

## Protein $^{19}\text{F}$ NMR in *Escherichia coli*

Conggang Li,<sup>†</sup> Gui-Fang Wang,<sup>†</sup> Yaqiang Wang,<sup>†</sup> Rachel Creager-Allen,<sup>‡</sup>  
Evan A. Lutz,<sup>†</sup> Heidi Scronce,<sup>†</sup> Kristin M. Slade,<sup>†</sup> Rebecca A.S. Ruf,<sup>†</sup>  
Ryan A. Mehl,<sup>§</sup> and Gary J. Pielak<sup>\*,†,‡,||</sup>

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, Department of Chemistry, Franklin and Marshall College, P.O. Box 3003, Lancaster, Pennsylvania, 17604, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, North Carolina 27599

Received September 22, 2009; E-mail: gary\_pielak@unc.edu

**Abstract:** Although overexpression and  $^{15}\text{N}$  enrichment facilitate the observation of resonances from disordered proteins in *Escherichia coli*,  $^{15}\text{N}$  enrichment alone is insufficient for detecting most globular proteins. Here, we explain this dichotomy and overcome the problem while extending the capability of in-cell NMR by using  $^{19}\text{F}$ -labeled proteins. Resonances from small ( $\sim 10$  kDa) globular proteins containing the amino acid analogue 3-fluoro-tyrosine can be observed in cells, but for larger proteins the  $^{19}\text{F}$  resonances are broadened beyond detection. Incorporating the amino acid analogue trifluoromethyl-L-phenylalanine allows larger proteins (up to 100 kDa) to be observed in cells. We also show that site-specific structural and dynamic information about both globular and disordered proteins can be obtained inside cells by using  $^{19}\text{F}$  NMR.

Most proteins function inside cells under crowded and complex conditions, where the concentration of macromolecules can reach  $\sim 400$  g/L.<sup>1–3</sup> Studying proteins in the cellular interior, although difficult, is important for understanding how environments affect functions. In-cell NMR provides a means to assess protein structure, function, and interactions with other proteins, DNA, and small molecules at atomic resolution in living cells.<sup>4–19</sup> Recently, the high-resolution NMR structure of a small, 66-residue protein in the cytosol of *Escherichia coli* has been reported.<sup>20</sup>

The success of in-cell NMR experiments depends on overcoming several obstacles. As currently practiced, in-cell NMR in *E. coli* requires protein overexpression, which may lessen its biological significance. Current practice also requires growth on nutrients enriched in NMR-active nuclei, usually  $^{15}\text{N}$  or  $^{13}\text{C}$ . Normal metabolism of these nutrients causes a background spectrum that obscures signals from the protein being studied. Proteins that leak from the cell also cause artifacts.<sup>21</sup> Further-

more, the crowded intracellular environment broadens resonances from globular proteins, lowering the sensitivity of NMR experiments. For instance, specific methyl labeling, a technique typically used only for large proteins, was required to obtain sufficient long-range structure restraints for the small protein cited above. Augustus et al. showed that the repressor protein MetJ is completely undetectable in *E. coli* because of weak, nonspecific DNA binding,<sup>4</sup> and indetectability has been reported for other globular proteins.<sup>11</sup>

<sup>†</sup> Department of Chemistry, University of North Carolina at Chapel Hill.

<sup>‡</sup> Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill.

<sup>§</sup> Department of Chemistry, Franklin and Marshall College.

<sup>||</sup> Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill.

(1) Luby-Phelps, K. *Int. Rev. Cytol.* **2000**, *192*, 189–221.

(2) Malmström, J.; Beck, M.; Schmidt, A.; Lange, V.; Deutsch, E. W.; Aebersold, R. *Nature* **2009**, *460*, 762–765.

(3) Zimmerman, S. B.; Trach, S. O. *J. Mol. Biol.* **1991**, *222*, 599–620.

(4) Augustus, A. M.; Reardon, P. N.; Spicer, L. D. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5065–5069.

(5) Brindle, K. M.; Williams, S.-P.; Boulton, M. *FEBS Lett.* **1989**, *255*, 121–124.

(6) Burz, D. S.; Dutta, K.; Cowburn, D.; Shekhtman, A. *Nat. Protoc.* **2006**, *1*, 146–152.

(7) Dedmon, M. M.; Patel, C. N.; Young, G. B.; Pielak, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12681–12684.

(8) Hubbard, J. A.; MacLachlan, L. K.; King, G. W.; Jones, J. J.; Fosberry, A. P. *Mol. Microbiol.* **2003**, *49*, 1191–1200.

(9) Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M. *Nature* **2009**, *458*, 106–109.

(10) McNulty, B. C.; Young, G. B.; Pielak, G. J. *J. Mol. Biol.* **2006**, *355*, 893–897.

(11) Pielak, G. J.; Li, C.; Miklos, A. C.; Schlesinger, A. P.; Slade, K. M.; Wang, G.; Zigaretanu, I. G. *Biochemistry* **2008**, *48*, 226–234.

(12) Sakai, T.; Tochio, H.; Tenno, T.; Ito, Y.; Kokubo, T.; Hiroaki, H.; Shirakawa, M. *J. Biomol. NMR* **2006**, *36*, 179–188.

(13) Selenko, P.; Serber, Z.; Gadea, B.; Ruderman, J.; Wagner, G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11904–11909.

(14) Selenko, P.; Wagner, G. *Nat. Methods* **2006**, *3*, 80–81.

(15) Serber, Z.; Dötsch, V. *Biochemistry* **2001**, *40*, 14317–14323.

(16) Serber, Z.; Keatinge-Clay, A. T.; Ledwidge, R.; Kelly, A. E.; Miller, S. M.; Dötsch, V. *J. Am. Chem. Soc.* **2001**, *123*, 2446–2447.

(17) Serber, Z.; Selenko, P.; Hansel, R.; Reckel, S.; Lohr, F.; Ferrell, J. E., Jr.; Wagner, G.; Dötsch, V. *Nat. Protoc.* **2006**, *1*, 2701–2709.

(18) Serber, Z.; Straub, W.; Corsini, L.; Nomura, A. M.; Shimba, N.; Craik, C. S.; Ortiz de Montellano, P.; Dötsch, V. *J. Am. Chem. Soc.* **2004**, *126*, 7119–7125.

(19) Xie, J.; Thapa, R.; Reverdatto, S.; Burz, D. S.; Shekhtman, A. *J. Med. Chem.* **2009**, *52*, 3516–3522.

(20) Sakakibara, D.; Sasaki, A.; Ikeya, T.; Hamatsu, J.; Hanashima, T.; Mishima, M.; Yoshimasu, M.; Hayashi, N.; Mikawa, T.; Walchli, M.; Smith, B. O.; Shirakawa, M.; Guntert, P.; Ito, Y. *Nature* **2009**, *458*, 102–105.

(21) Pielak, G. J. *Biochemistry* **2007**, *46*, 8206.

Because natural proteins contain no fluorine, this 100% abundant spin- $1/2$  nucleus with its high sensitivity (83% of  $^1\text{H}$ ), spectral simplicity, and large chemical shift range is attractive for protein NMR in vitro and in cells.<sup>22–25</sup>  $^{19}\text{F}$  in-cell NMR was first applied to detect protein mobility in the yeast *Saccharomyces cerevisiae*,<sup>26–28</sup> and a preliminary study has been reported in *E. coli*.<sup>29</sup> Here, we describe detailed studies using this bacterium.

We examine one disordered and five globular proteins containing  $^{15}\text{N}$  and/or  $^{19}\text{F}$ . More specifically, we incorporate the fluorinated amino acid analogues, 3-fluoro-tyrosine (3FY), and trifluoromethyl-L-phenylalanine (tfmF) into proteins ranging in size from 7 to 100 kDa.

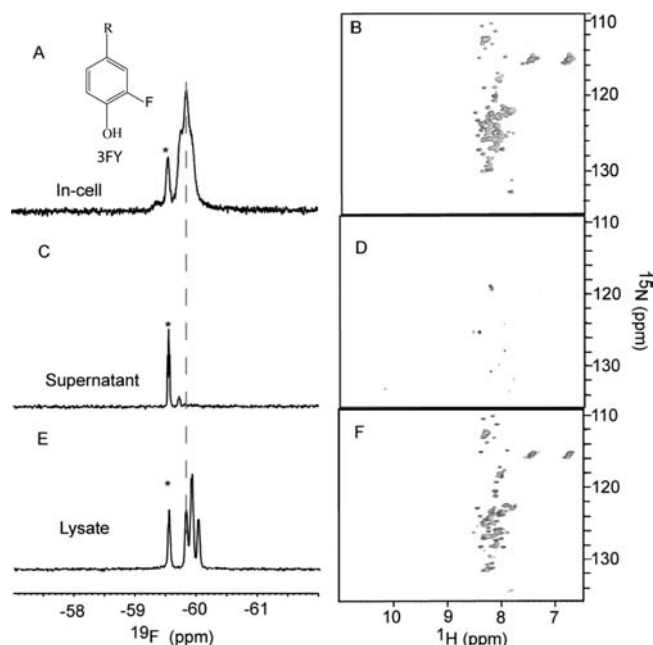
## Materials and Methods

**Expression Systems.** The ubiquitin (UBQ),<sup>6</sup> calmodulin (CAM),<sup>30</sup> and  $\alpha$ -synuclein ( $\alpha$ SYN) expression systems were gifts from Alexander Shekhtman (State University of New York at Albany), Anthony Persechini (University of Missouri, Kansas City), and Peter Lansbury (Harvard), respectively. pET28a plasmids (Novagen) containing the gene for the truncated chymotrypsin inhibitor 2 (CI2)<sup>31</sup> or the PDZ3<sup>32</sup> domain were a gift from Andrew Lee (University of North Carolina at Chapel Hill). The GFP and histidinol dehydrogenase (HDH) expression system has been described.<sup>29,33</sup> For  $^{15}\text{N}$  enrichment, the plasmids were transformed into BL-21(DE3-Gold) competent cells. The CAM, CI2, and PDZ3 transformants were spread onto Luria broth agar plates containing 60  $\mu\text{g}/\text{mL}$  kanamycin, and the others were spread onto plates containing 60  $\mu\text{g}/\text{mL}$  ampicillin.

**$^{15}\text{N}$  Enrichment and 3FY Labeling.** The procedure was similar to that described by Khan et al.<sup>34</sup> and Li et al.<sup>24</sup> Details are given in the Supporting Information.

**tfmF Labeling.** Amber stop codons (TAG) were incorporated at the sites for tfmF labeling by using site-directed mutagenesis (QuickChange, Stratagene) of the target genes, which are present in the arabinose-inducible expression vector, pBAD. The labeling procedure was similar to that described by Hammill et al.<sup>35</sup> Details are given in the Supporting Information.

**Preparing for In-Cell NMR.** Cultures (usually  $\sim 100$  mL) were centrifuged at 1 200g for 30 min at room temperature. The cell pellets were resuspended in 2 mL of LB media. The samples, comprising 90:10 mixtures of cell slurry: $\text{D}_2\text{O}$ , were placed in 5 mm NMR tubes for data acquisition. Supernatants were collected by centrifugation (Eppendorf, model 5418, 2 000g for 10 min) after



**Figure 1.**  $^{19}\text{F}$  spectra of 3FY-labeled  $\alpha$ SYN (left panels) and  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectra of  $^{15}\text{N}$ -enriched  $\alpha$ SYN (right panels). Panels A and B show in-cell spectra. The insert in panel A shows the structure of 3FY. Panels C and D show spectra of supernatants collected immediately after completing the in-cell spectra. Panels E and F show spectra of supernatants from the clear lysates. The asterisks indicate the free 3FY resonances. The dashed vertical line shows the upfield shift on cell lysis.

the experiments to assess leakage.<sup>21</sup> The pellets were resuspended in buffer (50 mM Tris, pH 8.0) to a final volume of 1 mL. Lysates were made from the resuspended pellets by sonication (Fisher Scientific, Sonic Dismembrator Model 500) on ice for 10 min with a duty cycle of 2 s on, 5 s off. The lysate was collected after centrifugation at 16 000g for 10 min. Viscosities were measured with a Viscolite 700 viscometer (Hydramotion Ltd., England).

**NMR.**  $^{15}\text{N}$ - $^1\text{H}$ -HSQC spectra were acquired on a cold-probe equipped Varian Inova 500 MHz spectrometer at 25  $^{\circ}\text{C}$ . The  $^1\text{H}$  dimension had a sweep width of 8 401 Hz and comprised 1 024 complex points. The  $^{15}\text{N}$  dimension had a sweep width of 2 200 Hz and comprised 64 complex points. The data were processed with *NMRPipe*<sup>36</sup> and *NMRDraw*.<sup>37</sup>  $^{19}\text{F}$  spectra were acquired at 37  $^{\circ}\text{C}$  on a Varian Inova 600 MHz spectrometer equipped with a 5 mm  $^{19}\text{F}$  {H} z-gradient probe. The spectra comprised 128 to 2 048 transients, a 30 kHz sweep width, and a 2 s delay before acquisition.  $^{19}\text{F}$  chemical shifts are referenced to trifluoroethanol at 0 ppm.

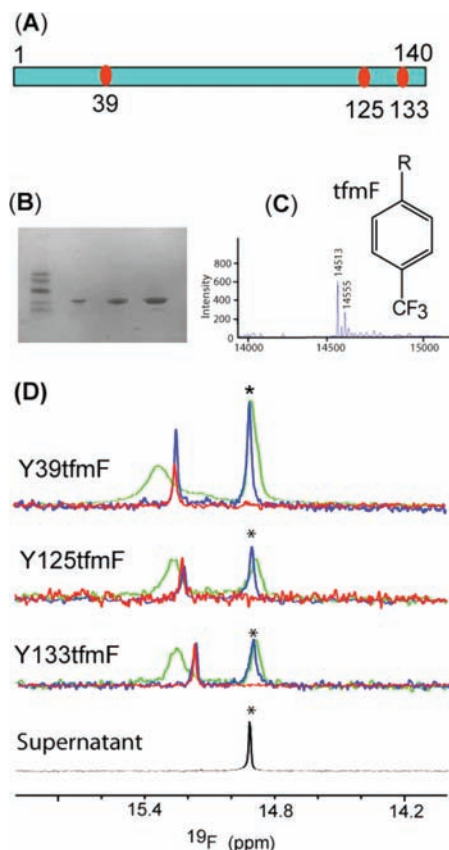
## Results

**3FY-Labeled,  $^{15}\text{N}$  Enriched  $\alpha$ -Synuclein ( $\alpha$ SYN).** There are four tyrosines in this 140 residue, intrinsically disordered protein, one at position 39 and three near the C-terminus, at positions 125, 133, and 136. We labeled all of these residues with 3FY. As shown in part A of Figure 1, the  $^{19}\text{F}$  spectrum of the cell slurry shows a broad protein resonance at  $\sim -60$  ppm and a sharp resonance from free 3FY at  $-59.6$  ppm. The assignment of the  $\alpha$ SYN resonance was confirmed by comparison to the spectrum of the purified protein. The assignment of the free 3FY resonance was confirmed by comparison to the supernatant spectrum. The  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of the cell slurry (part B of Figure 1) shows numerous  $\alpha$ SYN crosspeaks, consistent

- (22) Cellitti, S. E.; et al. *J. Am. Chem. Soc.* **2008**, *130*, 9268–9281.  
 (23) Danielson, M. A.; Falke, J. *J. Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 163–195.  
 (24) Li, C.; Lutz, E. A.; Slade, K. M.; Ruf, R. A.; Wang, G.; Pielak, G. J. *Biochemistry* **2009**, *48*, 8578–8584.  
 (25) Li, H.; Frieden, C. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11993–11998.  
 (26) Haggie, P. M.; Brindle, K. M. *J. Biol. Chem.* **1999**, *274*, 3941–3945.  
 (27) Williams, S. P.; Fulton, A. M.; Brindle, K. M. *Biochemistry* **1993**, *32*, 4895–4902.  
 (28) Williams, S. P.; Haggie, P. M.; Brindle, K. M. *Biophys. J.* **1997**, *72*, 490–498.  
 (29) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166.  
 (30) Pervushin, K. V.; Wider, G.; Riek, R.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9607–9612.  
 (31) Charlton, L. M.; Barnes, C. O.; Li, C.; Orans, J.; Young, G. B.; Pielak, G. J. *J. Am. Chem. Soc.* **2008**, *130*, 6826–6830.  
 (32) Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18249–18254.  
 (33) Slade, K. M.; Baker, R.; Chua, M.; Thompson, N. L.; Pielak, G. J. *Biochemistry* **2009**, *48*, 5083–5089.  
 (34) Khan, F.; Kuprov, I.; Craggs, T. D.; Hore, P. J.; Jackson, S. E. *J. Am. Chem. Soc.* **2006**, *128*, 10729–10737.  
 (35) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. *Nat. Protoc.* **2007**, *2*, 2601–2607.

(36) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.

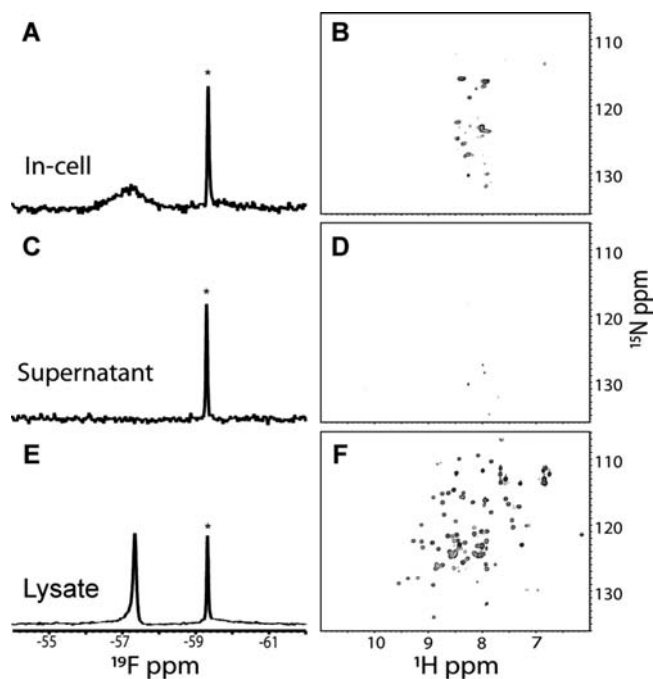
(37) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603–614.



**Figure 2.** Sites of tfmF incorporation in  $\alpha$ -synuclein (A). SDS-PAGE of the three purified tfmF-labeled  $\alpha$ -synucleins (B). ESI-mass spectrum of tfmF39-labeled  $\alpha$ -synuclein (C). The insert shows the structure of tfmF. <sup>19</sup>F spectra of labeled synuclein (D). Spectra from cell slurries are shown in green. Spectra from clear lysates are shown in blue. Spectra from purified tfmF-labeled proteins are shown in red. Spectra from supernatants collected immediately after the in-cell NMR experiments are shown in black. The asterisks indicate the free tfmF resonances.

with previous work.<sup>10,38</sup> To check for leakage, we subjected the cell slurry to centrifugation and examined the supernatant. The presence of only the free 3FY resonance in the <sup>19</sup>F spectrum (part C of Figure 1) and the near absence of crosspeaks in the HSQC spectrum (part D of Figure 1) indicate that little or no  $\alpha$ SYN had leaked. The cells were then lysed by sonication, the cellular debris removed by centrifugation, and the clear lysate examined by NMR. The <sup>19</sup>F resonances sharpened (part E of Figure 1), revealing three  $\alpha$ SYN peaks that shifted upfield by  $\sim 0.1$  ppm. The <sup>19</sup>F spectrum has been assigned.<sup>24</sup> The middle peak comprises 3FY resonances from residues 39 and 125. The downfield and upfield resonances correspond to residues 136 and 139, respectively. The crosspeaks in the HSQC spectrum of the lysate are sharper than those from the cell slurry but the spectrum is essentially unchanged. The limited chemical shift dispersion of the <sup>19</sup>F and <sup>1</sup>H resonances show that  $\alpha$ SYN is disordered in cells, consistent with other work.<sup>10,38</sup>

**tfmF-Labeled  $\alpha$ SYN.** To overcome the incomplete resolution of the four <sup>19</sup>F resonances from 3FY-labeled tyrosines in cells (part A of Figure 2), we labeled the protein with tfmF at three of the four tyrosines by using an orthogonal aminoacyl synthase system.<sup>29</sup> Before performing in-cell NMR experiments, we



**Figure 3.** <sup>19</sup>F- and <sup>1</sup>H-<sup>15</sup>N HSQC spectra of <sup>15</sup>N-enriched, 3FY-labeled UBQ. The panels are labeled as described in the caption to Figure 1.

assessed the system by purifying the labeled protein (part B of Figure 2) and using mass spectrometry to confirm the expected 52 Da increase in mass, from 14 461 to 14 513 Da. The peak at 14 555 is the labeled form of acetylated  $\alpha$ SYN.

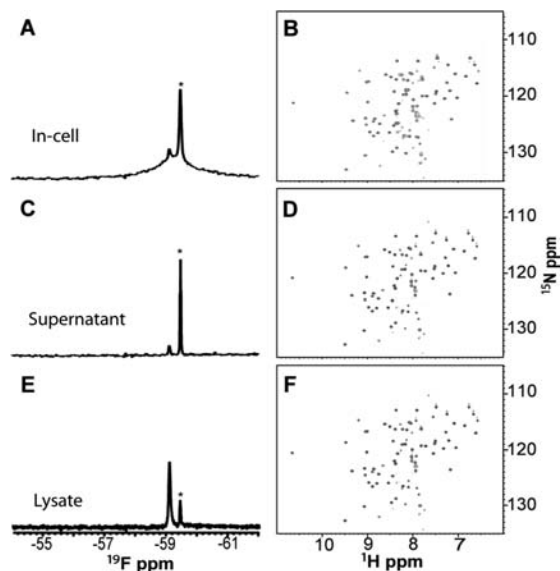
The in-cell <sup>19</sup>F spectra for proteins labeled at positions 39, 125, and 133 are shown as green traces in part D of Figure 2. The tfmF 39 resonance is broader than the tfmF 125 and 133 resonances in the cell slurry. The resonances from the lysates and from the purified proteins are narrower and shift upfield by  $\sim 0.1$  compared to those from the cell slurry. Only a free tfmF resonance is observed in the supernatants, showing that the proteins do not leak.

**<sup>19</sup>F-Labeled, <sup>15</sup>N-Enriched Ubiquitin (UBQ).** This 8 kDa globular protein has one tyrosine. Part A of Figure 3 shows the <sup>19</sup>F spectrum of the cell slurry. The spectrum contains a sharp free 3FY resonance and a broad protein resonance. The identity of these resonances was confirmed by comparisons to spectra of the purified protein and the supernatant. Part B of Figure 3 shows the HSQC spectrum from the slurry. Only metabolite signals<sup>39</sup> are observed. Parts C and D of Figure 3 show the <sup>19</sup>F and HSQC spectra from the supernatant collected immediately after the in-cell NMR experiment. Only the free 3FY resonance is present in the <sup>19</sup>F spectrum, and the HSQC spectrum is nearly blank. These observations show that UBQ has not leaked from the cells. The cells were then lysed. The <sup>19</sup>F lysate spectrum (part E of Figure 3) shows a single sharp protein resonance and the HSQC spectrum (part F of Figure 3) closely resembles that of pure UBQ.<sup>40</sup>

We also collected HSQC data on the globular, 11 kDa PDZ3 domain of PSD95<sup>32,41</sup> in cells. Like UBQ, the PDZ3 domain is expressed at mM levels (Figure S1 and Table S1 of the

(38) Li, C.; Charlton, L. M.; Lakkavaram, A.; Seagle, C.; Wang, G.; Young, G. B.; Macdonald, J. M.; Pielak, G. J. *J. Am. Chem. Soc.* **2008**, *130*, 6310–6311.

(39) Bryant, J. E.; Lecomte, J. T.; Lee, A. L.; Young, G. B.; Pielak, G. J. *Biochemistry* **2005**, *44*, 9275–9279.  
 (40) Schneider, D. M.; Dellwo, M. J.; Wand, A. J. *Biochemistry* **1992**, *31*, 3645–3652.  
 (41) Law, A. B.; Fuentes, E. J.; Lee, A. L. *J. Am. Chem. Soc.* **2009**, *131*, 6322–6323.



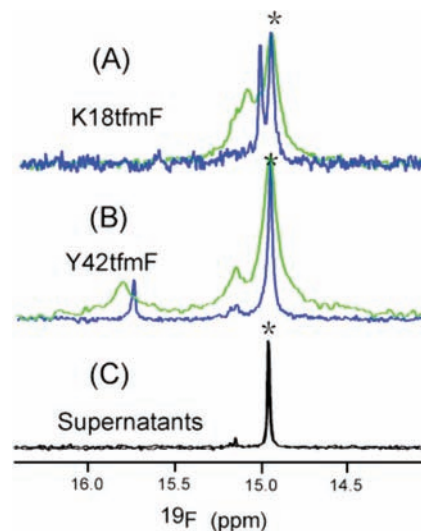
**Figure 4.**  $^{19}\text{F}$  spectra of 3FY-labeled CI2 (left panels) and  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -enriched CI2. The panels are labeled as described in the caption to Figure 1.

Supporting Information) but its HSQC spectrum cannot be obtained from the cell slurry. The protein signals, however, appear upon lysis (Figure S2 of the Supporting Information).

We do not understand our inability to reproduce the published results on UBQ, which has been reported to yield high-resolution spectra in *E. coli*.<sup>6,42</sup> Our use of a different growth medium is not the reason because we obtain similar results to those shown in Figure 3 when we use the media described in the publications. We also tried expressing the protein at different temperatures without success. Our studies were conducted on a cold-probe equipped 500 MHz instrument. Lack of sensitivity does not explain our inability to detect UBQ in cells because we obtain the same results with a cold-probe equipped 700 MHz spectrometer.

Freezing cells prior to in-cell NMR studies has been suggested.<sup>6</sup> We prepared another  $^{15}\text{N}$ -enriched UBQ sample for in-cell experiments but stored the sample at  $-20^\circ\text{C}$  overnight. The sample was thawed and used to collect an in-cell spectrum. The spectrum of native UBQ,<sup>40</sup> which is not observed in a fresh sample (part B of Figure 3), is visible in the previously frozen sample (Figure S3 of the Supporting Information). Storing the sample at  $-80^\circ\text{C}$  gives the same result. Adding 10% (v/v) glycerol decreases, but does not always prevent, leakage. We conclude that cells should not be frozen if they are to be studied by using in-cell NMR.

**3FY-Labeled,  $^{15}\text{N}$ -Enriched Chymotrypsin Inhibitor 2 (CI2).** This 7 kDa globular protein has one tyrosine. Part A of Figure 4 shows the  $^{19}\text{F}$  spectrum of 3FY-labeled cell slurry. Three resonances are observed. The sharpest resonance is from free 3FY. The other two resonances are from CI2. Both have a chemical shift of  $-59.2$  ppm. One protein resonance is broad with a width at half height of  $\sim 1.5$  ppm. The other resonance is sharper and superimposed on the broad resonance. The HSQC spectrum of the cell slurry (part B of Figure 4) shows a spectrum almost identical to that of purified CI2.<sup>31</sup> The spectrum from the supernatant collected immediately after the in-cell experiment contains a resonance from both free 3FY and 3FY-labeled



**Figure 5.**  $^{19}\text{F}$  spectra of K18tfmF CI2 in cells (green) and lysates (blue) (A), Y42tfmF CI2 in cells and in lysates (B), and supernatants collected after the in-cell NMR experiments (C). The asterisks indicate the free tfmF resonances.

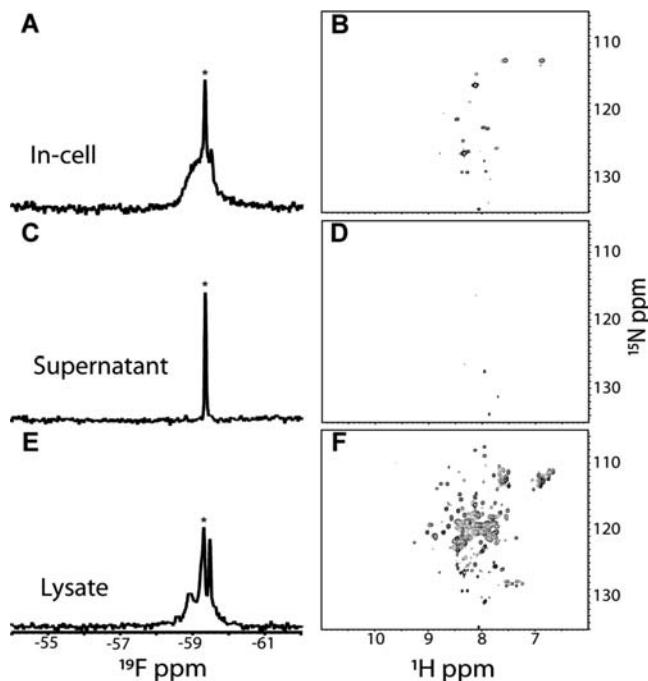
CI2 (part C of Figure 4). The HSQC spectrum of the supernatant (part D of Figure 4) is almost identical to the spectrum from the cell slurry (part A of Figure 4). These data show that CI2 has leaked from the cells, consistent with previous work.<sup>38</sup> After lysis (part E of Figure 4), only free 3FY and a single sharp resonance from the labeled protein is observed. The HSQC spectrum of the lysate is identical to the HSQC spectrum from the cell slurry. Comparing the three  $^{19}\text{F}$  spectra suggests that the broad resonance at  $-59.2$  ppm in the cell slurry is intracellular CI2 and the superimposed sharper resonance is from CI2 that has leaked from the cells.

**tfmF-Labeled CI2.** Figure 5 shows the  $^{19}\text{F}$  spectra of CI2 labeled at positions 18 and 42 in cells and lysates. The protein resonances have a width at half height of  $\sim 0.20$  ppm in cells. They shift upfield by  $0.15$ – $0.20$  ppm and narrow to  $\sim 0.03$  ppm upon lysis. There are no protein signals from the supernatants collected after the NMR experiments, indicating that tfmF-labeled CI2 does not leak. This result is surprising considering the results from the 3FY-labeled protein (Figure 4). As discussed below, a lower expression level may explain the absence of leakage. The small signals near the free tfmF may be a degradation product of labeled CI2 or a tfmF metabolite.

**3FY-Labeled,  $^{15}\text{N}$ -Enriched Calmodulin (CAM).** This 16 kDa two-lobed globular protein has two tyrosine residues. Part A of Figure 6 shows the  $^{19}\text{F}$  spectrum of the 3FY-labeled cell slurry. Three resonances are evident. The sharp resonance is from free 3FY. The other two, one on either side of the 3FY resonance, are from the protein.<sup>43</sup> The HSQC spectrum from the slurry (part B of Figure 6) shows only metabolite signals.<sup>38,39</sup> Parts C and D of Figure 6 show the  $^{19}\text{F}$  spectrum and the HSQC spectrum from the supernatant collected immediately after the in-cell NMR experiment. Only free 3FY is observed in the  $^{19}\text{F}$  spectrum, and the HSQC spectrum is nearly devoid of cross-peaks, indicating that CAM does not leak from the cells. Part E of Figure 6 shows the  $^{19}\text{F}$  spectrum of the clear lysate. The broad protein resonances observed in the cell slurry narrow on lysis but the width at half height for the broadest resonance is

(42) Burz, D. S.; Shekhtman, A. *PLoS ONE* **2008**, *3*, e2571.

(43) Kitevski-LeBlanc, J. L.; Evanics, F.; Prosser, R. S. *J. Biomol. NMR* **2009**, *45*, 255–264.



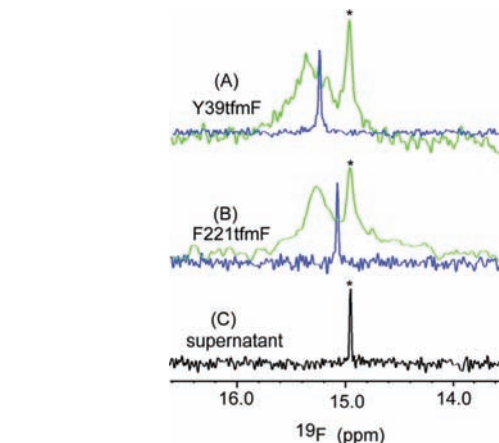
**Figure 6.** <sup>19</sup>F- and <sup>1</sup>H–<sup>15</sup>N HSQC spectra of <sup>15</sup>N-enriched, 3FY-labeled CAM. The panels are labeled as described in the caption to Figure 1.

still >0.5 ppm. The observation of CAM crosspeaks<sup>43,44</sup> in the HSQC spectrum of the lysate (part F of Figure 6) proves that detectable amounts of the protein are present. The HSQC spectrum of the lysate also shows that CAM is not fully Ca<sup>2+</sup> loaded.<sup>43</sup>

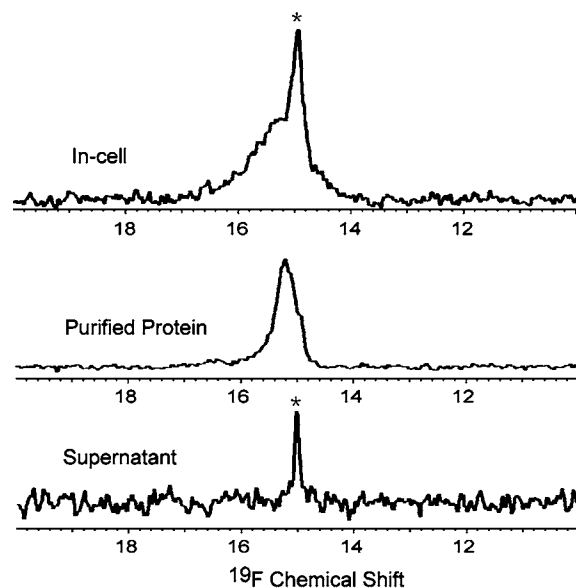
**3FY-Labeled, <sup>15</sup>N-Enriched Green Fluorescence Protein (GFP).** This 27 kDa globular protein contains 12 tyrosines. The resonances in the <sup>19</sup>F and <sup>15</sup>N–<sup>1</sup>H HSQC spectra are too broad to detect in cells and lysates (Figure S4 of the Supporting Information) but SDS-PAGE analysis and the fluorescence of the samples show that the protein is overexpressed (Figure S1 and Table S1 of the Supporting Information).

**tfmF-Labeled GFP.** tfmF might be a better label for larger globular proteins because the trifluoromethyl group adds rotational motion that is independent of molecular tumbling. The green traces in Figure 7 show the <sup>19</sup>F spectra of GFP labeled at position 39 and position 221 in cells. The <sup>19</sup>F resonances from the two proteins in cells are broad, with widths at half height of ~0.4 ppm, but observable. The corresponding resonances from the purified protein are narrower with widths of <0.1 ppm. The only resonance in the supernatant from the cell slurry is from free tfmF, which shows that labeled GFP does not leak. Lysis caused an upfield shift of 0.10–0.15 ppm.

**tfmF-Labeled Histidinol Dehydrogenase (HDH).**<sup>29</sup> We applied the tfmF labeling method to this 98 kDa homodimer. The <sup>19</sup>F spectra are presented in Figure 8. As shown by comparisons to spectra for the purified protein and the supernatant, the sharp resonance in the cell slurry spectrum is from free tfmF and the broad resonance (width at half height of ~1.0 ppm) is from HDH. The only resonance in



**Figure 7.** <sup>19</sup>F spectra of tfmF 39 labeled GFP in cells (green) and the purified protein in solution (blue) (A), tfmF 221 labeled GFP in cells and the purified protein in solution (B), and supernatants collected after the in-cell NMR experiments (C). The asterisks indicate the free tfmF resonances.



**Figure 8.** <sup>19</sup>F spectra of L225tfmF HDH. In-cell sample (A), purified protein (B), supernatant collected after the in-cell NMR experiments (C). The asterisks indicate the resonance from free tfmF.

the supernatant is from free tfmF, showing that labeled HDH does not leak.

### Discussion

We used NMR to study six proteins enriched in <sup>15</sup>N and/or labeled with <sup>19</sup>F in *E. coli* cells. The proteins are present in the cytoplasm [although some αSYN<sup>33</sup> and CI2 (Figure S5 of the Supporting Information) is periplasmic]. Two <sup>19</sup>F labeling strategies were used. One strategy, incorporating 3FY in place of tyrosine, was accomplished by expressing the protein in <sup>15</sup>N-enriched minimal media containing 3FY, phenylalanine, tryptophan, and *N*-(phosphomonoethyl) glycine.<sup>34</sup> The other strategy involved an orthogonal tRNA synthase system<sup>29</sup> to replace residues with tfmF.

A 1D <sup>19</sup>F spectrum can be acquired in minutes (compared to an hour for <sup>15</sup>N–<sup>1</sup>H HSQC spectra shown here), which allows the study of proteins near their physiological concentrations. We estimate an intracellular concentration of the tfmF-labeled

(44) Jaren, O. R.; Kranz, J. K.; Sorensen, B. R.; Wand, A. J.; Shea, M. A. *Biochemistry* **2002**, *41*, 14158–14166.

proteins of 50 to 100  $\mu\text{M}$  from the areas of the free *tfmF* and the protein resonances, the *tfmF* concentration in the media, and the fact that the cells occupy half the slurry volume of NMR samples. This concentration equals that of the most abundant soluble *E. coli* proteins.<sup>45,46</sup> Furthermore, these <sup>19</sup>F experiments can be performed as a function of time to obtain data on signal transduction and metabolism.

<sup>19</sup>F labeling is well suited to assess leaking. Controls must be performed to ensure the protein of interest is inside the cells during the NMR experiment.<sup>21</sup> For CI2 (Figure 4), we see a sharp 3FY resonance from leaked protein and a broad resonance from intracellular protein. By comparing the signal intensity in the supernatant (leaked CI2) to that in the lysate (total CI2), we estimate that 5–10% of the protein leaks from the cells. Importantly, this small fraction of leaked CI2 accounts for 100% of the CI2 signal in the HSQC spectrum of the cell slurry.

To assess the effect of the expression system on leakage, we repeated the experiments in the same *E. coli* strain [BL21(DE3)] with CI2 under control of the *araBAD* promoter<sup>33</sup> rather than the T7 promoter. We did not observe CI2 resonances in the HSQC spectrum from the cell slurry, but we did observe CI2 resonances in the lysate (Figure S6 of the Supporting Information). These experiments confirm the indetectability of CI2 HSQC spectra in cells and suggest that the expression system affects protein leakage. For our pBAD experiment, however, leaked protein might not be detected because CI2 expression was also lower [ $\sim 0.2$  mM compared to 1 mM in BL21(DE3)]. We repeated the experiment using strain DH10B, but expression was so low that CI2 crosspeaks were not observable even in the lysate. Additionally, in opposition to what has been recommended,<sup>6</sup> storing cells in the freezer should be avoided. As we have shown for UBQ (Figure S3 of the Supporting Information), the freeze–thaw cycle disrupts a fraction of the cells, spilling the enriched protein into the surrounding dilute solution. We have shown elsewhere that encapsulating the cells controls leakage.<sup>38</sup> In summary, although more experiments are required to deconvolute the effects of the expression system, expression level, and strain, <sup>19</sup>F provides a straightforward assay for leakage.

<sup>19</sup>F labeling extends the utility of in-cell NMR for studying intrinsically disordered proteins. For  $\alpha\text{SYN}$ , we observe both backbone <sup>15</sup>N and side chain <sup>19</sup>F signals in cells, although the resonances from the cells are broader than those from lysates and dilute samples (Figure 1). Because of the limited chemical shift dispersion of disordered proteins, *tfmF* labeling (Figure 2) is preferred over 3FY labeling (Figure 1) because any natural, ribosomally encoded, amino acid can be replaced with *tfmF*. Moreover, *tfmF* labeling provides dynamic information at a single site. For  $\alpha\text{-synuclein}$  in cells, the *tfmF* 39 resonance is broader than the C-terminal *tfmF* resonances. This observation indicates constrained motion at position 39, consistent with reports that position 39 has residual structure while the C-terminal region is completely disordered.<sup>24,47,48</sup> By the same reasoning, the increased width of the *tfmF* 39 resonance in GFP compared to the *tfmF* 221 resonance (Figure 7) indicates that the side chain at position 221 is more mobile than the side chain

at position 39. Such dynamic information is masked in dilute solution studies of purified proteins because the difference in the intrinsic line width is small in dilute solution.

Counter to the utility of <sup>15</sup>N enrichment for in-cell studies of disordered proteins, we do not observe <sup>15</sup>N signals from globular proteins in cells. Similar problems have been observed elsewhere,<sup>4,11</sup> and Sakakibara et al. report the instance of a 7 kDa globular protein that is amenable to in-cell NMR in one *E. coli* strain, but not another.<sup>20</sup> We can exclude several causes for our failure to detect the HSQC spectra of the globular proteins studied here. It is not insufficient expression. The data in Table S1 of the Supporting Information show that the <sup>15</sup>N-enriched proteins are expressed at mM levels, which should allow detection. We can also rule out overexpression because the low concentration of *tfmF*-labeled proteins in cells (50–100  $\mu\text{M}$ ) relative to <sup>15</sup>N-enriched proteins (mM) still leads to broad <sup>19</sup>F resonances. Augustus et al.<sup>4</sup> showed that DNA binding explains the absence of an in-cell HSQC spectrum from the MetJ protein, but this is not a reasonable explanation for our results because the proteins are not DNA binders and all have pI values of 6.5 or less. We can rule out insolubility because the proteins are found in the supernatant of the lysates, not in the pellets. Strong membrane binding is also excluded because membranes are found in the pellets. The proteins also appear to be mostly, and perhaps completely, in their native states because HSQC spectra from lysates of cells expressing UBQ, CI2, CAM, apocytochrome *b<sub>5</sub><sup>21</sup>*, and PDZ3 (Figure S2 of the Supporting Information) are like those of the native proteins in dilute solution. Even though the HSQC spectrum of GFP is not observed in cell slurries or lysates, we know that GFP is in its native state because the samples fluoresce.

Increased viscosity in cells is one reason for our inability to observe HSQC spectra of globular proteins in *E. coli*. High viscosity slows molecular tumbling, increasing the breadth of crosspeaks, which decreases their detectability.<sup>38</sup> The measured viscosity of the clear supernatants from the lysates is 2–4 fold times that of water. Because we did not add liquid during lysis, we can use the protein concentration in the cells and the lysate (Table S1 of the Supporting Information) to estimate that the cytoplasm is diluted 1.5–3.0 fold in the lysates. Combining these ranges, and assuming a direct relationship between viscosity and concentration, gives a crude estimated intracellular viscosity of 3–12 times that of water.

The Stokes–Einstein–Debye equation<sup>49</sup> predicts a direct linear dependence between viscosity and the apparent molecular size of globular proteins. If we assume this equation is valid inside cells, the apparent molecular weight of UBQ, the smallest protein studied here, would be 24–96 kDa. The lower value is compatible with the detection by the NMR methods we used, but the upper value is too large to yield an HSQC spectrum of the protein. This increase in apparent molecular weight will be even greater for the larger proteins. In summary, if our assumptions are valid, the increased viscosity in cells can explain our inability to detect globular proteins. At least one assumption, however, is suspect.

The Stokes–Einstein–Debye relationship breaks down in cells and in lysates because the definition of the viscosity assumes that the species increasing the viscosity (the viscosogen) is infinitely smaller than the test molecule. This definition is valid in systems comprising small viscosogens (like glycerol) and

(45) Pedersen, S.; Bloch, P. L.; Neidhardt, F. C. *Cell* **1978**, *14*, 179–190.

(46) Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Hartl, F. U.; Kerner, M.; Frishman, D. *BMC Genomics* **2008**, *9*, 102.

(47) Eliezer, D.; Kutluay, E.; Bussell, R., Jr.; Browne, G. *J. Mol. Biol.* **2001**, *307*, 1061–1073.

(48) Wu, K. P.; Kim, S.; Fela, D. A.; Baum, J. *J. Mol. Biol.* **2008**, *378*, 1104–1115.

(49) Debye, P. J. *Polar Molecules*; Chemical Catalog Company: New York, 1929.

a globular test protein, but it is invalid in cells where the viscosogens and the test protein are approximately the same size. Such macromolecular crowding can cause negative deviation from the Stokes–Einstein–Debye law, at least when synthetic polymers are used as crowding agents.<sup>50</sup> That is, increases in viscosity decrease the tumbling rate by a smaller amount than is predicted by the equation. Given this negative deviation, the apparent molecular sizes will be less than the estimates given above, providing confidence that, if viscosity were the only factor, we should have observed at least the smallest protein, UBQ, in cells.

We suggest that nonspecific interactions also contribute to our inability to detect globular protein HSQC spectra in *E. coli* cells. There is precedence for the idea that weak, nonspecific interactions are a feature of the cellular interior. It was suggested in the 1930s that the cells might be highly organized, and in the 1940s the complete enzymic repertoire of the Krebs cycle was isolated as a whole.<sup>51</sup> Recent NMR studies indicate that 50% of the proteins in bacterial cells are completely immobile<sup>52</sup> and that nonspecific interactions occur in vitro when proteins are used as crowding agents.<sup>53</sup> In summary, we suggest that a combination of increased viscosity and nonspecific protein interactions explains our inability to obtain high-quality solution-state NMR spectra from globular proteins in *E. coli*.

<sup>19</sup>F labeling not only facilitates leakage detection but also overcomes the problem of detecting small globular proteins in cells (e.g., Figures 3, 4, and 6). Even though the 3FY resonances in UBQ, CI2, and CAM are broad, they are detectable in cells. We attribute this detectability to the low background and high sensitivity afforded by <sup>19</sup>F and the limited number of labels in the proteins. The width of 3FY resonances in UBQ, CI2, and CAM were used to provide a rough estimate of ~40–100 ns for the rotational correlation times.<sup>34</sup> These values represent a 10-fold increase in correlation time compared to dilute solution. Such an increase is also consistent with our inability to detect crosspeaks from <sup>15</sup>N-enriched globular proteins in cell slurries.

For larger globular proteins like GFP, even the <sup>19</sup>F resonances from 3FY are too broad to observe (Figure S4 of the Supporting Information). A clue to overcoming this problem came from our work on disordered proteins. NMR spectra of disordered proteins are observable in cells because the disorder facilitates internal protein motions.<sup>38</sup> These motions are almost completely damped by the inherent order of globular proteins, such that global motion of globular proteins determines the line width of their resonances<sup>54</sup> and hence their detectability.

We reasoned that the independent internal motion of the trifluoromethyl group of tfmF would sharpen the <sup>19</sup>F resonances, thereby facilitating the detection of larger globular proteins. This prediction is borne out. Resonances from tfmF-labeled GFP and HDH are observed in cells (Figures 7 and 8). Furthermore, as

expected, the <sup>19</sup>F peak width of tfmF is molecular weight dependent, increasing from 0.2 ppm for CI2 to 0.5 ppm for GFP, to ~0.6 ppm for the 48 kDa homodimer, nitroreductase,<sup>29,35</sup> and ~1 ppm for the 98 kDa homodimer, histidinol dehydrogenase<sup>29,35</sup> (Figure 8). This idea of increased sensitivity via increased internal motion is used in the labeling of methyl groups with <sup>13</sup>C,<sup>18</sup> but tfmF offers the advantage that the labeled compound is less susceptible to metabolic scrambling.

The chemical shift of <sup>19</sup>F is sensitive to its environment. This sensitivity is readily seen upon cell lysis. Lysis causes a ~0.1 ppm upfield shift of resonances from the intracellular protein for every protein investigated. An upfield shift was also observed for small <sup>19</sup>F containing molecules by Xu et al.<sup>55</sup> By comparing the <sup>31</sup>P to <sup>19</sup>F shifts for compounds containing both nuclei, these authors showed that the difference between extracellular and intracellular compounds arises because of differences in protein hydration inside and outside of cells. These differences in hydration may help explain the changes in protein stability in cells compared to the dilute solution.<sup>56,57</sup>

## Conclusions

The high viscosity and weak interactions in the cytoplasm can make routine <sup>15</sup>N enrichment a poor choice for in-cell NMR studies of globular proteins in *E. coli*. We demonstrated that <sup>19</sup>F labeling is a suitable labeling method for studying not only globular proteins but also disordered proteins in cells with NMR. The <sup>19</sup>F chemical shift and line width provides site-specific structural and dynamic information in cells. In addition, we have shown that the increased motion of tfmF expands the application of in-cell NMR to larger globular proteins. Finally, the decreased rotational motion of globular proteins suggests that high-resolution magic angle spinning<sup>58</sup> might be well suited for in-cell NMR.

**Acknowledgment.** We thank Andrew Lee, Alexander Shekhtman, Anthony Persechini, and Peter Lansbury for supplying the expression systems, Gregory B. Young and Marc ter Horst for NMR training and spectrometer maintenance, and Elizabeth Pielak for comments on the manuscript. This work was supported by the National Science Foundation (MCB-051647) and the National Institutes of Health (5DP1OD783).

**Supporting Information Available:** Complete ref 22, additional descriptions of sample preparation, NMR parameters, and additional in-cell spectra. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA907966N

(50) Li, C.; Wang, Y.; Pielak, G. J. *J. Phys. Chem. B* **2009**, *113*, 13390–13392.

(51) Srere, P. A. *Trends Biochem. Sci.* **2000**, *25*, 150–153.

(52) Persson, E.; Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6266–6271.

(53) Li, C.; Pielak, G. J. *J. Am. Chem. Soc.* **2009**, *131*, 1368–1369.

(54) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624–1671.

(55) Xu, A. S. L.; Waldeck, R.; Kuchel, P. W. *NMR Biomed.* **1993**, *6*, 136–143.

(56) Ghaemmaghami, S.; Oas, T. G. *Nat. Struct. Mol. Biol.* **2001**, *8*, 879–882.

(57) Ignatova, Z.; Gierasch, L. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 523–528.

(58) Lindon, J. C.; Beckonert, O. P.; Holmes, E.; Nicholson, J. K. *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55*, 79–100.