A Method for Assigning Phenylalanines in Proteins

Hong Wang, David A. Janowick, Jeffery M. Schkeryantz,
Xiaohong Liu,[†] and Stephen W. Fesik*

*Abbott Laboratories, Pharmaceutical Discovery Division
D47G, AP10, Abbott Park, Illinois 60064*

*Received November 11, 1998
Revised Manuscript Received January 4, 1999*

The ability to label proteins with NMR active isotopes has greatly facilitated the study of large proteins by NMR spectroscopy. Uniform ^{13}C and ^{15}N isotope enrichment combined with heteronuclear multidimensional NMR experiments has allowed the determination of NMR structures for proteins up to approximately 20 kDa.[1-3] In addition, the backbone resonances of even larger proteins have been assigned through the use of deuterium labeling,[4-6] and the structures of a few 30-kDa proteins have been solved.[7-9] However, the stucture determination of such large molecules is difficult because it requires the assignment of side-chain resonances and the identification of NOEs to side-chain protons that are removed or diluted by deuteration. To overcome this problem, labeling methods[10-12] have been utilized that incorporate protons into the methyl groups of valine, leucine, and isoleucine residues in an otherwise fully deuterated protein. These methyl groups are easier to assign, and NOEs involving these methyl groups are large due to favorable relaxation and reduced spin diffusion. Furthermore, since these hydrophobic residues are usually buried in the interior of a protein, their methyl protons exhibit a large number of long-range NOEs that can help determine the fold of a protein.

Phenylalanine is another hydrophobic residue that is found within the core of proteins and typically exhibits a large number of long-range NOEs that could be helpful in defining the overall fold. However, the assignment of the aromatic resonances of phenylalanine is often difficult due to the lack of ^{1}H and ^{13}C spectral dispersion, the broad line widths of the protons (especially when attached to ^{13}C), and the potential for nonfirst-order ^{13}C–^{13}C coupling. To assign the aromatic protons of phenylalanine, Bax and co-workers[13] proposed a reverse isotopic labeling scheme that incorporates unlabeled phenylalanine into an otherwise uniformly ^{13}C-labeled protein. Although the reverse labeling method alleviates the nonfirst-order ^{13}C–^{13}C coupling and reduces the proton line widths compared to the uniformly ^{13}C-labeled
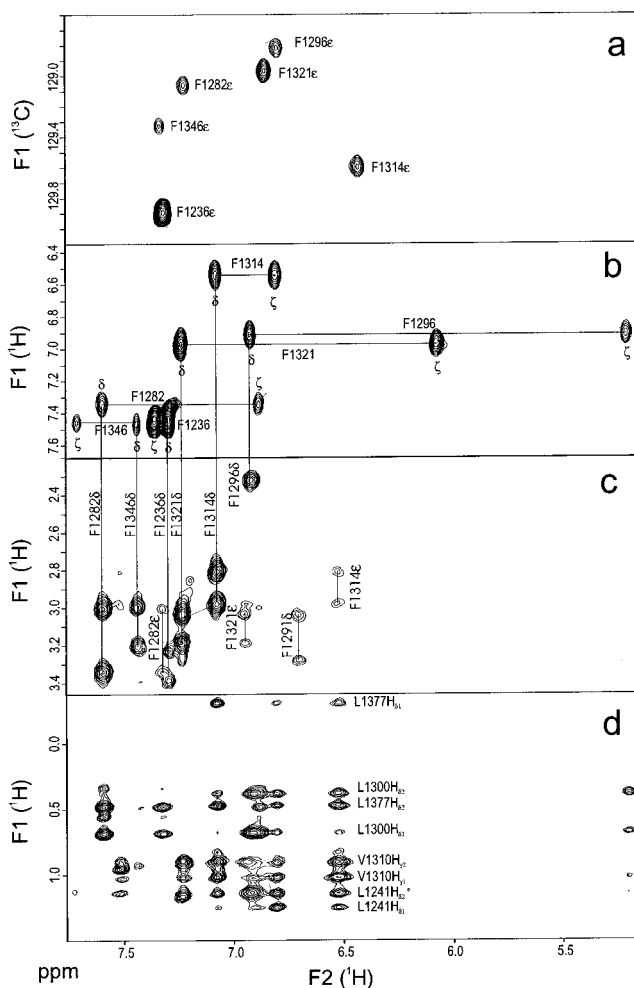
(1) Clore, G. M.; Gronenborn, A. *Science* **1991**, *252*, 795–796.
(2) Bax, A.; Grzesiek, S. *Acc. Chem. Res.* **1993**, *26*, 131–138.
(3) Clore, G. M.; Gronenborn, A. M. *Methods Enzymol.* **1994**, *239*, 349–363.
(4) Grzesiek, S.; Anglister, J.; Ren, H.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 4369–4370.
(5) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655–11666.
(6) Yamazaki, T.; Lee, W.; Revinton, M.; Mattiello, D. L.; Dahlquist, F. W.; Arrowsmith, C. H.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 6464–6465.
(7) Garret, D. S.; Seok, Y.-J.; Liao, D.-I., Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1997**, *36*, 2517–2530.
(8) Martin, J. R.; Mulder, F. A.; Karimi-Nejad, Y.; van der Zwan, J.; Mariani, M.; Schipper, D.; Boelens, R. *Structure* **1997**, *5*, 521–532.
(9) Yu, L.; Petros, A. M.; Schnuchel, A.; Zhong, P.; Severin, J. M.; Walter, K.; Holzman, T. F.; Fesik, S. W. *Nat. Struct. Biol.* **1997**, *4*, 483–489.
(10) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, E. *J. Mol. Biol.* **1996**, *263*, 627–636.
(11) Gardner, K. H.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 7599–7600.
(12) Zwahlen, C.; Vincent, S. J. F.; Gardner, K. H.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 4825–4831.
(13) Vuister, G. W.; Kim, S.-J.; Wu, C.; Bax, A. *J. Am. Chem. Soc.* **1994**, *116*, 9206–9210.

Figure 1. (a) Aromatic region of a ^{13}C/^{1}H HSQC spectrum of the N-terminal DH domain of Trio recorded using the TROSY method[9] to obtain narrower ^{13}C line widths. The digital resolution for the spectrum was 0.05 ppm in the ^{13}C dimension and 0.006 ppm in the ^{1}H dimension. The total acquisition time was 10 h. (b) A ^{13}C-edited (F1) and ^{12}C-filtered (F2) TOCSY spectrum of the N-terminal DH domain of Trio. The chemical shifts of the ϵ aromatic protons are labeled in the F1 dimension, and the chemical shifts of the δ and ζ aromatic protons are observed in the F2 dimension. The peaks of the same aromatic spin system are connected by a horizontal line. The ^{13}C-editing and ^{12}C-filtering were optimized for the aromatic ^{1}H–^{13}C scalar coupling using a 3 ms delay time. The spectrum was recorded in 10 h with a t_1(max) = 20 ms and a t_2(max) = 104 ms (c) A region of a 2D NOESY spectrum of the DH domain showing NOEs between the phenylalanine H$^\beta$ protons and the aromatic protons of the same residue. (d) A region of the a 2D NOESY spectrum showing NOEs between phenylalanine aromatic protons and methyl protons of valine and leucine residues in the DH domain. The NOEs involving H$^\epsilon$ of F1314 are labeled. All spectra were recorded on a Bruker DRX800 at 32 °C with a highly deuterated DH domain sample (0.4 mM, pH 6.5) that contains protonated phenylalanine residues ^{13}C-labeled at the ϵ aromatic position and valine and leucine residues that contain protonated methyl groups.

protein, overlap problems still exist for the phenylalanine ring protons which typically resonate within a small frequency range. Here we describe an approach for rapidly assigning the aromatic protons of phenylalanine in proteins. These assignments are useful for interpreting long-range NOEs involving phenylalanines which will aid in determining the protein fold.

Our approach involves the use of phenylalanine residues that are ^{13}C-labeled only at the ϵ aromatic position. The ϵ-^{13}C-labeled

phenylalanine was prepared from ϵ-^{13}C-labeled tyrosine (Cambridge Isotope, Inc.), using slight modifications of Viswanatha and Hruby's procedure.[14,15] This labeling scheme eliminates the potential for strong ^{13}C–^{13}C coupling while affording reduced line widths of the H$^\delta$ and H$^\epsilon$ protons compared to that observed for uniformly ^{13}C-labeled proteins. In addition, since phenylalanine is only labeled at the ϵ position, spectral overlap of the ^{13}C signals is reduced, and by ^{13}C-editing and ^{12}C-filtering,[16–18] the aromatic protons can be readily resolved.

The approach is demonstrated in our recent structure determination of the N-terminal Dbl homology (DH) domain of Trio,[19] a guanine nucleotide exchange factor for the GTPase Rac1.[20,21] This DH domain is a 21-kDa all α-helical protein that contains seven phenylalanine residues. Because of the signal overlap and broad NMR line widths, the aromatic protons of the phenylalanine residues could not be assigned by conventional methods using a uniformly ^{13}C-labeled sample. To assign the aromatic resonances of the phenylalanines and to rapidly determine the protein fold, we produced a perdeuterated sample of the DH domain that contained protonated phenylalanines with ^{13}C labeling at the ϵ position.[22] The protein also contained ^{13}C-labeled valine and leucine residues with protonated methyl groups. Figure 1a depicts the aromatic region of the ^{1}H/^{13}C correlation spectrum recorded using the TROSY method.[23] The absence of ^{13}C–^{13}C couplings allowed a high resolution spectrum to be recorded with good signal-to-noise ratio without the need of a constant time evolution period. Six out of the seven phenylalanines in the protein are clearly separated and easily identified in the spectrum. The ϵ signals of F1291 were not observed, presumably because of exchange broadening. To assign the aromatic spin systems, a ^{13}C-edited and ^{12}C-filtered 2D TOCSY spectrum (Figure 1b) was

(14) Viswanatha, V.; Hruby, V. J. *J. Org. Chem.* **1980**, *45*, 2010–2012.

(15) Method used to generate ϵ-^{13}C-labeled phenylalanine. Reagents and conditions: (a) CuSO$_4$, NaOH/H$_2$O, 55 °C, 30 min. (b) 2 M NaOH, DMF, 5-chloro-1-phenyl-1H-tetrazole, 18-crown-6, 0 °C, 12 h. (c) EDTA, H$_2$O, 95 °C, 30 min, 49% overall. (d) 5% Pd/C, cyclohexene, EtOH/AcOH, reflux, 2h, 42%.

(16) Wörgötter, E.; Wagner, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1986**, *108*, 6162–6164.

(17) Gemmecker, G.; Olejniczak, E. T.; Fesik, S. W. *J. Magn. Reson.* **1992**, *96*, 199–203.

(18) Bax, A.; Grzesiek, S.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. A* **1994**, *106*, 269–273.

(19) Liu, X.; Wang, H.; Eberstadt, M.; Schnuchel, A.; Olejniczak, E. T.; Meadows, R. P.; Schkeryantz, J. M.; Janowick, D. A.; Harlan, J. E.; Haris, E. A. S.; Staunton, D. E.; Fesik, S. W. *Cell* **1998**, *95*, 269–277.

(20) Debant, A.; Serra-pages, C.; Seipel, K.; O'Brien, S.; Tang, M.; Park, S.; Streuli, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5466–5471.

(21) Bellanger, J.; Lazaro, J.; Diriong, S.; Fernandez, A.; Lamb, N.; Debant, A. *Oncogene* **1998**, *16*, 147–152.

(22) The N-terminal DH domain was subcloned into the pET30b vector (Novagen, Inc.) and transformed into *Escherichia coli* strain BL21(DE3) for the expression of the recombinant protein with a C terminal His tag. The *E. coli* were grown in a minimal medium in ^{2}H$_2$O (>98%) containing uniformly ^{13}C- and ^{2}H-labeled glucose (Cambridge Isotopes) supplemented with methyl-protonated valine (Cambridge Isotopes) and ϵ-^{13}C-labeled phenylalanine. The recombinant DH domain was purified by affinity chromatography on a nickel-IDA column (Invitrogen, Inc.). NMR samples contained 0.4 mM protein in 50 mM sodium phosphate (pH 6.5) and 20 mM NaCl in ^{2}H$_2$O.

(23) Pervushin, K.; Rick, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366–12371.

recorded. In this spectrum, the chemical shifts of the H$^\epsilon$ protons are frequency-labeled in the ^{13}C-edited F1 dimension, and the chemical shifts of the scalar coupled H$^\delta$ and H$^\zeta$ protons are recorded in the ^{12}C-filtered F2 dimension. The ability to use ^{13}C-editing and ^{12}C-filtering greatly reduces the spectral overlap in the TOCSY spectrum, allowing the rapid identification of the spin systems. The ^{12}C-filtering also eliminates the strong diagonal peaks that could obscure the assignments of the H$^\delta$ and H$^\zeta$ protons that resonate close to H$^\epsilon$ (e.g., the H$^\delta$ of F1296 and F1346 in the DH domain). The sequence specific assignments of the aromatic spin systems can be obtained by connecting the H$^\delta$ protons to the H$^\beta$ and/or H$^\alpha$ protons in a 2D NOESY spectrum (Figure 1c). For the DH domain, we illustrated the utility of the method for assigning the phenylalanines in a predominantly deuterated protein. However, the same strategy could be employed in NMR studies of a fully protonated, ^{15}N-,^{13}C-labeled protein in which ϵ-^{13}C-labeled phenylalanine was incorporated. In this case, the ^{13}C-edited and ^{12}C-filtered TOCSY experiment will eliminate all other aromatic signals except that of the phenylalanines, and a doubly ^{12}C-filtered 2D NOESY experiment[13] could be used to connect the aromatic spin systems of phenylalanine to their H$^\alpha$ and H$^\beta$ protons.

The complete assignments of the phenylalanine ring protons allow long-range NOEs involving these protons to be interpreted. Figure 1d depicts the methyl-aromatic region of the 2D NOESY spectrum of the perdeuterated DH domain containing protonated ϵ-^{13}C-labeled phenylalanine and ^{13}C-labeled valine and leucine with protonated methyl groups. Many NOEs are observed between the aromatic protons of the phenylalanines and the methyl protons of the valine and leucine residues. For example, the H$^\epsilon$ of F1314 shows long-range NOEs to both methyl groups of L1241, L1300, V1310, and L1377. These long-range NOEs, together with methyl–methyl, methyl–NH, NH–NH, and Phe–Phe NOEs that can be observed and easily assigned using this labeling scheme, helped us quickly obtain an initial fold of the DH domain.[19]

In summary, a strategy is described that uses phenylalanine residues that are ^{13}C-labeled only at the ϵ aromatic position to quickly assign the aromatic signals of phenylalanine in proteins. Combining this labeling method with other protein-labeling schemes that introduce protons in an otherwise perdeuterated protein can dramatically shorten the time requirement for NMR structure determination of proteins by allowing the initial fold to be rapidly obtained. In addition, this method has the potential to increase the precision and accuracy of NMR-derived structures by enabling more NOEs to be assigned unambiguously and to facilitate the structure determination of larger proteins by NMR.

Supporting Information Available: A listing of reaction conditions for the synthesis of the ϵ-^{13}C-labeled phenylalanine is available (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA983897X