Segmental Isotope Labeling for Protein NMR Using Peptide Splicing

Toshio Yamazaki,*,† Takanori Otomo,† Natsuko Oda,† Yoshimasa Kyogoku,† Koichi Uegaki,‡ Nobutoshi Ito,§ Yoshizumi Ishino,§ and Haruki Nakamura§

Institute for Protein Research, Osaka University 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan Biomolecular Engineering Research Institute 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan Osaka National Research Institute AIST 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan

Received March 9, 1998

The stable isotope labeling of peptide segments in a protein sample is achieved by means of peptide splicing to observe NMR signals of a part of larger protein molecules. The developments of the isotope-aided NMR spectroscopy allowed determination of the three-dimensional structures of proteins of up to 25 kDa,[1] but the thousands of NMR signals from a protein of this size or larger cause high degeneracy of their chemical shifts, bringing about ambiguity as to interpretation of the structural information. The technique of selective isotope labeling can reduce the number of signals. Currently, techniques for amino acid specific labeling,[2] its dual use[3] to identify a single residue, and incorporation of a labeled amino acid at a single position[4] are available. These selective labeling techniques give unambiguous information but are impractical for obtaining the resonance assignments of an entire molecule and for determining the structure of a protein.

An ideal way to approach solving the structure of a large protein is segmental labeling along the peptide chain. For example, labeling on the N-terminal side of a specific amino acid residue will reduce the number of signals in the spectra to a manageable number. The standard NMR techniques for backbone assignment of a uniformly labeled sample can work on it, because all signals come from the continuous peptide chain. The existence of the other part can be completely neglected. Structural information can be collected by means of the standard isotope-edited multidimensional NMR for the labeled part and by means of the isotope-filtered experiments for the contacting area between the labeled and unlabeled segments. Segmental labeling is the most efficient method for dividing a large target into parts of manageable size.

Here, we demonstrate a new method for the ligation of labeled and unlabeled peptide fragments using a peptide splicing element, intein (Protozyme).[5−8] Inteins are insertion sequences which are cleaved off after translation. Of particular interest is that the preceding and succeeding fragments are ligated, leaving a continuous peptide sequence (called an extein) without the intein.

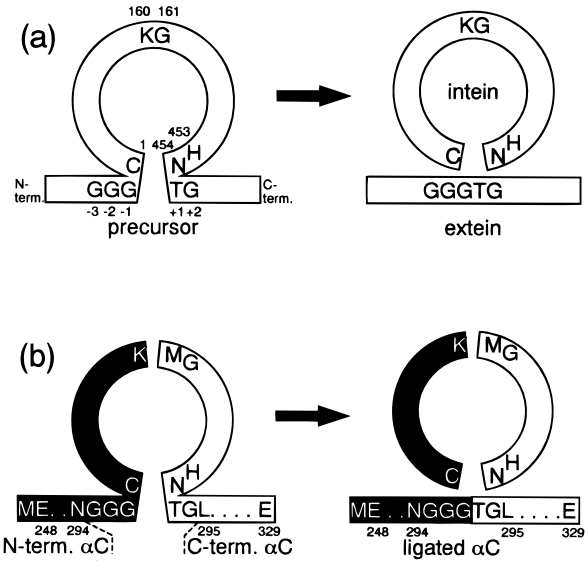

Figure 1. (a) A schematic representation of the reaction of peptide splicing. The amino acid sequence of PI-PfuI around the splicing points is given. The amino acid sequence is numbered from the N-terminus of the intein. The sequence of the extein is numbered with + and − signs. Amino acid residues, C1, H453, N454, and T(+1) are crucial for the reaction. The reason for the importance of the glycine residues is not known. (b) A schematic representation of segmental labeling. The N-terminal fragment (filled) is composed of the N-terminal part of αC (M + E248−N294), three residues of the extein (G(−3)−G(−1)), and the N-terminal part of PI-PfuI (C1−K160). The C-terminal fragment (open) is composed of the C-terminal part of PI-PfuI (M + G161−N454), two residues of the extein (T(+1) and G(+2)), and the C-terminal part of αC (L295−E329). They were expressed in *Escherichia coli* transformed with separate expression plasmids. For the labeling with ^{15}N of the N-terminal half, the N-terminal fragment was produced in the ^{15}N-labeled medium and the other fragment in the labeled medium for the labeling of the other. After partial purification of the fragments, they were mixed, refolded, and then heated for the ligation reaction. The product was purified with the standard procedure for αC.

Since only one amino acid residue in the extein part is involved in the splicing reaction, any C-terminal peptide fragment beginning with a cysteine, serine, or threonine is expected to be ligated with any N-terminal peptide fragment (Figure 1a). For different types of labeling of the N- and C-terminal extein, the precursor is cut at the middle of the intein sequence and thus divided into separate fragments. The fragments are separately produced in culture media with different isotopes. Successively, they are mixed in vitro and refolded, which leads to activation of the ligation of the target peptide fragments (Figure 1b).

The N-terminal intein (PI-PfuI) of the two in the ribonucleotide reductase[9] from *Pyrococcus furiosus* was used to generate a splicing function in this study. PI-PfuI was 454 amino acids long and was fragmented at K160-G161 (counting from the N-terminal cysteine of the intein) (Figure 1b). This position was the most sensitive to proteolysis by Lys-C protease and trypsin. We reasoned that such a flexible loop would not be important structurally or functionally. The fragmented and refolded intein retained its splicing activity.

As the target protein for segmental isotope labeling and NMR observation, the C-terminal domain of the RNA polymerase α subunit (αC)[10] is used, which has a single structural domain

* Author to whom correspondence should be addressed.
† Institute for Protein Research.
‡ Osaka National Research Institute.
§ Biomolecular Engineering Research Institute.

(1) Clore, G. M.; Gronenborn, A. M. *Methods Enzymol.* **1994**, *239*, 349−363.
(2) Muchmore, D. C.; McIntosh, L. P.; Russell, C. B.; Anderson, D. E.; Dahlquist, F. W. *Methods Enzymol.* **1989**, *177*, 44−73.
(3) Kainosho, M.; Tsuji, T. *Biochemistry* **1982**, *21*, 6273−6279.
(4) Ellman, J. A.; Volkman, B. F.; Mendel, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1992**, *114*, 7959−7961.
(5) Hirata, R.; Ohsumi, Y.; Nakano, A.; Kawasaki, H.; Suzuki, K.; Anraku, Y. *J. Biol. Chem.* **1990**, *265*, 6726−6733.
(6) Kane, P. M.; Yamashiro, C. T.; Wolczyk, D. F.; Neff, N.; Goebl, M.; Stevens, T. H. *Science* **1990**, *250*, 651−657.
(7) Xu, M.; Southworth, M. W.; Mersha, F. B.; Hornstra, L. J.; Perler, F. B. *Cell* **1993**, *75*, 1371−1377.
(8) Southworth, M. W.; Adam, E.; Panne, D.; Byer, R.; Kautz, R.; Perler, F. B. *EMBO J.* **1998**, *17*, 918−926.
(9) Riera, J.; Robb, F. T.; Weiss, R.; Fontecave, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 475−478.
(10) Jeon, Y. H.; Negishi, T.; Shirakawa, M.; Yamazaki, T.; Fujita, N.; Ishihama, A.; Kyogoku, Y. *Science* **1995**, *270*, 1495−1497.

S0002-7863(98)00776-8 CCC: $15.00
Published on Web 05/22/1998

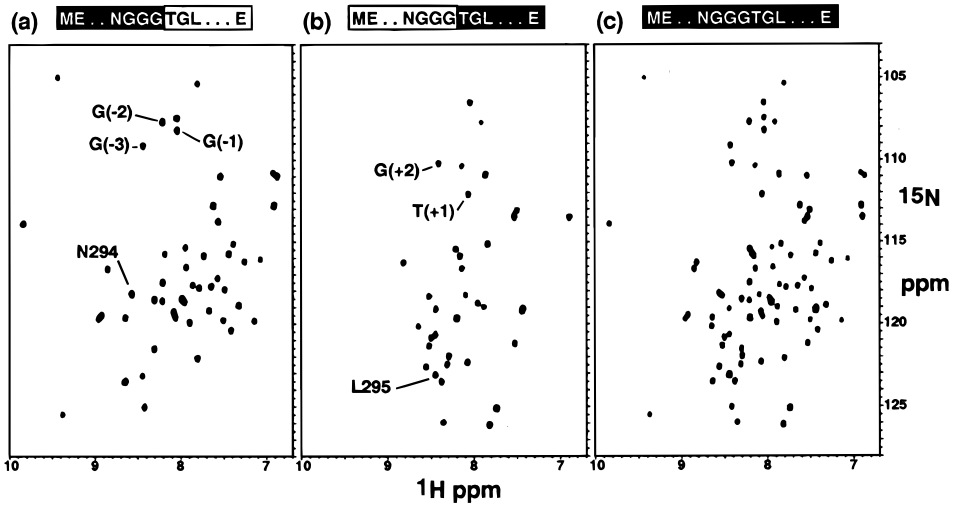

Figure 2. 2D $^{15}N-^{1}H$ HSQC spectra of the (a) N-terminal segmentally labeled and (b) C-terminal segmentally labeled αC. (c) Reference spectrum of the uniformly ^{15}N labeled αC produced from the continuous gene including the GGGTG insertion. The assignments for peaks around the joint were indicated. They were connected by regular peptide bonds, because their signals were traced on the 3D ^{15}N-edited NOESY spectrum. The N-terminal ^{15}N labeled sample only showed peaks for the N-terminal 51 amino acid residues (E248−N294 + GGG). The C-terminal ^{15}N labeled sample only showed peaks for the C-terminal 37 amino acid residues (TG + L295−E329). No signals at all were observed for the unlabeled part. All assignments were confirmed using the 3D NOESY spectrum.

composed of four α-helices. We designed to divide the target protein between N294 and L295 (counting from the N-terminus of the full α subunit) in the short flexible loop between helices 3 and 4 and to ligate them by connecting the N- and C-terminal PI-PfuI fragments. Several residues of the extein sequence were also included (Figure 1b). The insertion of a few extra amino acids at this position was not expected to perturb the entire structure of αC. We successfully obtained two segmentally ^{15}N-labeled protein samples, one labeled on the N-terminal segment composed of M + E248−N294 + GGG (Figure 2a) and the other labeled on the C-terminal segment composed of TG + L295−E329 (Figure 2b). The sum of these spectra completely coincided with the spectrum of the reference sample produced from the continuous gene for the total 88 amino acid residues (Figure 2c), indicating that chemically and structurally identical protein samples except for the patterns of isotope labeling were obtained through the ligation reaction. The efficiency of the ligation reaction was good. We expect that 1 L of culture for each fragment will provide enough of a sample for the triple resonance experiments. Though we had problems on chemical stability and lower yield in the M9 culture medium of the N-terminal fragment, these specific problems can be overcome easily (see the Supporting Information).

The three-dimensional structure of αC was not perturbed by the insertion of the GGGTG sequence to the short flexible loop as judged on a 3D NOESY spectrum. The chemical shift changes of the amide protons and nitrogens from those of wild-type αCwere negligible except in the case of several residues around the joint. This suggests that a flexible loop is an appropriate position for ligation. The minimal amino acid sequence to be inserted at the right and the applicability of the ligation reaction to less flexible positions will be examined in the coming experiments.

This method is expected to facilitate the application of NMR to proteins as large as 50 kDa. The NMR technique using deuterium labeling[11,12] is applicable to such proteins, but the actual applications are still difficult because of signal overlap. Segmental labeling will make it possible to assign signals unambiguously in a shorter time and to determine the structures of such large proteins to a higher precision.

Acknowledgment. We thank Dr. K. Komori for helpful discussions.

Supporting Information Available: Description of the constructs of vector plasmids, biochemical procedure, and the parameters of the spectra given in Figure 2 and comments on the yield of the current experiment and its improvement (3 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA980776O

(11) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655−11666.
(12) Shan, X.; Gardner, K. H.; Muhandiram. D. R.; Rao, N. S.; Arrowsmith, C. H.; Kay, L. E. *J. Am. Chem. Soc.* **1996**, *118*, 6570−6579.