{+}_{lab} - RP^{-}_{lab}$) ^{13}CO difference intensity should therefore be the signal of the labeled (*lab*) ^{13}CO nuclei of the RP. Comparison with a standard curve of ^{13}CO intensity versus moles of ^{13}CO allows for conversion to the mass of RP per liter of bacterial culture, which is a common metric of RP expression. The variation in cell mass between the RP^{+}_{lab} and RP^{-}_{lab} samples is accounted for by matching the intensities of the two samples in the 0−90 ppm aliphatic region. This aliphatic ^{13}C signal serves as an internal standard because it is due to natural abundance (*na*) nuclei whose numbers should be comparable in a RP^{+} cell and a RP^{-} cell.

The second HCN approach applies rotational-echo double-resonance (REDOR) SSNMR to one RP^{+}_{lab} sample labeled with either a ^{13}CO- and ^{15}N-labeled amino acid or a ^{13}CO-labeled amino acid and a ^{15}N-labeled amino acid.[18] Separate S_0 and S_1 REDOR data are acquired with 1 ms dephasing. All ^{13}C signals are detected in S_0, whereas there is specific attenuation in S_1 of signals of directly bonded ^{13}CO−^{15}N spin pairs because of the ~1 kHz dipolar coupling. The $\Delta S = S_0 - S_1$ ^{13}CO spectrum is therefore dominated by these pairs.[19] The ΔS ^{13}CO signal intensity is converted to the mass of RP per liter of culture using a method analogous to that of the HC approach. Relative to HC, the HCN RP quantitation has the advantage of one rather than two samples. The HC variant has the advantage of requiring a double- rather than a triple-resonance SSNMR spectrometer and probe. Triple-resonance SSNMR instruments are less common and can have a lower ^{13}C sensitivity.

The Supporting Information provides detailed protocols for sample preparation and SSNMR. The ~2 day experiment is mostly unattended. The approach is inexpensive with a culture volume of only ~50 mL and ~10 mg of labeled amino acid. There is also a protocol for suppressing scrambling of the ^{13}CO and/or ^{15}N labels to other amino acid types.[20] A "whole cell" (WC) sample is the centrifugation pellet of the bacterial culture. Cell lysis is conducted prior to centrifugation for the "insoluble cell pellet" (ICP) sample. The RP^{+}_{lab} ICP is therefore enriched in IB RP. The 9.4 T magnetic field, 8 kHz MAS frequency, and ~50 kHz radiofrequency fields are moderate and accessible for many NMR facilities, including those with an SSNMR probe (cost of about $100000) on an otherwise liquid-state NMR instrument.

Figure 1 displays results from HC RP quantitation. Panel a displays ^{13}C spectra of the [^{13}CO]Leu RP^{-}_{lab} and RP^{+}_{lab} samples. Although there are differences in plasmid and *E. coli* strain types among the samples, the corresponding spectra have similar aliphatic ^{13}C signal intensities in the 0−90 ppm region. The isotropic ^{13}CO signals of the spectra are near 175 ppm with much weaker spinning sideband ^{13}CO signals near 95 and 255 ppm. Relative to those of RP^{-}_{lab}, there are much larger ^{13}CO signals in the RP^{+}_{lab} spectra, which support significant expression of all the RPs. There are also differences among the ^{13}CO intensities of the different RP^{+}_{lab} samples, which support RP-dependent variation in expression. Panel b displays the expression levels in milligrams of RP per liter of culture as determined from (1) measurements of I_{Al^-} and I_{CO^-}, the integrated aliphatic and isotropic ^{13}CO intensities of the RP^{-}_{lab} spectrum, respectively; (2) measurements of the corresponding I_{Al^+} and I_{CO^+} of the RP^{+}_{lab} spectrum; and (3) calculation of the ^{13}CO intensity from [^{13}CO]Leu samples in the RP using $I_{Al^0}(I_{CO^+}/I_{Al^+} - I_{CO^-}/I_{Al^-})$, where I_{Al^0} is the value for a typical sample and $I_{Al^0}(I_{CO^+}/I_{Al^+})$ and $I_{Al^0}(I_{CO^-}/I_{Al^-})$ are ^{13}CO intensities normalized to NMR sample mass. The expression level is calculated using (3) and (i) an experimentally determined mole ^{13}CO/I_{CO} conversion factor; (ii) MW_{RP}/N_{Leu}, where N_{Leu} is the number of Leu residues in the RP sequence; and (iii) the NMR sample is from cells in a culture volume of ~25 mL.

The SSNMR-determined expression levels (panel b) are 100−450 mg of RP/L of culture. These levels are very high relative to the reported ~5 mg/L purified yields for Fgp41, FHA2, and HPI.[1,9,15] The most common current approach to assessing RP expression is SDS−PAGE. Panel c displays SDS−PAGE results for boiled ICPs. Relative to the background, there are clear bands for HPI and Hairpin and much fainter and more ambiguous bands for FHA2 and Fgp41. The variation of the RP band intensities in the SDS−PAGE gel is more reflective of differences in RP IB solubilization than differences in expression levels.

dx.doi.org/10.1021/bi4007034 | *Biochemistry* 2013, 52, 4285−4287

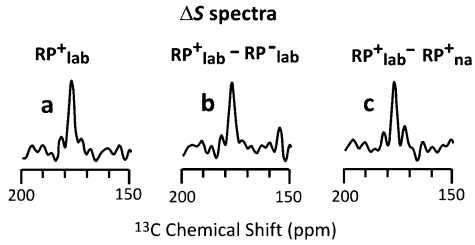

Figure 2. ^{13}CO ΔS spectra based on the S_0 and S_1 spectra of three different ICP samples. The RP^+_{lab} and RP^-_{lab} plasmids had and lacked the Fgp41 insert, respectively. The *lab* and *na* expression media contained [^{13}CO,^{15}N]Leu and unlabeled Leu, respectively. (a) The $\Delta S(RP^+_{lab}) = S_0 - S_1$ signal represents directly bonded ^{13}CO−^{15}N spin pairs of the RP^+_{lab} sample. (b) The $\Delta S = \Delta S(RP^+_{lab}) - \Delta S(RP^-_{lab})$ signal is from spin pairs of IB Fgp41. (c) The $\Delta S = \Delta S(RP^+_{lab}) - \Delta S(RP^+_{na})$ signal is from *lab* spin pairs of the RP^+_{lab} sample.

FHA2 and Fgp41 are membrane proteins, while HPI and Hairpin are not; therefore, the membrane RP IBs appear to be less well-solubilized. The SSNMR approach has the important advantage of being independent of IB solubilization.

The HCN approach is based on the $\Delta S = S_0 - S_1$ ^{13}CO REDOR difference spectrum of the RP^+_{lab} ICP sample. This spectrum is dominated by directly bonded *lab* ^{13}CO−^{15}N spin pairs in the IB RP. For Figure 2a, RP ≡ Fgp41, *lab* ≡ [^{13}CO,^{15}N]Leu, and the ΔS spectrum is mostly due to the N-terminal Leu residues of the six LL dipeptides in the Fgp41 sequence. One control is the RP^-_{lab} ΔS spectrum that is dominated by LL dipeptides of proteins other than Fgp41 produced during expression. However, there is no $\Delta S(RP^-_{lab})$ signal (Supporting Information), or equivalently, Figure 2b shows a $\Delta S(RP^+_{lab}) - \Delta S(RP^-_{lab})$ spectrum very similar to the $\Delta S(RP^+_{lab})$ spectrum that must therefore be dominated by the IB Fgp41 signals. Another control is the $\Delta S(RP^+_{na})$ spectrum of a sample prepared with unlabeled Leu and reflecting signals of *na* ^{13}CO−^{15}N spin pairs. However, there is little $\Delta S(RP^+_{na})$ signal as reflected in Figure 2c, where the $\Delta S(RP^+_{lab}) - \Delta S(RP^+_{na})$ spectrum is similar to the $\Delta S(RP^+_{lab})$ spectrum.

The HCN approach to quantitation of RP expression is detailed in the Supporting Information. For a particular RP^+_{lab} sample, the HC and HCN expression levels typically agree within a factor of 2. Quantitative labeling of the RP is assumed for both approaches, so the levels are likely lower limits of expression but probably within a factor of ∼2. Incomplete labeling will have a larger effect on HCN quantitation because the ΔS signal is only observed for dipeptides with both residues labeled.

Most of folded Fgp41 is a thermostable six-helix bundle that includes the six LL dipeptides.[13] The ΔS spectrum was previously obtained for [^{13}CO,^{15}N]Leu Fgp41 that had been purified, refolded, and reconstituted in membranes.[1] There was a single peak with a 178 ppm shift and a 3 ppm width that is consistent with folded helical structure. The $\Delta S(RP^+_{lab})$ spectrum of Fgp41 in IBs (Figure 2) is very similar and supports formation of folded Fgp41 structure in the IBs. For other RPs in IBs, the ΔS spectral widths are sometimes much broader, e.g., ∼7 ppm for HPI (Supporting Information). This breadth is consistent with unfolded RP structure in the IBs. SSNMR quantitation of RP expression by either the HC or HCN approaches is independent of the degree of RP folding in the IBs.

For all the RPs of this study, the SSNMR spectra demonstrated high expression, i.e., ≥100 mg of IB RP/L of culture, so the main obstacle to purified RP is solubilization of the IBs. For other RPs that are produced at much lower levels, SSNMR could also be applied to optimize RP production. [^{13}CO]RP^+_{lab} samples would be prepared with different growth and/or expression parameters and expression levels determined from the ^{13}CO intensities. In summary, this paper describes general, inexpensive, rapid, and straightforward SSNMR approaches to RP quantitation in whole cells and cell extracts without purification.

■ ASSOCIATED CONTENT

Supporting Information

Additional NMR spectra, cell growth and labeling, sample preparation, plasmids and protein sequences, NMR methods, and analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: weliky@chemistry.msu.edu. Phone: (517) 355-9715.

Funding

Supported by National Institutes of Health Grant AI047153.

Notes

The authors declare no competing financial interests.

■ ACKNOWLEDGMENTS

Drs. R. Mackin, Y.-K. Shin, and K. Sackett provided plasmids.

■ REFERENCES

(1) Vogel, E. P., Curtis-Fisk, J., Young, K. M., and Weliky, D. P. (2011) *Biochemistry 50*, 10013−10026.
(2) Miles, A. P., and Saul, A. (2009) in *Protein Protocols Handbook*, 3rd ed., pp 487−496, Humana, Totowa, NJ.
(3) Gross-Selbeck, S., Margreiter, G., Obinger, C., and Bayer, K. (2007) *Biotechnol. Prog. 23*, 762−766.
(4) Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) *J. Am. Chem. Soc. 130*, 12568−12569.
(5) Wang, J., Balazs, Y. S., and Thompson,, L. K. (1997) *Biochemistry 36*, 1699−1703.
(6) Kim, S. J., Cegelski, L., Stueber, D., Singh, M., Dietrich, E., Tanaka, K. S. E., Parr, T. R., Far, A. R., and Schaefer, J. (2008) *J. Mol. Biol. 377*, 281−293.
(7) Reckel, S., Lopez, J. J., Lohr, F., Glaubitz, C., and Dotsch, V. (2012) *ChemBioChem 13*, 534−537.
(8) Zhou, X. X., and Cegelski, L. (2012) *Biochemistry 51*, 8143−8153.
(9) Mackin, R. B., and Choquette, M. H. (2003) *Protein Expression Purif. 27*, 210−219.
(10) Yang, Y. W., Hua, Q. X., Liu, J., Shimizu, E. H., Choquette, M. H., Mackin, R. B., and Weiss, M. A. (2010) *J. Biol. Chem. 285*, 7847−7851.
(11) Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) *Biochemistry 48*, 2714−2722.
(12) Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) *J. Mol. Biol. 397*, 301−315.
(13) Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) *J. Struct. Biol. 126*, 131−144.
(14) Curtis-Fisk, J., Preston, C., Zheng, Z. X., Worden, R. M., and Weliky, D. P. (2007) *J. Am. Chem. Soc. 129*, 11320−11321.
(15) Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) *Protein Expression Purif. 61*, 212−219.
(16) Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) *J. Biol. Chem. 286*, 13226−13234.
(17) Chen, J., Skehel, J. J., and Wiley, D. C. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 8967−8972.
(18) Guillion, T., and Schaefer, J. (1989) *J. Magn. Reson. 81*, 196−200.
(19) Yang, J., Parkanzky, P. D., Bodner, M. L., Duskin, C. G., and Weliky, D. P. (2002) *J. Magn. Reson. 159*, 101−110.
(20) Tong, K. I., Yamamoto, M., and Tanaka, T. (2008) *J. Biomol. NMR 42*, 59−67.