Isotope Reverse-Labeled Infrared Spectroscopy as a Probe of In-Cell Protein Structure

Jacob H. Wat, Nicolas J. Pizzala, and Mike Reppert*

Department of Chemistry, Purdue University, West Lafayette, IN

E-mail: reppertm@purdue.edu

Abstract

While recent years have seen great progress in determining the three-dimensional structure of isolated proteins, monitoring protein structure inside live cells remains extremely difficult. Here we examine the utility of Fourier transform infrared (FTIR) spectroscopy as a probe of protein structure in live bacterial cells. Selective isotope enrichment is used both to distinguish recombinantly expressed NuG2b protein from the cellular background and to examine the conformation of specific residues in the protein. To maximize labeling flexibility and to improve spectral resolution between label and main-band peaks, we carry out isotope-labeling experiments in "reverse-labeling" mode: cells are initially grown in ^{13}C-enriched media, with specific ^{12}C-labeled amino acids added when protein expression is induced.[1] Because FTIR measurements require only around 20 µL of sample and each measurement takes only a few minutes to complete, isotope-labeling costs are minimal, allowing us to label multiple different residues in parallel in simultaneously grown cultures. For the stable NuG2b protein, isotope-difference spectra from live bacterial cultures are nearly identical to spectra from isolated proteins, confirming that the structure of the protein is unperturbed by the cellular environment. By combining such measurements with site-directed mutagenesis, we further demonstrate that the local conformation of individual amino acids

can be monitored, allowing us to determine, for example, whether a specific site in the protein contributes to α-helix or β-sheet structures.

Introduction

A protein's three-dimensional structure is the key to understanding and controlling its biological function. Modern structure-determination tools such as X-ray diffraction (XRD), nuclear magnetic resonance (NMR), and cryogenic electron microscopy (CryoEM) have revolutionized our understanding of how proteins function by revealing the modes by which proteins interact with one another, bind molecular substrates, and catalyze chemical reactions.[2–4]

Despite the dramatic advances enabled by these techniques, the "holy grail" of monitoring the structure of proteins within their native environment - living cells - remains profoundly challenging.[5–12] While many proteins are stable enough to maintain the same fold both in and out of the cellular milieu, exceptions to this trend undoubtedly exist, particularly for disordered proteins whose conformation can depend strongly on their surroundings.[13–15] Interest in such "difficult" systems has expanded dramatically in recent years as their significance in neurodegenerative diseases has become clear.[16–19]

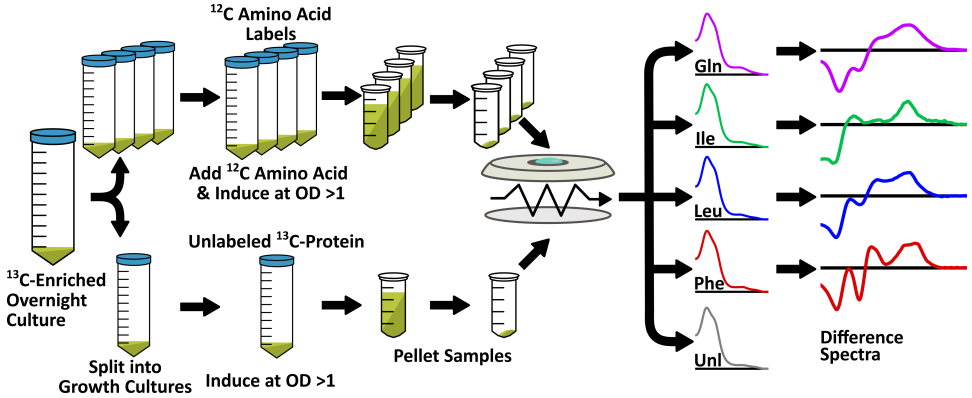


Figure 1: Work flow for live-cell FTIR Spectroscopy. A single ^{13}C-enriched overnight culture is split into labeled (^{12}C Amino Acid Labels) and unlabeled (Unlabeled ^{13}C Protein) subcultures at the time of induction. FTIR difference spectra between the labeled and unlabeled cultures offer structural insight into the recombinantly expressed target protein.

Perhaps the most widely used technique for studying protein behavior in live cells is fluorescence tagging, often in combination with Förster resonance energy transfer (FRET) experiments that can be used to monitor protein-protein association or even the macroscopic conformation (e.g., "folded" vs "unfolded") of large proteins.[20–27] Although challenging to perform, NMR spectroscopy can also provide important insights protein structure in live cells.[6,28,29] Yet both techniques face challenges: FRET is primarily useful for measuring large-scale conformational changes and is largely blind to atomistic structural details. In-cell NMR offers much greater detail but often suffers from severe line broadening due to the restricted motion of proteins in crowded cellular environments.[11,29–32] These limitations impose serious constraints on both the quality and quantity of structural data available for proteins in live cells.

In light of these challenges, we set out to explore the suitability of isotope-labeled Fourier Transform Infrared (FTIR) spectroscopy as a probe of protein structure in live cells. Infrared spectroscopy (IR) is well situated as a complementary method to FRET and NMR. Although FTIR does not provide the atomistc detail of NMR, it has long been recognized as a useful probe of protein secondary structure content (particularly in detecting the formation of β sheets), and when combined with isotope labeling can also provide local structural details such as hydrogen bonding and solvent exposure.[33–45] FTIR instrumentation is relatively inexpensive and widely available; and FTIR measurements require only rather small quantities of sample. In attenuated total-internal reflectance (ATR) mode,[46] for example, a few microliters of 10 mg/mL protein solution – less than 100 µg of protein – is sufficient to take high-quality spectra.

This last observation inspires our approach to in-cell IR spectroscopy. Perhaps the core challenge in any in-cell method is the difficulty of separating the signal of the analyte of interest from that of the cellular background. In principle, this separation could be accomplished (at least in systems with very tight control over expression) by simply comparing the IR spectra of cells grown with and without overexpression of the protein of interest. In

practice, however, we found that differences in growth rates between expression-induced and uninduced cell cultures makes this comparison somewhat unreliable since the "background" culture may not truly match the cellular background of the analyte of interest.

Fortunately, the small sample volume requirements of IR spectroscopy – and thus the cost-feasibility of routine isotope-labeling experiments – offers a more reliable approach to eliminating cellular background signals, illustrated schematically in Fig. 1. By growing two bacterial cultures, *both* of which over-express the analyte of interest but only *one* of which contains isotope-enriched material (added simultaneously with inducing expression so as to selectively label the protein of interest), we hypothesized that it should be possible to construct isotope-label difference spectra in which the cellular background is reliably eliminated. In fact, since isotope-enriched materials are added only at induction, the two cultures can be taken from a shared stock, ensuring that the "sample" and "reference" are truly identical up to the last possible moment.

This strategy rests on a long history of isotope-labeling of recombinant proteins for NMR spectroscopy, including some in-cell experiments.[47–50] Selective labeling in these experiments is typically achieved by adding ^{13}C-enriched amino acids to a minimal-medium growth culture, where virtually all carbon – including that used to produce amino acids in proteins – comes from a glucose feedstock. Uniform labeling is achieved by using ^{13}C-enriched glucose (in which case all proteins are uniformly ^{13}C-enriched) or selective labeling by introducing a single ^{13}C-enriched amino acid into a ^{12}C-glucose background.[47] Although *Escherichia coli* cells are capable of producing all 20 common proteinogenic amino acids, the selective addition of excess alanine, for example, to the growth medium will induce the cells to incorporate the excess alanine directly into expressed proteins, instead of synthesizing new alanine from the ^{12}C glucose feedstock.[47] In this way, selective ^{13}C enrichment of only a single amino acid can be achieved, allowing targeted information from a single type of residue.

This selective labeling strategy has proven particularly useful for in-cell NMR measurements on over-expressed proteins in bacterial cells.[51] Because the labeled amino acids can be

directly incorporated into the over-expressed target protein (without additional metabolic steps that risk conversion into other biomolecules), residue-selective labeling tends to minimize unwanted "background" signal from other cellular components.[49] While the exact level of background labeling is difficult to quantify, residue-selective labeling (in combination with uniform ^{15}N labeling) has been shown to offer sufficiently clean backgrounds even to allow the determination of the structure of small, over-expressed proteins directly in live prokaryotic cells.[5,52]

The purpose of this work is to test the feasibility of this strategy for in-cell IR measurements in a reverse-labeling mode,[1] i.e., using ^{12}C labels in a ^{13}C background, using the 57-residue NuG2b[53–56] protein as a test case. Although the cost of isotope-enriched materials is our primary motivation for reverse-labeling (as opposed to "forward" labeling, with ^{13}C labels in a ^{12}C background), we find that the reverse-labeling approach also offers other advantages, notably better separation between the ^{12}C label peaks and the red-shifted profile of the ^{13}C-enriched protein main band. We obtain nearly identical isotope-difference spectra for NuG2b in live cells and from purified protein in solution, confirming both the feasibility of the approach and that the structure of NuG2b is unaffected by cellular crowding. Finally, by combining in-cell isotope reverse-labeling with single-point mutagenesis in NuG2b, we demonstrate that in-cell measurements can be used to monitor the conformation even of individual protein residues in live cells.

Methods

Overview

Before describing in detail our experimental methods, we first outline the basic assumptions on which our approach rests and the tests we designed to validate those assumptions. Our

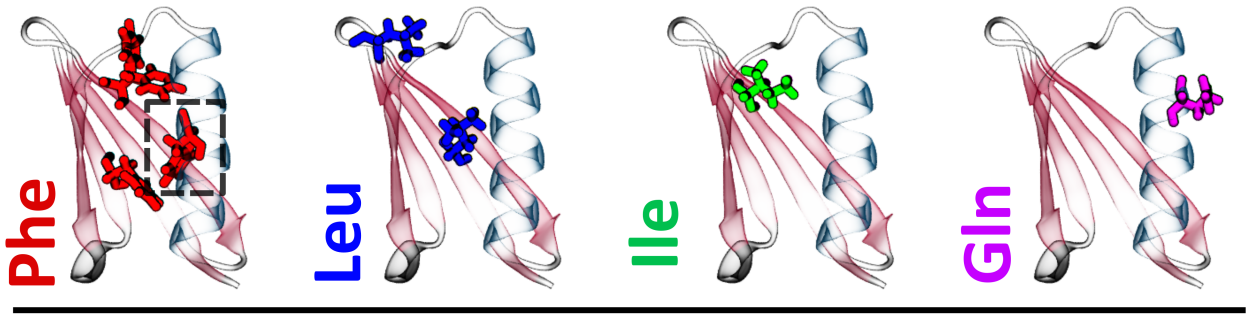
approach to isolating NuG2b signals in live cells rests on two primary assumptions:

1. That amino acids added to the growth medium will be directly incorporated into the NuG2b protein, without substantial cross-labeling into other amino acids.

2. That post-induction cells will produce a strong excess of NuG2b (relative to other proteins or biomolecules), so that isotope-difference spectra represent primarily the properties of NuG2b and not other proteins in the cellular background.

The validity of the first assumption is well-documented in the NMR literature, albeit with the caveat that metabolic conversion rates vary widely across different amino acids.[47,57,58] For example, Gln is readily converted to Glu, while amino acids that are not major reactants in the synthesis of other amino acids, like Leu and Ile, are much more stable.[58–60] The second assumption is less well documented, and its validity will depend strongly on how well the particular target protein expresses.[52,58]

To validate these assumptions in our present context, we first purified NuG2b protein selectively enriched with four different amino acids – Leu, Ile, Phe, and Gln – and used mass spectrometry to characterize the labeling efficiency and to check for interconversion between different amino acids. This allows a direct test of assumption #1. For reference, Figure 2 illustrates the occurrence of each of the four selected labels in the NuG2b structure and sequence; these particular residues were selected due to their appearance in only a few sites with well-defined structures in the NuG2b protein. Based on literature results, the Phe, Leu, and Ile labels are expected to be quite stable, with minimal interconversion to other amino acids, while Gln has a much higher risk of interconversion (and thus unwanted cross-labeling with) Glu.[57–59,61] Quantitative validation of assumption #2 is much more difficult since it requires characterization of not just the over-expressed protein but also of the rest of the cellular background. While such quantification could in principle be accomplished via mass spectrometry, it would be quite complex and is beyond the scope of our present work.



Phe: MDTYKLVIVLNGTT**F**TYTTEAVDAATAEKV**F**KQYANDAGVDGEWTYDAATKT**F**TVTE
Leu: MDTYK**L**VIV**L**NGTTFTYTTEAVDAATAEKVFKQYANDAGVDGEWTYDAATKTFTVTE
Ile: MDTYKLV**I**VLNGTTFTYTTEAVDAATAEKVFKQYANDAGVDGEWTYDAATKTFTVTE
Gln: MDTYKLVIVLNGTTFTYTTEAVDAATAEKVFK**Q**YANDAGVDGEWTYDAATKTFTVTE

Figure 2: Top panel: Structure of the NuG2b protein with reverse-isotope-labeled residues highlighted. The dashed square marks the phenylalanine 31 (F31) residue. Structures were rendered using the software VMD[62] from PDB entry 1MI0,[63] after incorporating the three single point mutations N38A, A47D and D48A.[56] Bottom panel: NuG2b WT sequence with labeled residues highlighted with colored text.

Instead, we chose here to validate assumption #2 using two qualitative tests. First, we compare isotope-difference spectra collected from live *E. coli* cells with isotope-difference curves collected for purified protein, reasoning that close agreement between the spectra offers strong support for the idea that our in-cell signals come primarily from the target protein.

Second, as a more rigorous test, we performed a site-directed mutagenesis experiment where we replaced one of the three Phe residues in NuG2b with Tyr. If the in-cell signal is dominated by over-expressed NuG2b, this change should induce strong and specific changes to the measured spectrum; if instead the cellular background dominates the signal, very little change should be observed.

The remainder of this section provides technical information on our experiments. The outcome of the validation tests are described in the Results section.

NuG2b Expression and Purification

As a reference for in-cell measurements, we collected FTIR spectra for purified NuG2b protein recombinantly expressed in BL21-CodonPlus(DE3)-RIL *E. coli*. This section provides details on NuG2b recombinant expression, purification, and characterization.

Expression

Protein expression followed the protocol provided in Ref[47] scaled down for 100 mL growth cultures. ^{13}C glucose was purchased from Cambridge Isotopes Laboratories while all other chemicals were purchased from Fisher Scientific. The protein NuG2b and the single point mutant F31Y were expressed in BL21-CodonPlus(DE3)-RIL *Escherichia coli* using the pET28a plasmid with an insert coding for NuG2b synthesized by Genscript. The sequence of the wild-type expressed protein is MDTYKLVIVLNGTTFTYTTEAVDAATAEKVFKQYANDAGVDGEW-TYDAATKTFTVTE. The NuG2b protein is a mutant of NuG2 produced by introducing three single point mutations N38A, A47D and D48A.[55,56] To produce protein, 5 mL of ^{13}C minimal media was inoculated with a stab from a glycerol stock and was grown overnight in a 50 mL Falcon tube at 37 °C rotating at 180 rpm. The next day the 5 mL of overnight culture was transferred into 95 mL of ^{13}C-enriched minimal media in a baffled 500 mL Erlenmeyer flask. Once the samples reached an optical density (OD) of 0.6 at 600 nm, 10 mL of 2 $\frac{\text{mg}}{\text{mL}}$ amino acids dissolved in ultrapure water were filtered using a 0.2 µm sterile filter and then added to the culture. 30 minutes after adding the amino acids, 100 µL of 1M isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added. After 3 hours the cells were pelleted at 7,000 rpm for 10 minutes. After discarding the supernatant, the cell pellet was resuspended in 20 mL of ultrapure water and transferred to a 50 mL Falcon tube for storage; after again pelleting the sample at 7,000 rpm for 10 minutes and discarding the supernatant, the cell pellet was stored at -20°C until extraction and purification.

Purification

Frozen cell pellets were resuspended in 10 mL of 5% acetic acid and then transferred to a 15 mL falcon tube. The samples were then lysed through sonication using a Fisherbrand FB-505 Sonicator with a 1/8 in. Microtip Probe (418-A). The cells were lysed for a total time of 6 minutes with a 30% pulse amplitude set for 2 seconds on/ 1 second off (total active sonication time 4 minutes). The samples were then transferred to a 50 mL falcon tube and diluted with water to a total volume of 20 mL. The samples were then pelleted for 30 minutes at 12,000 rpm. The liquid fraction was dialyzed against pure water for 6 hrs with water changes every 2 hrs using Thermo Scientific^{TM} SnakeSkin^{TM} Dialysis tubing with a 3.5 kDa molecular cut off weight. The samples were then lyophilized before further purification using liquid chromatography on a GE Akta Pure housed in a 4 °C fridge. All organic solvents were chilled and degassed using sonication before usage. All aqueous solvents were filtered via vacuum filtration using 0.22 µm Fisherbrand^{TM} General Filtration membrane filters before degassing and chilling.

Crude protein was purified from the cell lysate using size exclusion chromatography (Cytiva Superdex 75 Increase 10/300 GL column). The lyophilized samples were dissolved in 0.5 mL of ultrapure water and injected on to the FPLC using an Air-Tite Products Henke-Ject Low Dead Space Syringe to minimize sample loss. Samples were run at a flow rate of 0.5 $\frac{\text{mL}}{\text{min}}$ with 100 mM Acetic Acid Buffer(pH 5) as the mobile phase. The mobile phase was prepared using HPLC grade glacial acetic acid, ultrapure water, and 10 M sodium hydroxide. Sodium hydroxide pellets and HPLC grade glacial acetic acid were purchased from Fisher Scientific. Elutions were monitored on FPLC using the absorbance at 280 nm. Before purification, the column was first flushed with 2 column volumes of the mobile phase at a flow rate 0.8 $\frac{\text{mL}}{\text{min}}$. Proteins eluted in 1 mL of volume and had an elution time of 15 minutes.

Protein pH Adjustment

After size exclusion purification, the protein was fully protonated due to the low pH of the buffer. The sample was then neutralized by adding sodium hydroxide until the elutions reached a pH between 7-7.5. The pH was tested using a Thermo Scientific Orion 9110DJWP Double Junction pH electrode that had a three point calibration. Samples were then stored for less then 24 hours at 4 °C until desalting could occur.

To desalt and eliminate the buffer, the protein was further purified using reverse phase chromatography (Hypersil Prep HS C18 10µM 250mm x 10 mm column). This was done in two 0.5 mL injections. The mobile phase was a combination of ultrapure water and HPLC grade acetonitrile and had a flowrate of 6 $\frac{\text{mL}}{\text{min}}$. The HPLC grade acetonitrile was purchased from Fisher Scientific. Following protein samples being injected, salt species were eluted first using 60 mL of water followed by protein elution through a gradient 0% to 100% water:acetonitrile. The protein eluted at ≈45% acetonitrile in two 5 mL aliquots totaling 10 mL. The elution times were monitored using the absorbance at 280 nm. The samples were then lyophilized in 1 mg aliquots.

Mass Spectrometry

To characterize the protein by mass spectrometry, all samples were washed using a 3kDa molecular weight cut-off (MWCO) filter 3 times and diluted to a final concentration of 5 µM. Acetic acid, 5% by total volume, was added to the sample solution to promote ionization. Nano-Electrospray Ionization, nESI, was used to generate ions and RF/DC isolation was used to isolate the 4+ charge state on an instrument previously described.[64] Fragment ions generated via Ion Trap Collision Induced Dissociation, IT-CID, were matched based on expected/calculated masses from a protein sequencing program and labeled for the ^{12}C protein control. Since the mass is shifted for protein samples made entirely from ^{13}C, or ^{13}C with ^{12}C amino acid tags, fragment assignments were done manually by comparing fragment patterns to the ^{12}C sample. Fragment ion mass shifts between the ^{13}C sample and the ^{13}C with ^{12}C

amino acid tag samples were attributed to the ^{12}C labels. Since the protein sequence was known a priori, these mass shifts attributed to the ^{12}C labels were calculated based on the number of a particular amino acid in the sequence, and the number of carbons involved with each amino acid. Experimental and theoretical mass values of the whole protein as well as individual fragments were congruent. Fragments involving the ^{12}C label were compared to fragments without it as further evidence the mass shift is caused by a specific ^{12}C label, i.e., b-type ions had a mass shift whereas y-type ions did not and vice versa. Comparison of the calculated and measured mass values indicated approximately 96.8% ^{13}C enrichment. Simulated isotopic distributions of varied ^{13}C percentage enrichment were compared to experimental data to further corroborate this value. All data was replicated with different samples on different days. Of the four isotope labels, leucine, phenylalanine, and isoleucine were found to be within 1 Da of the expected weight for each label with weight decreasing by 12 Da, 26 Da and 5 Da respectively. In contrast, the glutamine-labeled sample was found to have decreased by 29 Da, roughly 6 times larger than the 5 Da shift expected for labeling the single Gln residue in NuG2b; as discussed below, this suggests two types of conversion of isotope-enriched glutamine. Rapid interconversion of Gln with glutamate (Glu) is known to occur in *E. coli* and would account for an additional 20 Da mass shift, since NuG2b contains 4 Glu residues.[58,59] The remaining mass shift of 4 Da likely comes from non-specific labeling of other amino acids due to metabolism of labeled Glu via the citric acid cycle, which it can enter via interconversion with α-ketoglutarate.[57,58]

H/D Exchange

Deuterated proteins were made by dissolving 1 mg of protein in 1.5 mL of D_2O (Cambridge Isotopes) and heating for 4 hours at 75 °C. The samples were then lyophilized. Samples were then dissolved in 20 µL of D_2O for FTIR spectroscopy. To confirm that all hydrogen was exchanged, samples were acidified by adding 20 µL of 0.2 M DCl while the sample was still on the FTIR plate, following neutral pH FTIR measurements. For all H/D exchanged samples,

acidification completely eliminated absorption near 1530 cm^{-1} peak, which in ^{13}C-enriched samples is contributed to by both Amide II and carboxylic acid COO^- side groups from Glu and Asp residues. The absence of residual Amide II absorption confirms the completeness of H/D exchange.

FTIR Spectroscopy

For IR spectroscopy, 1 mg lyophilized protein samples were dissolved in 10 µL of ultrapure water for a final concentration of approximately 100 $\frac{mg}{mL}$. Spectra were collected using a Bruker Invenio S FTIR with a Pike MIRacle ATR (attenuated total reflection) insert. The ATR plate is zinc selenide with a diamond coating (Model Number: 250-2118). All spectra were collected using a liquid nitrogen cooled MCT (Mercury-Cadmium-Telluride) detector, with optics purged using dry N_2 gas (400 $\frac{Nl}{h}$ flow rate). All spectra were collected at 64 scans with a 2 cm^{-1} resolution against a water background.

In-Cell Spectroscopy

To produce samples for in-cell spectroscopy, overnight cultures were grown following the same protocol used for purified proteins. The next day, 0.5 mL of overnight culture was transferred to 5 mL of minimal media and grown to an OD between 0.6-1. Then 0.5 mL of sterile 2 $\frac{mg}{mL}$ amino acid stock solution was added to the culture together with IPTG. The culture was allowed to express for 3.5-4.5 hrs before the colonies were removed. 1.8 mL of culture was collected in a 2 mL centrifuge tube and the samples were centrifuged for 2 minutes at 7,000 rpm. After centrifuging the samples, 500 µL of the supernatant was collected while the remaining supernatant was disposed of. The volume of the remaining cell pellet varied from sample to sample, but was typically between 5 and 10 µL. The cell pellet was then resuspended in an additional 10 µL of supernatant, forming a suspended sample of $\approx$ 15 - 20 µL, which was then placed on the ATR plate for data collection. A 15 mL falcon tube cap was placed over the sample to prevent evaporation. FTIR spectra

were then collected against an atmosphere background. Spectra were collected at 64 scans while the background was collected at 128 scans. (The lower number of scans for the sample spectra is to avoid the possibility of the sample drying out during data collection.) Spectra of the supernatants were collected using the same settings after all cell sample spectra were collected. While concentration varied from sample to sample, the peak optical density (OD) of the protein Amide I band was around 0.025 after subtracting the supernatant background spectrum. The peak OD of the supernatant background in the Amide I region was typically around 0.4, due primarily to the water bend vibration.[65]

Spectrum Processing

Despite N_2 purging, sharp lines due to absorption from water vapor were present in some experimental spectra. Although the absolute amplitude of such signals was quite weak (less than 10^{-3} OD), they can be noticeable in some isotope-difference spectra. To correct for this, a pure water vapor spectrum $V(\omega)$ (obtained by breaking the N_2 purge on the ATR chamber) was subtracted from each raw experimental spectrum $S_{\text{raw}}(\omega)$ to produce the vapor-corrected spectrum

$$\mathbf{S}_{\text{corr}} = \mathbf{S}_{\text{raw}} - \alpha\mathbf{V} \tag{1}$$

with a prefactor α calculated as described below. In Eq. (1) and in what follows, boldface characters (e.g., $\mathbf{S}_{\text{raw}}$) represent discretely sampled spectra, with S_n the n^{th} sampled data point.

To calculate the optimal value for the coefficient α, a copy of the experimental spectrum was first high-frequency-filtered to accentuate the water-vapor contribution, which is much sharper than the protein/solvent spectra. The high-frequency spectrum was calculated as

$$\mathbf{S}_{\text{HF}} = \mathcal{F}^{-1}\left\{\mathbf{W} \times \mathcal{F}\left\{\mathbf{S}_{\text{raw}}\right\}\right\}. \tag{2}$$

Here $\times$ denotes element-wise multiplication, $\mathcal{F}\{...\}$ denotes a discrete Fourier transform, $\mathcal{F}^{-1}\{...\}$ denotes the discrete inverse transform, and $\mathbf{W}$ is a high-frequency filter constructed as

$$\mathbf{W} = 1 - |\mathcal{F}\{\mathbf{G}\}|, \tag{3}$$

where $\mathbf{G}$ is a normalized Gaussian profile with a standard deviation of 2.5 cm^{-1}. Roughly speaking, this high-frequency filter eliminates signals *broader* than 2.5 cm^{-1}. The coefficient α was then calculated as

$$\alpha = \frac{\mathbf{S}_{\mathrm{HF}} \cdot \mathbf{V}}{\|\mathbf{V}\|^2}, \tag{4}$$

where "$\cdot$" denotes the inner product, and $\|\mathbf{V}\|$ is the l^2 norm

$$\|\boldsymbol{V}\| = \sqrt{\sum_n V_n^2}. \tag{5}$$

Note that Fourier filtering is used only to accurately estimate the optimal subtraction coeffiicent α. The final corrected spectrum is unfiltered. To offer a sense for the magnitude of these corrections, Figure 3 compares typical raw (black line) and vapor-corrected (red dashed line) spectra. Water vapor signals are visible only in the inset which zooms in on the blue-boxed region near the baseline. In the inset, the four small bumps are water-vapor lines which are eliminated in the red dashed curve. Relative to the main Amide I band, such signals are negligible, but vapor-correction does improve the line shapes of some isotope-difference curves.

To correct for variable sample concentration and minor baseline offsets (see, e.g., raw data in Figure 3), vapor-corrected spectra were "zeroed" at 1725 cm^{-1} (i.e., the value at 1725 cm^{-1} was subtracted from each curve) and then normalized to have unit area between 1568 - 1725 cm^{-1} (the approximate minimum between Amide I and Amide II bands in

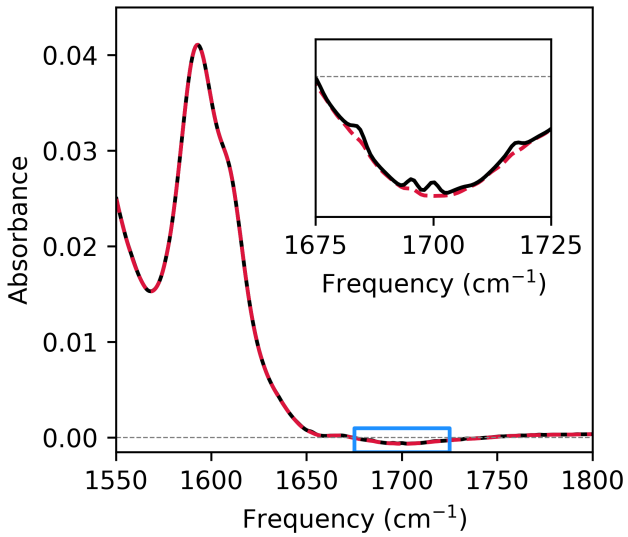


Figure 3: Illustration of our water vapor correction procedure. Vapor-corrected spectra (red dashed line) are produced from raw spectra (black) by subtracting an atmospheric reference spectrum (not shown) as per Eq. (1). The inset gives a zoomed-in view of the blue box near 1700 cm^{-1}.

^{13}C-enriched proteins) before subtraction. Purified-protein spectra were averaged over two independent experiments before subtraction. Live-cell spectra were averaged over three independent measurements. Spectra of purified proteins were averaged *before* subtraction, while live-cell spectra were averaged *after* subtraction to avoid mixing different levels of cellular background signal.

Results and discussion

Uniform Isotope Labeling

Amide I isotope-labeling experiments are most frequently performed in a "forward-labeling" mode, where a small number of ^{13}C- or ^{13}C^{18}O-enriched residues are incorporated into a predominantly ^{12}C protein.[33,39–42,42–45,47,54] While this approach could potentially be adapted for in-cell experiments, we opted instead for a "reverse-labeling" scheme where ^{12}C amino acids are incorporated into a ^{13}C-enriched cellular environment prepared by growing bacterial

cultures in ^{13}C-enriched glucose.[1,47] The primary motivation for this choice is flexibility: while ^{13}C-enriched preparations of some select amino acids (e.g., Ala, Gly, and Val) are affordable, ^{13}C-enriched stocks of other residues can be prohibitively expensive or difficult to obtain commercially. In addition, ^{13}C labels often provide limited separation from the Amide I main band, since β-sheet structures feature a strong absorption band near 1630 cm^{-1}, in the same range where isolated ^{13}C-labeled amide bonds absorb. Much better spectral separation is provided by ^{13}C^{18}O-enriched labeling, but ^{13}C^{18}O-enriched amino acids are not generally commercially available and so must be synthesized in-house using ^{18}O-enriched water.[54] In contrast, reverse-labeling allows us to work with a single, readily available isotope-enriched feedstock (^{13}C glucose) for all samples. By avoiding overlap with the main β-sheet band (see below), this approach also offers a cleaner label window compared to ^{13}C forward labeling.

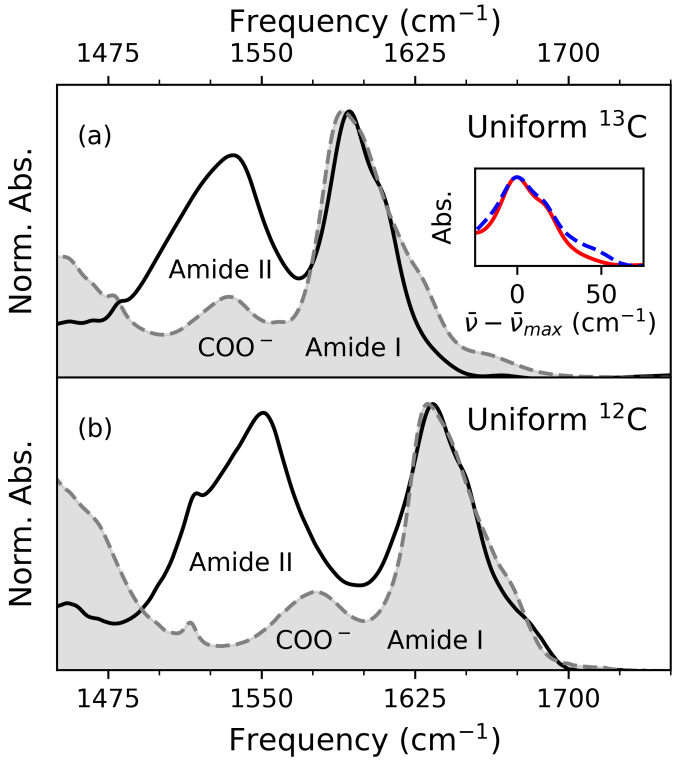


Figure 4: (a) Uniform labeled FTIR spectra of purified ^{13}C-enriched NuG2b in H_2O (black) and D_2O (shaded gray). (b) Spectra for purified ^{12}C NuG2b in H_2O (black) and D_2O (shaded gray). A direct comparison of ^{12}C (blue) to ^{13}C (red) spectra in H_2O is shown in the inset; for comparison both spectra have been shifted so that the Amide I maxima are at 0 frequency.

To orient the reader, Figure 4 presents IR absorption spectra of purified NuG2b protein under four different isotopic conditions: Frame (a) shows spectra for uniformly ^{13}C-enriched NuG2b in both H_2O (black line) and after deuterium exchange in D_2O (shaded gray curve). Frame (b) shows the same spectra for NuG2b without ^{13}C enrichment. Before discussing residue-specific isotope labeling data, we first describe the key spectroscopic features of these isotopically uniform samples. For an overview of the interpretation of such spectra, the reader may refer to Ref.[36]

The large band near 1540 cm^{-1} that is present in H_2O but not D_2O is the Amide II vibration. Upon deuteration, this dominantly N-H wag vibration redshifts to ∼1460 cm^{-1},[36] and is no longer fully visible in the Figure. The smaller peaks near 1535 cm^{-1} in the $H_2O/^{13}$C sample and near 1575 cm^{-1} in the $H_2O/^{12}$C sample are due to the protein C terminus and carboxylic acid sidechains Asp and Glu;[36] at low pH, when these groups are protonated, these bands shift to the high-frequency side of the Amide I band. These bands are presumably also present in the deuterated samples but are then obscured by the strong Amide II absorption.

The structured band that shifts from ∼1630 cm^{-1} in Frame (b) to near 1590 cm^{-1} upon ^{13}C enrichment is the Amide I band, due primarily to the backbone C=O stretch vibration.[36] In NuG2b, the main Amide I peak shifts only slightly (by around 3 cm^{-1}) upon deuteration. The main peak is due to the β-sheet that dominates the NuG2b structure, while the weaker shoulder near 1650 cm^{-1} in the ^{12}C samples and near 1610 cm^{-1} in the ^{13}C samples corresponds to absorption from the shorter α helix structure.[36]

It is notable that both ^{13}C enrichment and deuteration produce changes to the structure of the Amide I band that cannot be described as uniform shifts. For both ^{12}C and ^{13}C proteins, α-helix (1650 cm^{-1} and 1610 cm^{-1}, respectively) and β-sheet (1630 cm^{-1} and 1590 cm^{-1}, respectively) features are better resolved in H_2O than in D_2O. And the overall width of the ^{13}C-enriched sample is noticeably narrowed in H_2O compared to D_2O. In fact, the β-sheet $\nu_{\|}$ mode (whose transition dipole is roughly parallel to the β-strand direction) that appears near 1680 cm^{-1} in the ^{12}C sample is essentially undetectable under ^{13}C enrichment

in H_2O. (Based on the roughly 40 cm^{-1} shift observed for the main α-helix and β-sheet peaks, this band would be expected near 1640 cm^{-1}.) The reasons for this apparent suppression are not obvious; one possibility is a change in the orientation of the Amide I transition dipole upon ^{13}C-enrichment in H_2O due to the altered normal mode structure. This would lead to different selection rules for both α-helix and β-sheet normal modes, and it could induce a more complete suppression of the high-frequency $\nu_{\parallel}$ mode.

The bend vibration of H_2O solvent also peaks near 1640 cm^{-1}, which might suggest a role for solvent/solute coupling or improper background subtraction in the loss of the $\nu_{\parallel}$ feature. For example, the baseline in Figure 3 goes slightly below zero near 1700 cm^{-1}, likely due to small changes in the width of the H_2O band due to solute-solvent interactions. In this case, however, one would expect that the structure of the ^{12}C main-band peaks near 1635 and 1650 cm^{-1} should likewise be affected, since these modes also overlap strongly with the H_2O bend vibration. This does not appear to be the case, as illustrated in the inset of Frame (a), which compares ^{12}C and ^{13}C line shapes, with frequency axes shifted so that the β-sheet peak appears at zero frequency. While the curves clearly differ near 50 cm^{-1} (the $\nu_{\parallel}$ mode region), their lineshapes are nearly identical near the main-band β-sheet and α-helix peaks. It thus appears unlikely that overlap with the H_2O bend vibration is alone sufficient to account for the non-appearance of the $\nu_{\parallel}$ band in our uniform ^{13}C spectra. It will be interesting in future work to explore this change in line shape more fully; such effects have little impact on our present study, apart from the helpful fact that the loss of the $\nu_{\parallel}$ offers a somewhat more uniform background in the isotope-label region.

Reverse Isotope Labeling in Isolated NuG2b in D_2O

For reverse-isotope-labeling measurements, the most important feature of the uniform-enrichment spectra from Figure 4 is that, in all cases, the Amide I main-band peaks are skewed toward low frequencies, a characteristic consequence of the β-sheet (and to a lesser extent α-helix) geometries.[33,37,39,42,66,67] (See Figure 1 of Ref.[39] for example spectra from both secondary

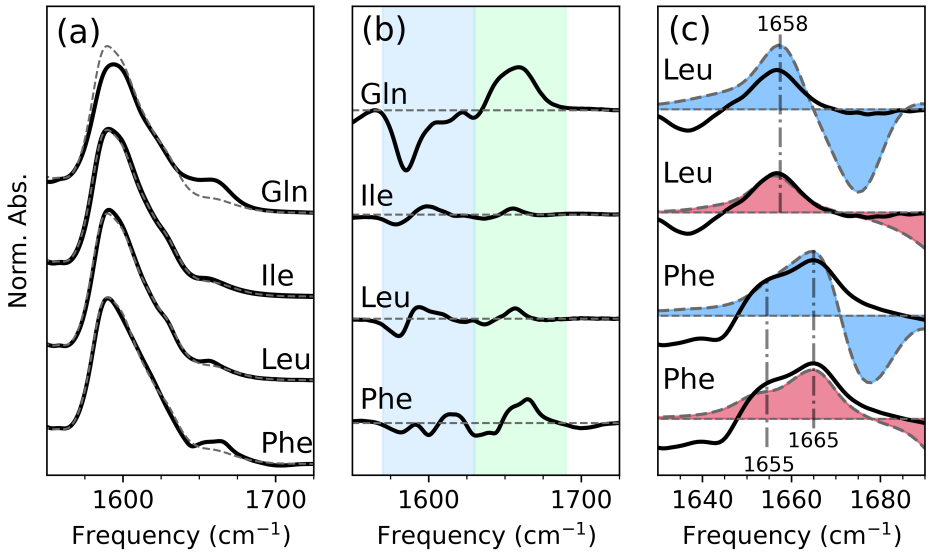


Figure 5: (a) Deuterated reverse-labeled purified protein spectra (black); gray dashed curves show the spectrum of the unlabeled ^{13}C-enriched protein for comparison; all spectra were collected in D_2O. (b) Difference spectra (labeled-minus-unlabeled) for each curve from frame (a); the green and blue shading represent the reverse-label (1630 to 1690 cm^{-1}) and main-band regions (1570 to 1630 cm^{-1}), respectively. (c) Comparison of Leu and Phe reverse-label difference spectra with the corresponding ^{13}C (blue) and $^{13}C^{18}O$ (red) forward-label difference spectra; forward-label spectra are from Ref.[54] and are shifted by 51 cm^{-1} (^{13}C labels) or 71 cm^{-1} ($^{13}C^{18}O$ labels). Vertical dashed lines mark peak maxima in the reverse-labeled spectra.

structure types.) As illustrated in Figures 5 and 6, this turns out to be quite advantageous for reverse-labeling experiments. Frame (a) of Figure 5 presents Amide I spectra of the purified NuG2b protein reverse-labeled at four different residues (Phe, Leu, Ile, and Gln; solid lines), with the spectrum of the uniformly ^{13}C-labeled protein shown for comparison as a dashed line. All spectra are normalized by the integrated area between 1568 and 1700 cm^{-1}, after subtracting off a flat baseline to "zero" each spectrum to the value at 1700 cm^{-1}. (The offset from zero in the raw data is small but becomes significant in isotope-difference spectra.) Thanks to the asymmetry of the Amide I main band, the reverse-label features near 1660 cm^{-1} appear on top of a relatively flat baseline; residual absorption in the unlabeled sample in this region is presumably due to the $\sim$3.2% non-specific ^{12}C content (as estimated from mass spectrometry) due to imperfect isotope-enrichment.

Frame (b) presents the same data as difference spectra (labeled minus unlabeled); the green and blue boxes represent the isotope-label and main-band region, respectively. In these reverse-labeled spectra, the selectively labeled features occur near 1660 cm^{-1}, highlighted by the gray band from 1630 to 1690 cm^{-1} in Frame (b). The intensity of the label feature depends on how many sites in the protein are selectively labeled. NuG2b has 3 Phe residues, resulting in a relatively intense band with lobes near 1655 and 1665 cm^{-1}. Only 2 Leu labels and a single Ile site are present in the NuG2b sequence, leading to relatively weaker label features for these constructs. This trend is broken by the Gln-labeled construct, which shows a very strong label band, despite the occurence of only a single Gln amino acid in the NuG2b sequence.

The large Gln label signal is explained by the fact that Gln is readily converted by *E. coli* to glutamate (Glu)[58] (of which 4 occur in the NuG2b protein) and, via the conversion of Glu to α-ketoglutarate, to many other cellular metabolites.[57,59,61] This conclusion is supported by the mass spectrometry data (see Methods) which shows a larger-than-expected mass shift from labeling Gln (29 Da instead of 5 Da). In contrast, mass spectrometry data for all other labels showed the mass shift expected from the NuG2b sequence within 1 Da. Thus the

strong label feature is presumably a result of cross-labeling with Glu and (to a lesser extent) other amino acids, while the label Leu, Ile, and Phe spectra appear to reflect selectively the properties of only the target residues.

In addition to the label features near 1660 cm^{-1}, the difference spectra in Frame (b) also reveal changes in the "main band" region between 1570 and 1630 cm^{-1} (blue shading), which are dominated by a β-sheet mode near 1590 cm^{-1} and a weaker α-helix mode near 1610 cm^{-1}. Physically, changes in this region are expected as a result of the disruption and weakening of the delocalized α-helix and β-sheet vibrational modes due to the removal of ^{12}C-labeled residues from from resonance with their ^{13}C-background neighbors.[37]

Significantly, but not surprisingly, the difference-spectrum features observed in the main-band region correlate with the secondary-structure type in which the labelled residues are embedded. For example, the Phe labels (two of which occur in sheet and one in helix structures) induce two features in the main-band region: one at 1610 cm^{-1} (corresponding to α-helix absorption) and the other at 1590 cm^{-1} (corresponding to the β sheet). The Leu and Ile spectra are dominated by β-sheet features near 1590 cm^{-1}, reflecting the fact that these labels occur either within β-sheet structures or the solvent-exposed loop region (one Leu site), which is expected to give a rather broad and featureless contribution to the Amide I band.[54] The Gln label again breaks the trend, giving a strong β-sheet feature, despite the situation of the NuG2b label in an α-helix structure, confirming again that substantial cross-labeling of other amino acids must occur in this construct.

To facilitate comparison between forward- and reverse-labelling data, Frame (c) of Figure 5 compares Leu and Phe reverse-labeled difference spectra (from Frame (b)) with the corresponding forward-labeled spectra from Ref.,[54] after shifting the forward-label curves to appear in the same frequency range as the reverse-label spectra. (Forward-labeling data for Gln and Ile are not available.) Forward-label difference spectra were calculated following Ref.[54] by subtracting the labeled from the unlabeled spectra after normalizing to have unit area from 1525-1700 cm^{-1}. Blue-shaded curves in Frame (c) represent ^{13}C forward-label

spectra shifted by 51 cm^{-1} to higher frequency; red-shaded curves are $^{13}C^{18}O$-labeled spectra, which are shifted by 71 cm^{-1}. These frequency shifts were chosen to approximately match the peak frequencies of forward- and reverse-labeled spectra. Apart from the overall shift, the $^{13}C^{18}O$ spectra are quite similar to their reverse-label equivalents. Differences on the low-frequency edge of the reverse-label spectra and on the high-frequency edge of the forward-label spectra are due to interference with the Amide I main band, which appears on opposite sides of the label region in the two curves.

The main label region from 1650 - 1680 cm^{-1}, however, appears largely the same in both cases. The ^{13}C forward-label spectra, on the other hand, miss key features – notably the double-peak structure of the Phe label – due to interference with the Amide I main band. Such features reflect the local spectroscopic properties of the labeled residues, and are strongly sensitive to environmental factors such as electrostatics and hydrogen bonding.[39] This avoidance of main-band effects in the reverse-labeled samples (without the need for costly ^{18}O-enriched material) is a notable benefit of reverse-labeling, made possible by the asymmetry of the Amide I main band away from the reverse-label region.

H_2O vs D_2O

While working in deuterated solvents like D_2O greatly simplifies Amide I data collection (by avoiding interference from the H_2O bend vibration), it is not always possible with biological tissues or live cell cultures. Washing live cells in D_2O can readily remove ambient water, but hydrogen atoms buried inside hydrophobic protein cores may take many hours to exchange at ambient temperatures. Although even partial exchange largely eliminates interference from the H_2O bend, incomplete H/D replacement creates new headaches, since deuterated Amide I modes (often denoted Amide I') are typically red-shifted by a variable amount compared with the equivalent Amide I vibration.[68] Incomplete H/D exchange in live-cell samples thus makes it difficult to tell which frequency shifts are due simply to H/D exchange as opposed to more interesting structural factors such as hydrogen bonding or charge interactions.

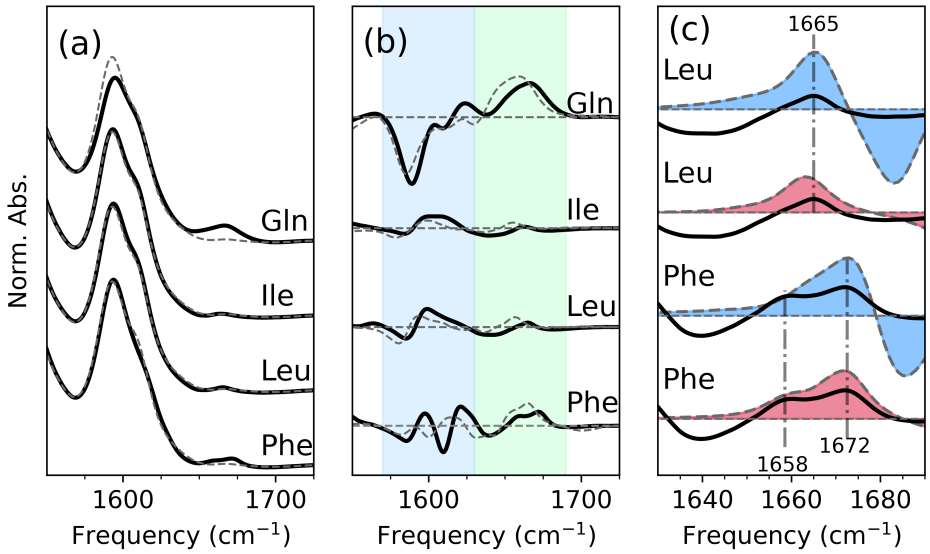


Figure 6: (a) Reverse-label spectra for purified proteins in H_2O; gray dashed curves show the spectrum of the unlabeled ^{13}C-enriched protein for comparison. (b) Difference spectra (labeled-minus-unlabeled) for each curve from frame (a); green and blue shading, respectively, highlight the isotope-label (1630 to 1690 cm^{-1}) and main-band (1570 to 1630 cm^{-1}) regions. Deuterated difference spectra are shown through the dashed curves. (c) Comparison of Leu and Phe reverse-label difference spectra with the corresponding ^{13}C (blue) and $^{13}C^{18}O$ (red) forward-label difference spectra; forward-label spectra are from Ref.[54] and are shifted by 59 cm^{-1} (^{13}C labels) or 78 cm^{-1} ($^{13}C^{18}O$ labels). Vertical dashed lines mark peak maxima in the reverse-labeled spectra.

To avoid this ambiguity, we chose to carry out in-cell experiments in H_2O solvent, without H/D exchange. In preparation for live-cell experiments, we thus repeated the experiments depicted in Figure 5 above for isolated NuG2b in H_2O without any prior deuteration. The resulting absorption spectra are plotted in frame (a) of Figure 6 (solid curves), with the uniform label-free (^{13}C-enriched) spectrum from Figure 5 depicted by dashed lines for comparison. Frame (b) presents the same data in the form of difference spectra (solid lines) in order to highlight the label features near 1650 cm^{-1}; dashed lines show the corresponding difference spectra for the deuterated protein from Figure 5 for comparison. Frame (c) zooms in on the label region from 1630 cm^{-1} to 1690 cm^{-1} (green-shaded area in Frame (b)), and compares the H_2O/reverse-labeled difference spectra from Frame (b) with the corresponding D_2O/forward-labeled spectra from the literature. To facilitate direct comparison, the forward-label literature spectra are shifted to higher frequencies by 59 cm^{-1} for ^{13}C spectra or 78 cm^{-1} for $^{13}C^{18}O$ spectra; these frequency shifts are chosen by eye to maximize overlap between forward- and reverse-label peaks. Blue-shaded curves represent ^{13}C forward-label spectra, shifted by 59 cm^{-1}; red-shaded curves represent $^{13}C^{18}O$ forward-label spectra, shifted by 78 cm^{-1}.

In general terms, the H_2O spectra are quite similar to their D_2O counterparts from Figure 5, featuring only minor (and expected) frequency shifts. One notable difference is that, due to the absence of a discernible β-sheet $\nu_{\|}$ mode, the isotope-label region (roughly 1640 to 1690 cm^{-1}) has a nearly flat baseline in the unlabeled sample, featuring only a slight bump due to incomplete isotope enrichment. It remains to be seen whether this flat baseline will be preserved in other ^{13}C-enriched proteins in H_2O, but at least for NuG2b, it offers a convenient baseline for monitoring reverse-label features.

In-Cell Experiments

With this basic characterization of reverse-labeled NuG2b in place, we turned next to see whether equivalent measurements could be performed in live *E. coli* cells. Our strategy for

distinguishing NuG2b signal from the cellular background capitalizes on the tight control of protein production afforded by recombinant expression. We for some time attempted to isolate NuG2b signal by comparing the IR absorption spectra of "induced" and "uninduced" cultures, i.e., cultures in which NuG2b expression either had or had not been induced by the addition of IPTG to the growth medium. Unfortunately, we had difficulty obtaining consistent difference spectra with this method, likely due to the different growth conditions of induced and uninduced cells. (Culture growth rates slow substantially upon IPTG-induction as resources are directed toward protein expression rather than cell growth and division.)

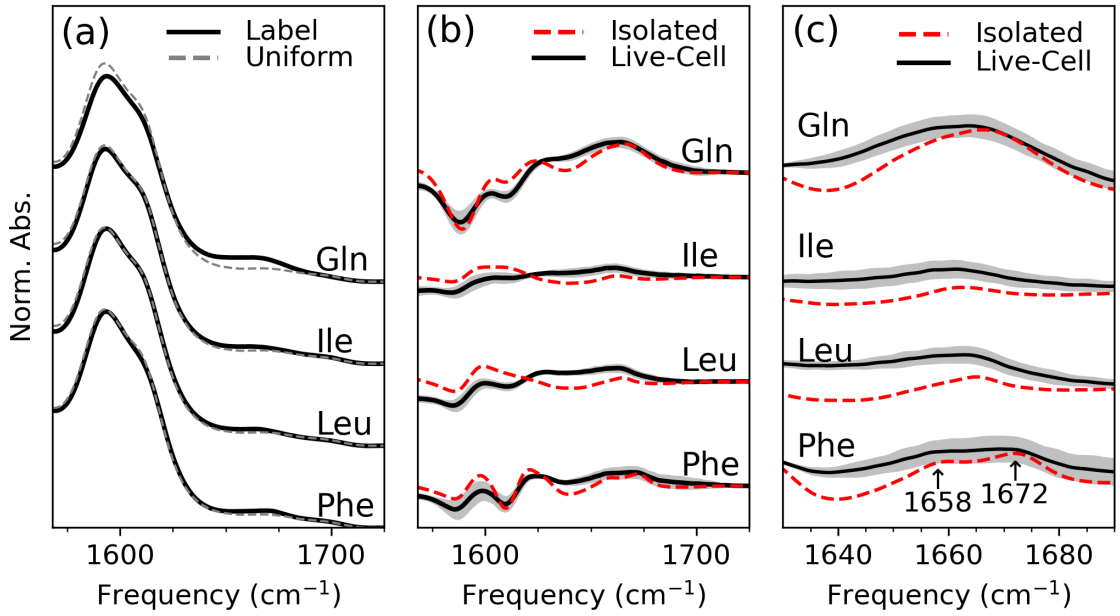


Figure 7: (a) FTIR spectra for *E. coli* cell pellets expressing NuG2b, either with uniform ^{13}C enrichment (dashed gray curves) or with specific reverse labels (black). (b) Isotope reverse-label difference spectra for NuG2b collected in-cell (black) or for isolated protein (dashed red). The gray shaded area around each curve represents $\pm\sigma$, where σ is the standard deviation at each frequency across difference spectra for three different cell cultures. (c) Close up of the 1640 - 1680 cm^{-1} region from Frame (b).

Fortunately, much more consistent results were obtained when we switched to directly comparing spectra from "labeled" versus "unlabeled" bacterial cultures. By splitting cultures into labeled and unlabeled flasks only at the time of IPTG induction (with simultaneous addition of ^{12}C amino acids in the "labeled" cultures), we obtained closely parallel growth rates

between each reverse-labeled sample and the unlabeled (uniformly ^{13}C-enriched) reference.

Figure 7 presents the results of such an experiment on *E. coli* cultures expressing NuG2b. Frame (a) presents absorption spectra of the entire resuspended cell pellet; solid black lines are for reverse-labeled cultures, with the uniformly ^{13}C-enriched culture shown as dashed black lines for reference. The magnitude of the reverse-label feature is noticeably weaker here compared with the purified protein spectra of Figure 6 since the whole-cell suspension contains contributions from many non-target proteins and non-protein biomolecules, including lipid ester groups and nucleic acids. Even in Frame (a), however, distinct label features are evident in comparing the two curves.

These features are brought out more clearly in Frames (b) and (c) which compare labeled-minus-unlabeled absorption difference spectra for each target amino acid. Solid lines here are live-cell difference curves, while dashed lines show the corresponding spectra for purified NuG2b for comparison. Although minor deviations in line shape are visible (particularly for the Ile label, which shows a weak in-cell α-helix feature near 1610 cm^{-1} that is not present in the isolated-protein spectrum), overall agreement between in-cell and purified NuG2b spectra is remarkably good. Given the distinctive line shapes of each reverse-label spectrum, this agreement strongly supports the hypothesis that we the observed difference spectra correspond to intact NuG2b inside the cellular environment.

Labeling with Mutagenesis

Encouraged by this result, we decided to use site-directed mutagenesis to both (a) confirm that the observed signals indeed correspond to intact NuG2b and (b) test a previous assignment that the higher-frequency Phe label feature (1672 cm^{-1} in Figure 7(c)) represents absorption by NuG2b's two β-sheet-embedded Phe residues, while the lower-frequency peak (1658 cm^{-1} in Figure 7(c)) corresponds to a lone Phe site that resides in an α helix.[54] To this end, we repeated our in-cell IR measurements using a plasmid coding for a NuG2b mutant (F31Y) in which Phe 31 (the lone Phe residue in the α helix) is replaced with a Tyr residue.

(Note that this corresponds to F30 in the numbering of Ref.,[54] in which the initial Met residue is indexed as "M0".) Assuming that our in-cell signals do indeed represent intact NuG2b, Phe-labeled spectra for the F31Y mutant should show a complete loss of difference features in the α helix region near 1610 cm^{-1}, since the F31Y construct has no α-helical Phe sites. A corresponding loss of signal in the reverse-label region should further reveal which label features are due to the α-helical Phe site in the WT protein.

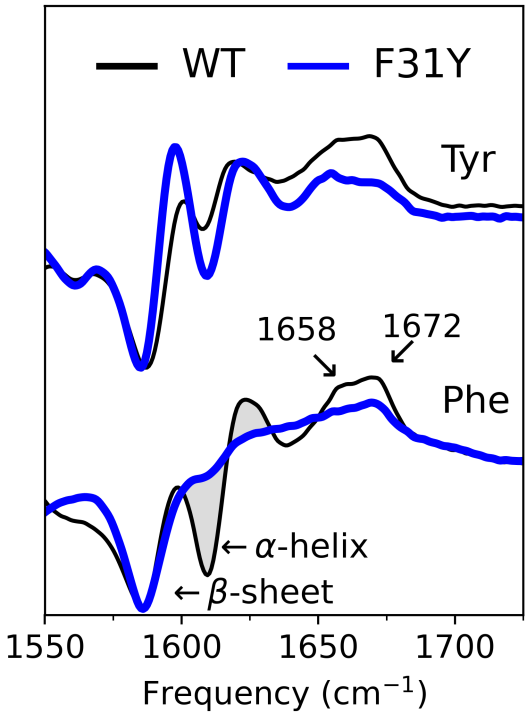


Figure 8: In-cell FTIR difference spectra of wild type NuG2b (black) and the mutant F31Y (blue). The gray shaded region highlights loss of α-helix signal in the F31Y mutant.

The results of this experiment are presented in Figure 8, which shows in-cell isotope-difference spectra for the WT protein as gray-filled curves and the corresponding spectra for the F31Y mutant as blue curves.The Phe-labeled spectra for F31Y show clearly the expected loss of signal near 1610 - 1625 cm^{-1} due to the elimination of the lone α-helical Phe site. Conversely, the Tyr-labeled spectra for F31Y shows enhanced features in the α-helix region due to the gain of an additional Tyr site in the NuG2b α helix. These correlated gain/loss features in the α-helix region provide strong evidence that our measured spectroscopic signals

indeed correspond to intact NuG2b, since it is difficult to rationalize any other mechanism by which the IR spectra of the whole bacterial cells would respond to the change of a single amino acid in the NuG2b plasmid.

The interpretation of the reverse-label region is more nuanced. On the one hand, the F31Y mutant does show a distinct loss of absorption near the lower-frequency (1658 cm^{-1}) Phe-label peak and a corresponding gain in absorption in this region in the Tyr-label spectrum. Both of these observations support the previous assignment of this feature to NuG2b's lone α-helix Phe site. However, a residual low-frequency shoulder persists in the F31Y Phe label spectra, suggesting that the remaining β-sheet labels also absorb in this region. Spectral differences between these two sites would indeed not be surprising, given that (although both C=O groups are interior to the β sheet) the amide N-H group of one of the two sites is solvent-exposed, while the other is embedded in the sheet structure.

Discussion and Conclusions

In-Cell IR Labeling

The goal of this work was to test whether isotope-difference FTIR spectroscopy offers a workable approach to monitoring protein structure in live bacterial cells. The results presented in Figures 7 and 8 answer in the affirmative, though with some caveats. The comparison of WT and F31Y spectra leave little doubt that the isotope-difference signals measured from the live cell culture originate largely from our target protein, since "background" signals would be unchanged by the single point mutation. On the other hand, the discrepancy near 1610 cm^{-1} between isolated and live-cell Leu spectra in frame (b) of Figure 7 suggests that some bleed-through of isotope labeling may occur to "background" proteins. This conclusion is supported by the slight broadening of features in the reverse labeling region between 1640

and 1680 cm^{-1}, which could easily be due to residue-specific enrichment of the cellular background. Overall, however, our results indicate that, at least for high-expression proteins, it is possible to detect structurally relevant spectroscopic features from live-cell measurements. In further developing this approach, it will be critical to quantify more carefully the rates of both isotope-label interconversion between different amino acids (such as the Gln $\rightarrow$ Glu interconversion observed here) and bleed-through incorporation of isotope labels into the cellular "background", since both will affect the specificity of the measured signal.

This in-cell labeling approach will likely be most useful when combined with site-directed mutagenesis, as in Figure 8. Since the cellular background is unaffected by point mutations in the target protein, comparison of wild-type and mutant isotope-difference spectra offer a means of identifying which spectral features arise from individual sites in the protein – assuming of course that mutants can be constructed that do not significantly perturb the protein structure. For residues like Phe and Tyr that have close chemical analogs, this is a relatively safe proposition; for more distinctive residues such as Pro (whose sidechain forms a 5-membered ring with the backbone N atom), constructing "conservative" mutations is more difficult.

Extension to 2D IR Spectroscopy

Given this success with FTIR measurements, it should in principle be straightforward to extend this approach to nonlinear measurements such as two-dimensional infrared (2D IR) spectroscopy, although 2D IR measurements will come with both challenges and advantages relative to FTIR. On the side of "challenges", 2D IR measurements are more susceptible to scatter than linear absorption, so the turbidity of concentrated cell cultures may pose some technical difficulties. But well-developed strategies already exist for handling highly scattering samples in 2D IR using beam chopping and/or phase cycling, and successful whole-cell 2D IR experiments have indeed already been reported.[69–73]

The inability to completely H/D exchange whole cells will likewise require some additional care when collecting 2D data, but such difficulties are by no means prohibitive.[66,74,75] Indeed, for live-cell samples, this difficulty should be mitigated to a significant extent by the fact that cell pellets can be prepared at very high concentrations without worrying about inducing aggregation inside the cellular environment. Thus background signal from H_2O can be considerably lessened simply by working with highly concentrated cell pellets. Dealing with the H_2O background is also made easier by the fact that 2D IR signals scale with the fourth power of the transition dipole moment, whereas linear absorption response scales with the second power. The H_2O bend vibration near 1640 cm^{-1} has a relatively weak transition dipole moment compared with protein Amide I transitions, but it nonetheless dominates linear absorption spectra for dilute proteins in solution due to the much higher molar concentration of H_2O compared to protein. In nonlinear signals, the relative magnitude of the protein signal increases substantially thanks to its stronger per-residue transition dipole moment.[66,75]

Finally, calculating accurate isotope-difference spectra in 2D IR is in some ways more complicated than in FTIR due to the absence of a simple oscillator-strength conservation rule. Isotope-induced shifts are not expected to change the integrated area under an FTIR absorption spectra, so that absorption spectra from different samples can be normalized to have the same area before subtraction. In contrast, the nonlinear nature of 2D IR signals means that there is no simple area-conservation rule for normalizing spectra, which can make difference-spectrum calculations more difficult. Again, however, 2D IR offers compensatory advantages, particularly increased resolution of individual spectral features, which also helps in normalizing spectra for subtraction.

H/D Frequency Shifts

Before closing, it is worth noting that comparison of the reverse-label absorption spectra in Figures 5 and 6 provides a convenient opportunity to check the H/D frequency shift for individual sites within a folded protein. Whereas H/D exchange is often assumed to introduce

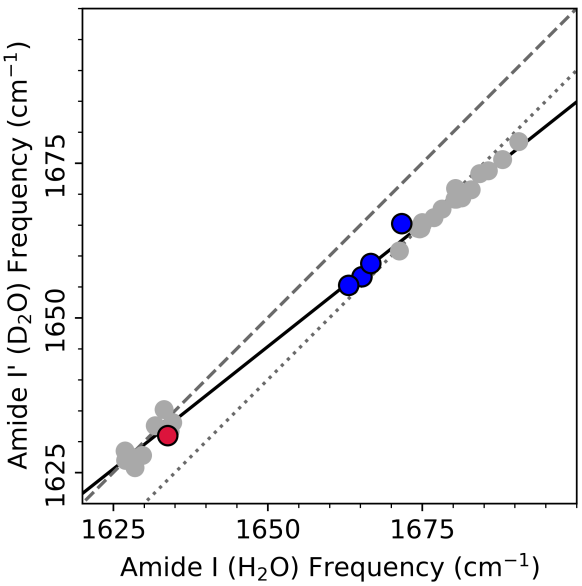


Figure 9: Scatter plot of Amide I frequencies in H_2O (horizontal axis) and D_2O (vertical axis); gray points correspond to dipeptide data from Ref.;[68] blue points are for individual isotope reverse-labels for NuG2b; the red point is for the NuG2b main band. The dashed gray line is the diagonal; dotted gray indicates a uniform 10 cm^{-1} shift; the solid black line is the best fit line from Ref.[68]

an average shift of roughly 10 cm^{-1} to the Amide I band, we found recently with a library of both capped and uncapped dipeptides that this shift varied systematically with the Amide I oscillator frequency.[68] This trend is illustrated in Figure 9, where gray data points indicate dipeptide frequencies measured in H_2O (horizontal axis) versus D_2O (vertical axis); all data are taken from Ref.[68] The solid black line represents the best-fit line from Ref.[68] for the dipeptide data set. Dashed and dotted gray lines represent, respectively, the diagonal (no shift) line and a uniform shift of -10 cm^{-1} between H_2O and D_2O. Whereas high-frequency oscillators undergo a frequency shift near of around 6 - 10 cm^{-1} upon deuteration, strongly red-shifted oscillators see almost no frequency shift on H/D exchange. In this data set, however, it was unclear whether or not this frequency-dependent shift is relevant to site-specific labels in proteins, or whether it might be an artifact of the small peptide fragments studied.

To test this, we tabulated reverse-label peak absorption frequencies for our four NuG2b

labels in H_2O versus D_2O (Frame (b) of Figure 6) and added them to the figure as blue data points. The red data point likewise represents the peak absorption frequency of the main Amide I band, which is dominated by the β sheet $\nu_\perp$ mode (see Frame (b) of Figure 4). At face value our reverse-label data set agrees quite well with the dipeptide-data trend, supporting the notion that H/D frequency shifts do indeed depend on the local environment of the amide unit. Note, however, that the main-band absorption frequencies (red data point) should be interpreted with some caution here, since these values will be influenced by site-to-site coupling in addition to the H/D shift for each individual amide unit.

Conclusions

In summary, our findings demonstrate that isotope-difference IR spectroscopy offers a viable route to monitoring protein structure in live bacterial cells, at least for recombinant proteins that express at high yields. Such insights are likely to be particularly valuable for intrinsically disordered proteins, whose structures can be highly sensitive to the molecular environment. In future work, it will be of great interest to see whether the technique can be extended to eukaryotic cells, building off of the existing strategies developed already for in-cell NMR. In combination with site-directed mutagenesis and nonlinear methods such as 2D IR spectroscopy, we anticipate that this approach will offer a valuable path to obtaining residue-specific insight into the impact of the cellular environment on protein structure and dynamics.

References

Acknowledgement

The authors thank Prof. Scott McLuckey (Purdue University) for helpful discussions regarding mass spectrometry data.

This work was supported by the National Science Foundation under Grant Number 2236625 and startup funding from Purdue University.

References

(1) Vuister, G. W.; Kim, S.-J.; Wu, C.; Bax, A. 2D and 3D NMR Study of Phenylalanine Residues in Proteins by Reverse Isotopic Labeling. *Journal of the American Chemical Society* **1994**, *116*, 9206–9210.

(2) Dyson, H. J.; Wright, P. E. Unfolded Proteins and Protein Folding Studied by NMR. *Chemical Reviews* **2004**, *104*, 3607–3622.

(3) Mielke, S. P.; Krishnan, V. Characterization of protein secondary structure from NMR chemical shifts. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2009**, *54*, 141–165.

(4) Bai, X.-c.; McMullan, G.; Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences* **2015**, *40*, 49–57.

(5) Sakakibara, D.; Sasaki, A.; Ikeya, T.; Hamatsu, J.; Hanashima, T.; Mishima, M.; Yoshimasu, M.; Hayashi, N.; Mikawa, T.; Wälchli, M.; Smith, B. O.; Shirakawa, M.; Güntert, P.; Ito, Y. Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **2009**, *458*, 102–105.

(6) Luchinat, E.; Banci, L. In-cell NMR: a topical review. *IUCrJ* **2017**, *4*, 108–118.

(7) Robinson, K. E.; Reardon, P. N.; Spicer, L. D. *Protein NMR Techniques*; Humana Press, 2011; p 261–277.

(8) Zhang, Z.; Zhao, Q.; Gong, Z.; Du, R.; Liu, M.; Zhang, Y.; Zhang, L.; Li, C. Progress, Challenges and Opportunities of NMR and XL-MS for Cellular Structural Biology. *JACS Au* **2024**, *4*, 369–383.

(9) Müntener, T.; Häussinger, D.; Selenko, P.; Theillet, F.-X. In-Cell Protein Structures from 2D NMR Experiments. *The Journal of Physical Chemistry Letters* **2016**, *7*, 2821–2825.

(10) Su, X.-C.; Chen, J.-L. Site-Specific Tagging of Proteins with Paramagnetic Ions for Determination of Protein Structures in Solution and in Cells. *Accounts of Chemical Research* **2019**, *52*, 1675–1686.

(11) Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M. NMR-Based Methods for Protein Analysis. *Analytical Chemistry* **2021**, *93*, 1866–1879, PMID: 33439619.

(12) Tanaka, T.; Ikeya, T.; Kamoshida, H.; Suemoto, Y.; Mishima, M.; Shirakawa, M.; Güntert, P.; Ito, Y. High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells. *Angewandte Chemie* **2019**, *131*, 7362–7366.

(13) Bai, Y.; Zhang, S.; Dong, H.; Liu, Y.; Liu, C.; Zhang, X. Advanced Techniques for Detecting Protein Misfolding and Aggregation in Cellular Environments. *Chemical Reviews* **2023**, *123*, 12254–12311.

(14) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; et al. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chemical Reviews* **2014**, *114*, 6661–6714.

(15) Sciolino, N.; Burz, D. S.; Shekhtman, A. In-Cell NMR Spectroscopy of Intrinsically Disordered Proteins. *Proteomics* **2019**, *19*, 1800055.

(16) Habashi, M.; Vutla, S.; Tripathi, K.; Senapati, S.; Chauhan, P. S.; Haviv-Chesner, A.; Richman, M.; Mohand, S.-A.; Dumulon-Perreault, V.; Mulamreddy, R.; Okun, E.; Chill, J. H.; Guérin, B.; Lubell, W. D.; Rahimipour, S. Early diagnosis and treatment of Alzheimer's disease by targeting toxic soluble Aβ oligomers. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2210766119.

(17) Uemura, N.; Uemura, M. T.; Luk, K. C.; Lee, V. M.-Y.; Trojanowski, J. Q. Cell-to-Cell Transmission of Tau and α-Synuclein. *Trends in Molecular Medicine* **2020**, *26*, 936–952.

(18) Uversky, V. N.; Davé, V.; Iakoucheva, L. M.; Malaney, P.; Metallo, S. J.; Pathak, R. R.; Joerger, A. C. Pathological Unfoldomics of Uncontrolled Chaos: Intrinsically Disordered Proteins and Human Diseases. *Chemical Reviews* **2014**, *114*, 6844–6879.

(19) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry* **2006**, *75*, 333–366.

(20) Szalai, A. M.; Zaza, C.; Stefani, F. D. Super-resolution FRET measurements. *Nanoscale* **2021**, *13*, 18421–18433.

(21) Dhar, A.; Girdhar, K.; Singh, D.; Gelman, H.; Ebbinghaus, S.; Gruebele, M. Protein Stability and Folding Kinetics in the Nucleus and Endoplasmic Reticulum of Eucaryotic Cells. *Biophysical Journal* **2011**, *101*, 421–430.

(22) Rust, M. J.; Bates, M.; Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **2006**, *3*, 793–796.

(23) Margineanu, A.; Chan, J. J.; Kelly, D. J.; Warren, S. C.; Flatters, D.; Kumar, S.; Katan, M.; Dunsby, C. W.; French, P. M. W. Screening for protein-protein interactions

using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). *Scientific Reports* **2016**, *6*, 28186.

(24) Godin, A. G.; Lounis, B.; Cognet, L. Super-resolution Microscopy Approaches for Live Cell Imaging. *Biophysical Journal* **2014**, *107*, 1777–1784.

(25) Ebbinghaus, S.; Dhar, A.; McDonald, J. D.; Gruebele, M. Protein folding stability and dynamics imaged in a living cell. *Nature Methods* **2010**, *7*, 319–323.

(26) Lippincott-Schwartz, J.; Snapp, E.; Kenworthy, A. Studying protein dynamics in living cells. *Nature Reviews Molecular Cell Biology* **2001**, *2*, 444–456.

(27) Ignatova, Z.; Gierasch, L. M. Monitoring protein stability and aggregationin vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences* **2003**, *101*, 523–528.

(28) Maldonado, A. Y.; Burz, D. S.; Shekhtman, A. In-cell NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2011**, *59*, 197–212.

(29) Banci, L.; Barbieri, L.; Bertini, I.; Luchinat, E.; Secci, E.; Zhao, Y.; Aricescu, A. R. Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nature Chemical Biology* **2013**, *9*, 297–299.

(30) Foster, M. P.; McElroy, C. A.; Amero, C. D. Solution NMR of Large Molecules and Assemblies. *Biochemistry* **2006**, *46*, 331–340.

(31) Barbieri, L.; Luchinat, E.; Banci, L. Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Scientific Reports* **2015**, *5*, 14456.

(32) Luchinat, E.; Barbieri, L.; Cremonini, M.; Banci, L. Protein in-cell NMR spectroscopy at 1.2 GHz. *Journal of Biomolecular NMR* **2021**, *75*, 97–107.

(33) Sonar, S.; Lee, C.-P.; Coleman, M.; Patel, N.; Liu, X.; Marti, T.; Khorana, H. G.; RajBhandary, U. L.; Rothschild, K. J. Site-directed isotope labelling and FTIR spectroscopy of bacteriorhodopsin. *Nature Structural Biology* **1994**, *1*, 512–517.

(34) Zuber, G.; Prestrelski, S.; Benedek, K. Application of Fourier transform infrared spectroscopy to studies of aqueous protein solutions. *Analytical Biochemistry* **1992**, *207*, 150–156.

(35) Miller, L. M.; Bourassa, M. W.; Smith, R. J. FTIR spectroscopic imaging of protein aggregation in living cells. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2013**, *1828*, 2339–2346.

(36) Barth, A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2007**, *1767*, 1073–1101.

(37) Miyazawa, T.; Blout, E. R. The Infrared Spectra of Polypeptides in Various Conformations: Amide I and II Bands. *Journal of the American Chemical Society* **1961**, *83*, 712–719.

(38) Baiz, C. R. et al. Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction. *Chemical Reviews* **2020**, *120*, 7152–7218.

(39) Reppert, M.; Tokmakoff, A. Computational Amide I 2D IR Spectroscopy as a Probe of Protein Structure and Dynamics. *Annual Review of Physical Chemistry* **2016**, *67*, 359–386.

(40) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy. *Methods* **2010**, *52*, 12–22.

(41) Lin, Y.-S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. Empirical Amide

I Vibrational Frequency Map: Application to 2D-IR Line Shapes for Isotope-Edited Membrane Peptide Bundles. *The Journal of Physical Chemistry B* **2008**, *113*, 592–602.

(42) Fang, C.; Hochstrasser, R. M. Two-Dimensional Infrared Spectra of the 13C18O Isotopomers of Alanine Residues in an α-Helix. *The Journal of Physical Chemistry B* **2005**, *109*, 18652–18663.

(43) Decatur, S. M. Elucidation of Residue-Level Structure and Dynamics of Polypeptides via Isotope-Edited Infrared Spectroscopy. *Accounts of Chemical Research* **2006**, *39*, 169–175.

(44) Buchanan, L. E.; Dunkelberger, E. B.; Zanni, M. T. *Protein Folding and Misfolding*; Springer Berlin Heidelberg, 2011; p 217–237.

(45) Ryan, M. J.; Gao, L.; Valiyaveetil, F. I.; Zanni, M. T.; Kananenka, A. A. Probing Ion Configurations in the KcsA Selectivity Filter with Single-Isotope Labels and 2D IR Spectroscopy. *Journal of the American Chemical Society* **2023**, *145*, 18529–18537.

(46) Cruz-Angeles, J.; Martínez, L. M.; Videa, M. Application of ATR-FTIR spectroscopy to the study of thermally induced changes in secondary structure of protein molecules in solid state. *Biopolymers* **2015**, *103*, 574–584.

(47) Samuel-Landtiser, M.; Zachariah, C.; Williams, C. R.; Edison, A. S.; Long, J. R. Incorporation of Isotopically Enriched Amino Acids. *Current Protocols in Protein Science* **2007**, *47*, 26.3.1–26.3.49.

(48) Luchinat, E.; Cremonini, M.; Banci, L. Radio Signals from Live Cells: The Coming of Age of In-Cell Solution NMR. *Chemical Reviews* **2022**, *122*, 9267–9306.

(49) Serber, Z.; Keatinge-Clay, A. T.; Ledwidge, R.; Kelly, A. E.; Miller, S. M.; Dötsch, V. High-Resolution Macromolecular NMR Spectroscopy Inside Living Cells. *Journal of the American Chemical Society* **2001**, *123*, 2446–2447.

(50) Serber, Z.; Corsini, L.; Durst, F.; Dötsch, V. *Nuclear Magnetic Resonance of Biological Macromolecules*; Methods in Enzymology; Academic Press, 2005; Vol. 394; pp 17–41.

(51) Theillet, F.-X. In-Cell Structural Biology by NMR: The Benefits of the Atomic Scale. *Chemical Reviews* **2022**, *122*, 9497–9570.

(52) Serber, Z.; Ledwidge, R.; Miller, S. M.; Dötsch, V. Evaluation of Parameters Critical to Observing Proteins Inside Living Escherichiacoli by In-Cell NMR Spectroscopy. *Journal of the American Chemical Society* **2001**, *123*, 8895–8901.

(53) Skinner, J. J.; Yu, W.; Gichana, E. K.; Baxa, M. C.; Hinshaw, J. R.; Freed, K. F.; Sosnick, T. R. Benchmarking all-atom simulations using hydrogen exchange. *Proceedings of the National Academy of Sciences* **2014**, *111*, 15975–15980.

(54) Reppert, M.; Roy, A. R.; Tokmakoff, A. Isotope-enriched protein standards for computational amide I spectroscopy. *The Journal of Chemical Physics* **2015**, *142*, 125104.

(55) Nauli, S.; Kuhlman, B.; Baker, D. Computer-based redesign of a protein folding pathway. *Nature Structural Biology* **2001**, *8*, 602–605.

(56) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.

(57) Mondal, S.; Shet, D.; Prasanna, C.; Atreya, H. S. High yield expression of proteins in E. coli for NMR studies. *Advances in Bioscience and Biotechnology* **2013**, *04*, 751–767.

(58) Sugiki, T.; Furuita, K.; Fujiwara, T.; Kojima, C. Amino Acid Selective 13C Labeling and 13C Scrambling Profile Analysis of Protein α and Side-Chain Carbons in Escherichia coli Utilized for Protein Nuclear Magnetic Resonance. *Biochemistry* **2018**, *57*, 3576–3589.

(59) Waugh, D. Genetic tools for selective labeling of proteins with 15N-amino acids. *Journal of Biomolecular NMR* **1996**, *8*, 184–192.

(60) Bellstedt, P.; Seiboth, T.; Häfner, S.; Kutscha, H.; Ramachandran, R.; Görlach, M. Resonance assignment for a particularly challenging protein based on systematic unlabeling of amino acids to complement incomplete NMR data sets. *Journal of Biomolecular NMR* **2013**, *57*, 65–72.

(61) Krishnarjuna, B.; Jaipuria, G.; Thakur, A.; D'Silva, P.; Atreya, H. S. Amino acid selective unlabeling for sequence specific resonance assignments in proteins. *Journal of Biomolecular NMR* **2010**, *49*, 39–51.

(62) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(63) Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D. Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2. *Protein Science* **2002**, *11*, 2924–2931.

(64) Bhanot, J. S.; Fabijanczuk, K. C.; Abdillahi, A. M.; Chao, H.-C.; Pizzala, N. J.; Londry, F. A.; Dziekonski, E. T.; Hager, J. W.; McLuckey, S. A. Adaptation and operation of a quadrupole/time-of-flight tandem mass spectrometer for high mass ion/ion reaction studies. *International Journal of Mass Spectrometry* **2022**, *478*, 116874.

(65) Hunt, N. T. Using 2D-IR Spectroscopy to Measure the Structure, Dynamics, and Intermolecular Interactions of Proteins in H2O. *Accounts of Chemical Research* **2024**, *57*, 685–692.

(66) Hume, S.; Hithell, G.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Hunt, N. T. Measuring proteins in H2O with 2D-IR spectroscopy. *Chemical Science* **2019**, *10*, 6448–6456.

(67) Ganim, Z.; Chung, H. S.; Smith, A. W.; DeFlores, L. P.; Jones, K. C.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Accounts of Chemical Research* **2008**, *41*, 432–441, PMID: 18288813.

(68) Chelius, K.; Wat, J. H.; Phadkule, A.; Reppert, M. Distinct electrostatic frequency tuning rates for amide I and amide I' vibrations. *The Journal of Chemical Physics* **2021**, *155*, 195101.

(69) Dicke, S. S.; Alperstein, A. M.; Schueler, K. L.; Stapleton, D. S.; Simonett, S. P.; Fields, C. R.; Chalyavi, F.; Keller, M. P.; Attie, A. D.; Zanni, M. T. Application of 2D IR Bioimaging: Hyperspectral Images of Formalin-Fixed Pancreatic Tissues and Observation of Slow Protein Degradation. *The Journal of Physical Chemistry B* **2021**, *125*, 9517–9525.

(70) Spector, I. C.; Olson, C. M.; Huber, C. J.; Massari, A. M. Simple fully reflective method of scatter reduction in 2D-IR spectroscopy. *Optics Letters* **2015**, *40*, 1850.

(71) Donaldson, P. M.; Howe, R. F.; Hawkins, A. P.; Towrie, M.; Greetham, G. M. Ultrafast 2D-IR spectroscopy of intensely optically scattering pelleted solid catalysts. *The Journal of Chemical Physics* **2023**, *158*, 114201.

(72) Procacci, B.; Rutherford, S. H.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Robinson, C. V.; Howle, C. R.; Hunt, N. T. Differentiation of bacterial spores via 2D-IR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2021**, *249*, 119319.

(73) Bloem, R.; Garrett-Roe, S.; Strzalka, H.; Hamm, P.; Donaldson, P. Enhancing signal detection and completely eliminating scattering using quasi-phase-cycling in 2D IR experiments. *Optics Express* **2010**, *18*, 27067.

(74) Butler, H. J.; Cameron, J. M.; Jenkins, C. A.; Hithell, G.; Hume, S.; Hunt, N. T.; Baker, M. J. Shining a light on clinical spectroscopy: Translation of diagnostic IR, 2D-IR and Raman spectroscopy towards the clinic. *Clinical Spectroscopy* **2019**, *1*, 100003.

(75) Chun, S. Y.; Son, M. K.; Park, C. R.; Lim, C.; Kim, H. I.; Kwak, K.; Cho, M. Di-

rect observation of protein structural transitions through entire amyloid aggregation processes in water using 2D-IR spectroscopy. *Chemical Science* **2022**, *13*, 4482–4489.

TOC Graphic