Selective and extensive ^{13}C labeling of a membrane protein for solid-state NMR investigations

M. Hong[a,*] & K. Jakes[b]

[a]*Departments of Polymer Science & Engineering and Chemistry, University of Massachusetts, Amherst, MA 01003, U.S.A.;* [b]*Department of Physiology and Biophysics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, U.S.A.*

Abstract

The selective and extensive ^{13}C labeling of mostly hydrophobic amino acid residues in a 25 kDa membrane protein, the colicin Ia channel domain, is reported. The novel ^{13}C labeling approach takes advantage of the amino acid biosynthetic pathways in bacteria and suppresses the synthesis of the amino acid products of the citric acid cycle. The selectivity and extensiveness of labeling significantly simplify the solid-state NMR spectra, reduce line broadening, and should permit the simultaneous measurement of multiple structural constraints. We show the assignment of most ^{13}C resonances to specific amino acid types based on the characteristic chemical shifts, the ^{13}C labeling pattern, and the amino acid composition of the protein. The assignment is partly confirmed by a 2D homonuclear double-quantum-filter experiment under magic-angle spinning. The high sensitivity and spectral resolution attained with this ^{13}C-labeling protocol, which is termed TEASE for *ten-a*mino acid *s*elective and *e*xtensive labeling, are demonstrated.

Structure determination of insoluble proteins such as membrane proteins and protein aggregates has been receiving increasing attention, since structural information is recognized to be important for understanding the functions of these proteins. Solid-state NMR is a powerful spectroscopic tool for studying the conformation and dynamics of such disordered biological systems in situ (Creuzet et al., 1991; Shon et al., 1991; Ketchem et al., 1993; McDowell et al., 1996). Recently, we showed (Hong, 1999b) that such solid-state NMR structure investigations can be made more efficient by using selectively and extensively ^{13}C labeled proteins (LeMaster and Kushlan, 1996; Lee et al., 1997). Supplying specifically ^{13}C-labeled glucose or glycerol as the sole carbon source in the expression medium, we can label certain carbon sites with ^{13}C at high levels while the other sites remain completely unlabeled. The selectivity of this ^{13}C labeling scheme

reduces line broadening, which would be encountered in uniformly ^{13}C-labeled proteins, simplifies the NMR spectra, facilitates resonance assignment, and potentially expedites the measurement of long-range distances. Meanwhile, the extensiveness of labeling allows multiple structural constraints to be obtained simultaneously from each experiment. Using selectively and extensively ^{13}C-labeled ubiquitin ($M_r = 8565$), we demonstrated the measurement of multiple backbone ϕ torsion angles and several resonance assignment techniques (Hong, 1999b).

The two original labeling schemes, which make use of [1-^{13}C] glucose and [2-^{13}C] glycerol as the carbon precursors, result in relatively extensive labeling of the amino acids produced from the citric acid cycle. This is due to the cyclic nature of the enzymatic reactions, which spreads the single label in the starting compound to multiple positions in the carbon skeletons of these ten amino acids. To simplify the NMR experiments, it would be advantageous to filter out the ^{13}C signals of these extensively la-

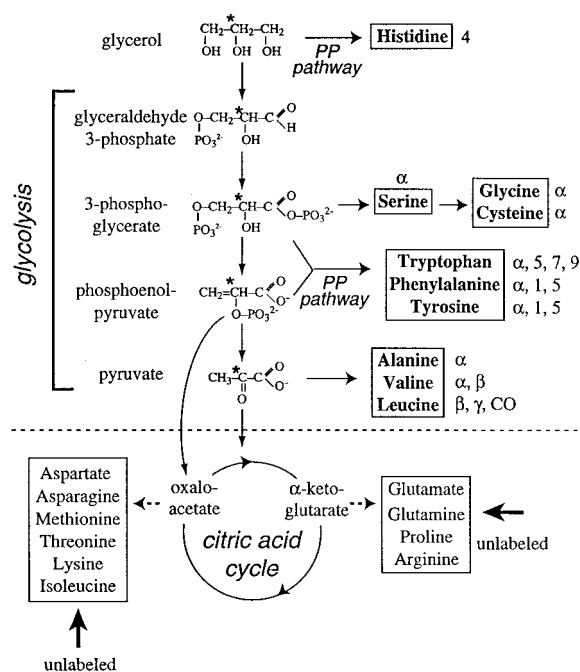*To whom correspondence should be addressed. E-mail: mhong@chem.umass.edu.

72

Figure 1. Amino acid biosynthetic pathways in bacteria, utilizing glycerol as the main carbon source. To achieve selective ^{13}C labeling of hydrophobic amino acids, which are largely produced from the glycolytic and the pentose phosphate (PP) pathways, the growth media is supplemented with unlabeled amino acid products of the citric acid cycle.
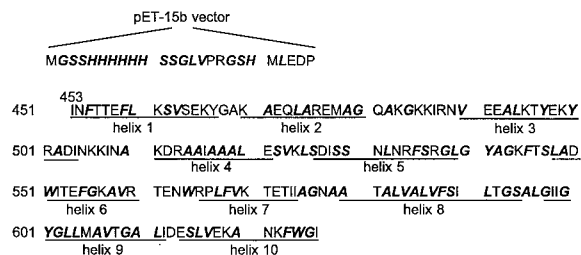

Figure 2. Amino acid sequence of the colicin Ia channel domain used in the NMR experiments. A sequence of 25 amino acids from the pET-15b vector is appended to the N-terminus of the protein to facilitate purification. The total molecular weight is 25 082 Da. Helix assignments are according to Wiener et al. (1997). The TEASE ^{13}C-labeled amino acids account for 110 (99) of the 198 (173) residues in the protein (numbers in brackets refer to the amino acids of the colicin Ia channel domain in the absence of the pET-15b vector). The distribution of the labeled amino acids is: A: 25, L: 19 (17), G: 17 (14), S: 16 (11), V: 10 (9), F: 8, H: 7, Y: 5, W: 3. The distribution of the unlabeled amino acids is: K: 17, I: 14, E: 13 (12), T: 11, N: 9, R: 9 (8), D: 6 (5), M: 4 (2), P: 3 (1), Q: 2. The labeled amino acids are shown in bold italics.

beled amino acids. This is particularly relevant to the study of membrane proteins, not only because of the severe resonance overlap resulting from the high molecular weights of these proteins, but also because the membrane-embedded segments of these molecules predominantly consist of hydrophobic amino acids, which are synthesized outside the citric acid cycle (Figure 1). The selective ^{13}C labeling of these hydrophobic amino acids should largely preserve the efficiency of the structure determination, since it is still possible to measure multiple structural constraints of the membrane-bound segments simultaneously. In this communication, we present a variant of the selective and extensive ^{13}C labeling approach that allows us to achieve this goal. The 10 predominantly hydrophobic amino acids to be labeled include Gly, Cys, Trp, Phe, Tyr, Ala, Val, Leu, Ser, and His.

The new ^{13}C labeling approach exploits the linear parts of the amino acid biosynthetic pathways (Figure 1) (Lehninger et al., 1993). Amino acids produced from glycolysis and the pentose phosphate pathways are to be labeled selectively and extensively by the specifically ^{13}C-labeled glycerol (or glucose).

Meanwhile, the production of the amino acids from the citric acid cycle, including Asp, Asn, Met, Thr, Lys, Ile, Glu, Gln, Pro, and Arg, can be blocked by supplementing the expression media with the unlabeled forms of these amino acids. With sufficient quantities of these unlabeled amino acids, the ^{13}C-labeled amino acid precursors, oxaloacetate and α-ketoglutarate, should be prevented from being converted into the amino acids. For convenience, we call this ^{13}C labeling scheme TEASE, which stands for *te*n-*a*mino-acid *s*elective and *e*xtensive labeling.

We tested the TEASE ^{13}C labeling protocol on the channel domain of colicin Ia, a bactericidal protein that forms voltage-gated channels in the plasma membranes of sensitive bacterial cells (Cramer et al., 1995). The C-terminal colicin Ia fragment was cloned by using site-directed mutagenesis to create a *Bam*HI site overlapping the codons for amino acids 450–452 in the colicin Ia gene cloned in pUC19 (Jakes et al., 1998). The resulting plasmid was digested with *Bam*HI and the fragment containing the 3′-end of the colicin Ia gene, beginning with the codon for Asp 451, as well as the Ia immunity gene, was ligated at the *Bam*HI site of the expression vector pET-15b (Novagen) to create pKSJ120. The primary sequence of the resulting protein is displayed in Figure 2.

The carboxy-terminal colicin Ia channel domain, with an amino-terminal His$_6$ tag, was expressed from pKSJ120-containing *E. coli* BL21(DE3). Cells were grown overnight at 37 °C in 10 ml of a modified M9 medium containing, per liter, 1 g NH$_4$Cl, 4 g glycerol,

3 g KH$_2$PO$_4$, 6 g Na$_2$HPO$_4$, 15 mg CaCl$_2$, 1 mM MgSO$_4$, the unlabeled amino acids Asp, Asn, Arg, Gln, Glu, Ile, Lys, Met, Pro and Thr at 150 µg/ml each, and 100 µg/ml ampicillin. The cells were pelleted and resuspended in 10 ml of the same medium, but containing ^{15}NH$_4$Cl and [2-^{13}C] glycerol (Cambridge Isotope Laboratories). Those cells were used to inoculate 300 ml of the stable-isotope-containing medium. The culture was grown at 37 °C to OD$_{660}$ = 0.4 and then induced with 1 mM IPTG. Cells were harvested after 3 h of induction at 37 °C. The soluble His-tagged colicin Ia channel domain was purified on His-Bind metal chelation resin (Novagen) exactly as specified in the Novagen pET System Manual, except that the wash step was performed at 40 mM imidizole. The yield of the pure soluble protein was approximately 7 mg from 300 ml of culture. A comparable amount of protein was unrecoverable from inclusion bodies. Purified protein eluted from the His-Bind resin in 1 M imidizole buffer was dialyzed extensively against distilled water and lyophilized.

A ^{13}C CPMAS spectrum of the TEASE ^{13}C-labeled colicin Ia channel domain (M_r = 25 082) is displayed in Figure 3a. It was obtained after 4096 scans on about 3 mg (equivalent to 0.14 µmol) of 30% (w/w) hydrated protein in the absence of lipids. For identifying the level of ^{13}C incorporation and the labeling selectivity, it is not necessary to reconstitute the protein into lipid bilayers. The spectrum was acquired on a Bruker DSX-300 spectrometer using a triple-resonance MAS probe equipped with a 4-mm spinner. All peaks were clearly visible after 128 scans, indicating the high levels of ^{13}C incorporation by this selective and extensive approach. Three dominant C$^\alpha$ peaks are observed above ∼ 43 ppm, which can be assigned to Ala (53 ppm), Val (66 ppm), and Gly (46 ppm) based on the characteristic chemical shifts of these amino acids in proteins. The resonances between the Ala and Val peaks are not clearly resolved but may be attributed to the C$^\alpha$ of Ser, Phe, Tyr, and Trp. Below 43 ppm, in the typical side-chain chemical shift region, the spectrum is even simpler, containing three highly resolved peaks with linewidths between 1.5 ppm and 2.7 ppm. According to the known labeling pattern of [2-^{13}C] glycerol (Hong, 1999b) and the characteristic chemical shifts of the amino acids, these resonances can be assigned to Leu C$^\beta$ (40 ppm), Leu C$^\gamma$ (26 ppm) and Val C$^\beta$ (31 ppm). The narrow widths of the C$^\beta$ peaks, whose chemical shifts are known to be influenced by the local backbone conformation (Spera and Bax, 1991; Havlin et al., 1997), strongly suggest that

these residues are located in similar secondary structures of the protein. This is consistent with the crystal structure of the soluble form of colicin Ia (Wiener et al., 1997), which indicates that the channel domain is predominantly α-helical and has few sheet or turn elements (Figure 2).

To confirm the assignment of the aliphatic peaks, we further carried out a 2D ^{13}C double-quantum-filter experiment. According to the labeling pattern of [2-^{13}C] glycerol, Ala, Ser, and Gly are labeled only at the ^{13}C$^\alpha$ site, while Val and Leu are doubly ^{13}C-labeled at directly bonded positions (Hong, 1999b). Thus a double-quantum-filter experiment should selectively excite the Val and Leu ^{13}C signals. The experiment was carried out using a dipolar-mediated INADEQUATE pulse sequence (Hong, 1999a), in which ^{13}C-^{13}C dipolar coupling was recoupled by a sevenfold phase-incremented pulse sequence, CMR7 (Rienstra et al., 1998). This recoupled dipolar interaction drives the excitation and reconversion of ^{13}C-^{13}C double-quantum coherence. A double-quantum excitation time of 571 µs, which is sufficient for recoupling the one-bond C-C dipolar interactions, was used. The resulting spectrum in Figure 3b exhibits two pairs of coupled ^{13}C spins, confirming our assignment of the Val C$^\alpha$/C$^\beta$ and Leu C$^\beta$/C$^\gamma$ resonances unambiguously.

Between 105 and 145 ppm, several aromatic carbon signals are observed. The main contributions to these signals are the phenylene C1 and C5 carbons of eight Phe residues and five Tyr residues. In addition, there are weak contributions from the C5, C7, and C9 carbons of the three Trp residues. The frequencies of these peaks agree well with the average chemical shifts of the carbon sites (Wüthrich, 1986; Ye et al., 1993) that are known to be labeled in these aromatic amino acids.

The ^{13}C spectra in Figure 3 are remarkably simple for a 25 kDa protein. This simplicity attests to the high selectivity of the TEASE ^{13}C-labeling scheme. Important potential applications of this labeling method are manifold. For example, the selective and extensive labeling of many C$^\alpha$ sites is particularly suitable for φ-torsion angle measurements using the HNCH correlation technique (Hong et al., 1997). In addition, the C$^\alpha$ chemical shifts are indicators of the secondary structure of the protein. The TEASE ^{13}C-labeling scheme can also be exploited in static NMR experiments of oriented samples, where segmental orientations are probed by anisotropic interactions (Marassi and Opella, 1998); since only isolated ^{13}C spins or isolated ^{13}C spin pairs are present in the TEASE-labeled
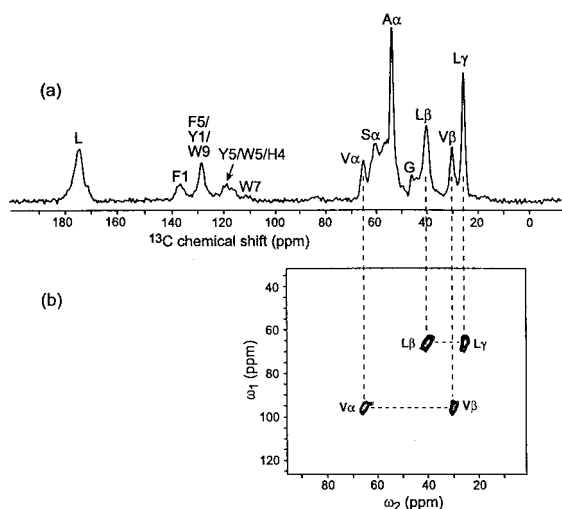
Figure 3. ^{13}C MAS spectra of the colicin Ia channel domain. The sample amount was 3 mg and the spinning speed was 7 kHz. (a) CP-MAS spectrum, 4096 scans. Note that the numbers associated with the aromatic residues refer to the labeled phenylene ring carbons. (b) 2D ^{13}C double-quantum-filtered INADEQUATE spectrum. The CMR7 sequence was used to recouple the ^{13}C-^{13}C dipolar interaction, which drives the double-quantum excitation and reconversion. The connectivity pattern confirms the assignment of all resolved aliphatic peaks in the 1D spectrum. The acquisition time was 21 h.

protein, dipolar broadening is minimized. Finally, the TEASE ^{13}C-labeling approach has a practical advantage over other selective labeling schemes in that the protein expression is more efficient than in minimal media due to the addition of 10 unlabeled amino acids.

In addition to ^{13}C labeling, we also analyzed the ^{15}N labeling level of this protein by acquiring ^{15}N MAS spectra and ^{15}N-^{13}C dipolar filtered ^{13}C spectra (not shown). The ^{15}N-^{13}C double- and zero-quantum filtered spectrum was obtained using a ^{15}N-^{13}C correlation pulse sequence (Hong and Griffin, 1998), except that the evolution period was set to zero to make it a 1D experiment. Since the ^{15}N-^{13}C filter efficiency is quite consistently around 25% under our experimental conditions, we can estimate the approximate level of ^{15}N incorporation by measuring the relative intensities of the ^{15}N-^{13}C-filtered ^{13}C spectrum with respect to the unfiltered spectrum. To rule out other factors that attenuate the signal intensities, we used the ^{15}N-^{13}C filtered ^{13}C spectrum of uniformly labeled and well-expressed ubiquitin as a reference, assuming that ubiquitin is 100% ^{15}N-labeled. Based on these considerations we found a ^{15}N labeling level of around 50% for the colicin Ia channel domain. The incomplete ^{15}N labeling (i.e. < 100%) may result from the presence of naturally occurring ^{14}N in the unlabeled

amino acids that are supplemented to the growth media. In particular, glutamate and glutamine are a source of nitrogen for many amino acids. This suggests that the ^{15}N labeling level can be improved by increasing the ratio of ^{15}NH$_4$Cl to the unlabeled amino acids and by using ^{15}N-labeled glutamate and glutamine in the growth media.

References

Cramer, W.A., Heymann, J.B., Schendel, S.L., Deriy, B.N., Cohen, F.S., Elkins, P.A. and Stauffacher, C.V. (1995) *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 611–641.

Creuzet, F., McDermott, A., Gebhard, R., van der Hoef, K., Spijker-Assink, M.B., Herzfeld, J., Lugtenburg, J., Levitt, M.H. and Griffin, R.G. (1991) *Science*, **251**, 783–786.

Havlin, R.H., Le, H., Laws, D.D., de Dios, A.C. and Oldfield, E. (1997) *J. Am. Chem. Soc.*, **119**, 11951–11958.

Hong, M. (1999a) *J. Magn. Reson.*, **136**, 86-91.

Hong, M. (1999b) J. Magn. Reson., submitted.

Hong, M. and Griffin, R.G. (1998) *J. Am. Chem. Soc.*, **120**, 7113–7114.

Hong, M., Gross, J.D. and Griffin, R.G. (1997) *J. Phys. Chem. B*, **101**, 5869–5874.

Jakes, K.S., Kienker, P.K., Slatin, S.L. and Finkelstein, A. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 4321–4326.

Ketcham, R.R., Hu, W. and Cross, T.A. (1993) *Science*, **261**, 1457–1460.

Lee, A.L., Urbauer, J.L. and Wand, A.J. (1997) *J. Biomol. NMR*, **9**, 437–440.

Lehninger, A.L., Nelson, D.L. and Cox, M.M. (1993) *Principles of Biochemistry,* 2nd ed., Worth Publishers, New York, NY.

LeMaster, D.M. and Kushlan, D.M. (1996) *J. Am. Chem. Soc.*, **118**, 9255–9264.

Marassi, F. and Opella, S. (1998) *Curr. Opin. Struct. Biol.*, **8**, 640–648.

McDowell, L.M., Klug, C.A., Beusen, D.D. and Schaefer, J. (1996) *Biochemistry*, **35**, 5395–5403.

Rienstra, C.M., Hatcher, M.E., Mueller, L.J., Sun, B.Q., Fesik, S.W. and Griffin, R.G. (1998) *J. Am. Chem. Soc.*, **120**, 10602–10612.

Shon, K.-J., Kim, Y., Colnago, L.A. and Opella, S.J. (1991) *Science*, **252**, 1303–1305.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Wiener, M., Freymann, D., Ghosh, P. and Stroud, R.M. (1997) *Nature*, **385**, 461–464.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* Wiley, New York, NY.

Ye, C., Fu, R., Hu, J., Hou, L. and Ding, S. (1993) *Magn. Reson. Chem.*, **31**, 699–704.