Supporting Information for "Native Conformation at Specific Residues in Recombinant Inclusion Body Protein in Whole Cells Determined with Solid-State Nuclear Magnetic Resonance Spectroscopy" by Jaime Curtis-Fisk, Ryan M. Spencer, and David P. Weliky

1. FHA2 sequence

GLF**G**AI**A**GFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING

KLNRVIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAE**L**LV

ALENQHTIDLTDSEMNKLFEKTRRQLRENAEEMGNGSFKIYHKCDNACIE

SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWVEHHHHHH

The first 185 residues in the sequence are the 185 N-terminal residues from the HA2 protein of the influenza virus, X31 strain, and the last 8 residues are non-native. The sequence is written in lines containing 50 residues and the Gly-1, Gly-4, Ala-7, and Leu-98 residues are in bold and underlined.

2. Expression and labeling of FHA2

The protocol for expression and labeling of FHA2 in inclusion bodies was similar to that described for FHA2 incorporated in bacterial membranes and subsequently purified in detergent and reconstituted in synthetic membranes, cf. Curtis-Fisk, J.; Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11320-11321 and Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. *Prot. Expr. Purif.* in press. Dr. Yeon-Kyun Shin at Iowa State University donated the FHA2 plasmid which contained the Lac promoter and kanamycin resistance and the plasmid was transformed into E. coli BL21(DE3) cells. The cells were first grown overnight to an OD_{600} of ~8 in 0.5 L media containing luria broth and 5 g glycerol. The cells were then switched into 0.5 L minimal media that contained 5 g glycerol. After resumption of growth in this media

for one hour, 30 mg aliquots of one or two labeled amino acids were added to the media

and FHA2 expression was induced for three hours with 1 mM isopropyl thiogalactoside.

The labeled amino acids were: (1) Gly-1 sample, 1-^{13}C Gly and ^{15}N-Leu; (2) Gly-4

sample, 1-^{13}C Gly and ^{15}N-Ala; (3) Ala-7 sample, 1-^{13}C Ala and ^{15}N-Gly; and (4) Leu-98

sample, 1-^{13}C, ^{15}N Leu. Expression was done at 37 ^{o}C rather than room temperature to

augment production of inclusion bodies, cf. Vera, A.; Gonzalez-Montalban, N.; Aris, A.;

Villaverde, A. *Biotechnology & Bioengineering* **2007**, *96*, 1101-1106. Because the

bacterial membrane space is limited, a greater fraction of FHA2 will be incorporated into

membranes at the beginning of the expression period and a greater fraction will be in

inclusion bodies at the end of the expression period. Complete labeling of the inclusion

body FHA2 therefore required additional aliquots of 30 mg/L of labeled amino acid at the

one and two hour times during the expression period. After three hours of expression, the

cells were centrifuged and the whole cell NMR sample was taken from the cell pellet.

The cell yield with this protocol was typically 10 g cells/L culture.

3. Separation and native purification of the bacterial membrane fraction of FHA2

and reconstitution of this fraction in synthetic membranes

The protocol for separation and native purification of the fraction of FHA2

incorporated in bacterial membranes has been previously described, cf. Curtis-Fisk, J.;

Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*,

11320-11321 and Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. *Prot. Expr. Purif.* in

press. The protocol relied on solubilization of this fraction in a lysis buffer that contained

0.5% w/v *N*-lauroylsarcosine detergent and subsequent affinity purification based on the

non-native histidine tag. Increased lysis time did not increase the amount of FHA2

purified from the soluble cell lysate and *N*-lauroylsarcosine is a very good detergent for solubilizing proteins in the bacterial membranes. The lysis was therefore considered to be complete in terms of breaking the cells and cell membranes and solubilizing the FHA2 which was embedded in the cell membranes. Membrane reconstitution of purified FHA2 also followed the previously published protocol. Briefly, a 5 mL solution was prepared which contained ~5 mg FHA2, 160 mg n-octyl-β-D-thioglucopyranoside detergent, 32 mg di-*O*-tetradecylphosphatidylcholine (DTPC) lipid, and 8 mg di-*O*-tetradecylphosphatidylglycerol (DTPG) lipid. The detergent was then removed by dialysis against a pH 5.0 buffer solution. The resultant membrane bilayer suspension with bound FHA2 was centrifuged and the pellet was transferred to a 4 mm diameter magic angle spinning (MAS) rotor that had ~40 µL active sample volume.

4. Preparation of inclusion body samples

Unlike the fraction of FHA2 incorporated in the bacterial membranes, the inclusion body FHA2 was not solubilized by the lysis buffer and was in the pellet obtained from centrifugation of the cell lysate at $50000g$ for 30 minutes. While optimizing the inclusion body protocol, pellets were also obtained with smaller centrifugation g forces and shorter times, but the SDS-PAGE gels of these pellets were very similar to the gel displayed in lane 3 of Fig. 1 in the main text. In addition, NMR samples were taken from different regions of the same pellet but the NMR spectra of these different samples were very similar. For the final protocol, the insoluble fraction NMR sample was taken from all regions of the pellet.

5. Solid-state nuclear magnetic resonance (NMR) experiments and analysis

Solid-state NMR rotational-echo double-resonance (REDOR) experiments were carried out using a 9.4 T spectrometer and a 4.0 mm MAS probe tuned to ^{13}C detection at 100.8 MHz, ^{1}H decoupling at 400.8 MHz, and ^{15}N dephasing at 40.6 MHz. The pulse sequence was ^{1}H-^{13}C cross-polarization followed by a REDOR dephasing period and ^{13}C-detection with ^{1}H decoupling during the latter two periods. Data were alternately acquired without (S_0) and with (S_1) the ^{15}N π pulses during the dephasing period and respectively represented the full ^{13}C signal and the ^{13}C signal minus ^{13}Cs close to ^{15}N nuclei. The experimental MAS frequency was 8.0 kHz, the dephasing time was 2.0 ms, the recycle delay was 1.0 s, and the ^{13}C chemical shifts were externally referenced to the methylene peak of adamantane at 40.5 ppm which allows direct comparison to databases of ^{13}C chemical shifts as a function of conformation, cf. Zhang, H. Y.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173-195 and Morcombe, C. R.; Zilm, K. W. *J. Magn. Reson*. **2003**, *162*, 479-486. Other details of the NMR experiments have been previously described, cf. Curtis-Fisk, J.; Preston, C.; Zheng, Z. X.; Worden, R. M.; Weliky, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11320-11321. Peak ^{13}C carbonyl chemical shifts were measured with ±0.3 ppm precision. During data acquisition, samples were cooled by flowing nitrogen gas whose temperature was –20 ^{o}C. For the inclusion body and whole cell samples, the NMR probe could not be tuned when the cooling gas temperature was higher than –20 ^{o}C presumably because of effects from the ~200 mM total salt concentration in these samples. This reasoning is evidenced by: (1) no problems with tuning the membrane-reconstituted samples at temperatures above –20 ^{o}C; and (2) the ~10 mM salt concentration in the membrane-reconstituted samples. In the future, it will

be possible to do the inclusion body and whole cell NMR experiments at temperatures higher than $-20\,^{\circ}\text{C}$ using recently developed NMR probes that minimize electric effects in conductive samples, cf. Stringer, J. A.; Bronnimann, C. E.; Mullen, C. G.; Zhou, D. H. H.; Stellfox, S. A.; Li, Y.; Williams, E. H.; Rienstra, C. M. *J. Magn. Reson.* **2005**, *173*, 40-48.

Because the FHA2 contained a $^{13}CO/^{15}N$ unique sequential pair, the $S_0 - S_1$ difference was predominantly the filtered signal of this pair, cf. Yang, J.; Parkanzky, P. D.; Bodner, M. L.; Duskin, C. G.; and Weliky, D. P. *J. Magn. Reson.* **2002,** *159,* 101-110. For spectra of labeled FHA2 which targeted the ^{13}CO of Gly-4, Ala-7, or Leu-98, the experimental S_1/S_0 integrated intensity ratio was respectively 0.93, 0.93, or 0.89 and correlated with the expected ratios of 0.94, 0.91, and 0.92 for putative 100% labeling. These expected ratios were based on an approximate model in which the S_1/S_0 intensity ratio was: (number of FHA2 residues of the same type as the ^{13}CO labeled amino acid minus one)/(number of FHA2 residues of the same type as the ^{13}CO labeled amino acid), e.g. 10/11 for Ala. The $S_0 - S_1$ signal for the sample targeting Leu-98 also had some contribution from the other Leus (~10% per Leu) in the sequence because of the 2.5 Å intra-residue $^{13}CO \cdots ^{15}N$ distance in $1\text{-}^{13}C$, ^{15}N-Leu, cf. Yang, J. Ph. D. thesis, Michigan State University, East Lansing, MI, 2003. These contributions are probably responsible for the experimental S_1/S_0 integrated intensity ratio which is smaller than the ratio predicted by the approximate model. Although some of these intra-residue Leu contributions would likely have helical shifts similar to the shift of Leu-98 (presuming that FHA2 remains folded), the contributions from other Leu residues would be dispersed over a range of non-helical shifts and would be less apparent in the overall appearance of

the difference spectrum. This idea is supported by Fig. 2b in which there is a large

apparent difference between the S_0 and S_1 intensities in the 177.5-179.5 ppm helical shift

region and much smaller difference in the non-helical lower shift region.

There will also be natural abundance contributions to the $S_0 - S_1$ difference signal.

A model is developed for these contributions in the context of the following

approximations: (1) the recombinant protein makes up a large portion of the mass of the

sample; and (2) only the recombinant protein is labeled and the labeling is close to 100%.

These approximations appeared to hold for the insoluble fraction of the cell lysate with

FHA2 labeling which targeted the Gly-4 or Ala-7 residues. The recombinant protein is

considered to have "D" residues that are ^{13}CO labeled and "E" residues that are ^{15}N

labeled. There is a single unique sequential ^{13}CO-^{15}N labeled pair and this pair

contributes 1.0 intensity to the $S_0 - S_1$ difference signal. The approximate contribution to

the difference signal intensity by natural abundance ^{15}N dephasing of the $D - 1$ other

labeled ^{13}COs will be $(D - 1) \times (1.1) \times 0.0037$ where the 1.1 factor considers the effect at

2 ms dephasing time of ^{15}N separated by one and two bonds from the labeled ^{13}CO and

0.0037 is the ^{15}N fractional natural abundance. Similarly, the approximate contribution to

the difference signal intensity from natural abundance ^{13}COs dephased by the $E - 1$ other

labeled ^{15}Ns is $(E - 1) \times (1.1) \times 0.0111$. The fractional natural abundance contribution to

the total $S_0 - S_1$ signal intensity is therefore:

$$\{[(D - 1) \times 0.00407] + [(E - 1) \times 0.0122]\}/$$
$$\{1.0 + [(D - 1) \times 0.00407] + [(E - 1) \times 0.0122]\}$$

For FHA2 with labeling which targeted the Gly-4 residue, $D = 16$, $E = 11$, and the ratio

was 0.15. A much larger natural abundance contribution to the $S_0 - S_1$ difference signal

will be observed if: (1) the recombinant protein is much larger than FHA2 and D and E

6

are much larger; (2) the recombinant protein is only a small fraction of the sample; and/or (3) proteins other than the recombinant protein are labeled during the expression period and are insoluble in the cell lysate.

In the context of the model, the lineshape of the natural abundance contribution to the $S_0 - S_1$ difference signal will primarily depend on the amino acid identities and the conformations of the residues in sequential pairs containing labeled residues. Because there will likely be a wide variety of amino acid types and perhaps conformations for these residues, there will likely be significant chemical shift dispersion in the lineshape. For large values of D and E, the $S_0 - S_1$ difference signal would therefore appear to contain a sharp signal from the ^{13}CO in the unique sequential pair and a broad signal from the contributions described in the aforementioned model.

Each of the membrane-reconstituted samples contained ~5 mg FHA2 and the corresponding spectrum was obtained with ~3 days of signal averaging. The signal-to-noises of the insoluble fraction and whole cell samples were typically comparable or higher than that of the corresponding membrane-reconstituted sample and were obtained in ~1½ days, cf. Fig. 1 in the main manuscript. There may be variation in motion and hence cross-polarization and REDOR efficiencies for the different sample types which would differentially affect signal intensities, but neglecting this potential variation, the signal-to-noises in the spectra suggest that the insoluble fraction and whole cell NMR samples contained 5-10 mg FHA2. The NMR active sample volume was ~40 µL which could likely contain ~50 mg of material. The overall analysis therefore suggested that 10-20% of the mass in the insoluble fraction and whole cell NMR samples was FHA2 and

7

this percentage was generally consistent with the apparent mass fraction of FHA2 in the gels in Fig. 1a of the main manuscript.

In the main manuscript, the reported distributions of ^{13}CO chemical shifts for Gly, Ala, and Leu residues in helical conformations were obtained from Table 7 in Zhang, H. Y.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173-195. The distributions were based on ~180 proteins for which both the ^{13}CO assignments and high-resolution structures were known. Inclusion of a residue in the helical distribution was based its hydrogen bonding and on its values of the backbone dihedral angles φ and ψ with typical helical ranges being $-120^{\circ} < \varphi < -34^{\circ}$ and $-80^{\circ} < \psi < -6^{\circ}$. These are generally consistent with the experimental ranges of dihedral angles in α helices, cf. Hovmoller, S; Zhou, T; Ohlson, T. *Acta Cryst.* **2002**, *D58*, 768-776. In the main manuscript, the reported ^{13}CO chemical shift ranges for β sheet/amyloid conformation were based on solid-state NMR-derived ^{13}CO shifts of the membrane-associated HIV fusion peptide, β amyloid peptide, and HET-s prion protein reported in references 5, 6, 19, and 20 from the main text. An increment of 2.0 ppm was added to the shifts from references 5 and 19 so that they would be referenced in the same way as the shifts in the helical distributions. The β sheet ^{13}CO shifts from these samples in ppm units were: Gly, 171.5, 170.3, 170.7, 171.4, 170.7, 172.0, 172.0; Ala, 174.2, 174.9, 175.6, 176.9, 174.7, 176.0, 175.2, 173.3; and Leu, 174.2, 174.7, 174.4, 174.8, 175.0, 173.0.