Journal of Biomolecular NMR, **14:** 105–114, 1999.
KLUWER/ESCOM

Accelerated publication

Improved segmental isotope labeling of proteins and application to a larger protein

Takanori Otomo[a], Kenta Teruya[a], Koichi Uegaki[b], Toshio Yamazaki[a,*] & Yoshimasa Kyogoku[a]
[a]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan;* [b]*Osaka National Research Institute AIST, 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan*

Received 8 January 1999; Accepted 15 April 1999

Key words: intein, larger protein, protein splicing, segmental isotope labeling

Abstract

A new isotope labeling technique for peptide segments in a protein sample was recently established using the protein splicing element intein [Yamazaki et al. (1998) *J. Am. Chem. Soc.*, **120**, 5591–5592]. This method makes it possible to observe signals of a selected amino (N-) or carboxyl (C-) terminal region along a peptide chain. However, there is a problem with the yield of the segmentally labeled protein. In this paper, we report an increase in the yield of the protein that enables the production of sufficient amounts of segmentally $^{13}C/^{15}N$-labeled protein samples. This was achieved by improvement of the expression level of the N-terminal fragment in cells and the efficiency of refolding into the active splicing conformation. The N-terminal fragment was expressed as a fused protein with the cellulose binding domain at its N-terminus, which was expressed as an insoluble peptide in cells and the expression level was increased. Incubation with 2.5 M urea and 50% glycerol increased the efficiency of the refolding greatly, thereby raising the final yields of the ligated proteins. The feasibility of application of the method to a high-molecular-weight protein was demonstrated by the results for a maltose binding protein consisting of 370 amino acids. All four examined joints in the maltose binding protein were successfully ligated to produce segmentally labeled protein samples.

Abbreviations: E. coli, *Escherichia coli*; αC, the carboxyl-terminal domain of the α subunit of RNA polymerase of *Escherichia coli*; MBP, maltose binding protein; GdnHCl, guanidine hydrochloride; CBD, cellulose binding domain; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; DTT, 1,4-dithiothreitol.

Introduction

The development of multinuclear, multidimensional NMR spectroscopy has made it possible to determine the tertiary structures of proteins of up to 25 kDa (Clore and Gronenborn, 1994) and the global folds for those of up to 40 kDa (Gardner and Kay, 1998). Many useful sample preparation methods for a variety of forms of isotope labeling, from uniform labeling (Redfield et al., 1989) to site-specific labeling (Kigawa et al., 1995), have been developed as well as many pulse sequences for analysis of the labeled proteins. A uniformly ^{13}C and ^{15}N labeled protein is basically used for assigning the resonances of the protein and for solving its structure. Additional ^{2}H labeling was applied for larger proteins (Yamazaki et al., 1994; Gardner and Kay, 1998). However, in the case of analysis of proteins of high molecular masses, NMR signal overlapping makes it difficult to assign resonances unambiguously. Amino acid-specific labeling (Muchmore et al., 1989) has often been employed in such a case in order to decrease the number of NMR signals. However, this type of selective labeling cannot be used to determine the connectivities between sequential residues through triple-resonance experiments. Moreover, since the same type of amino acids

*To whom correspondence should be addressed. E-mail: yamazaki@protein.osaka-u.ac.jp

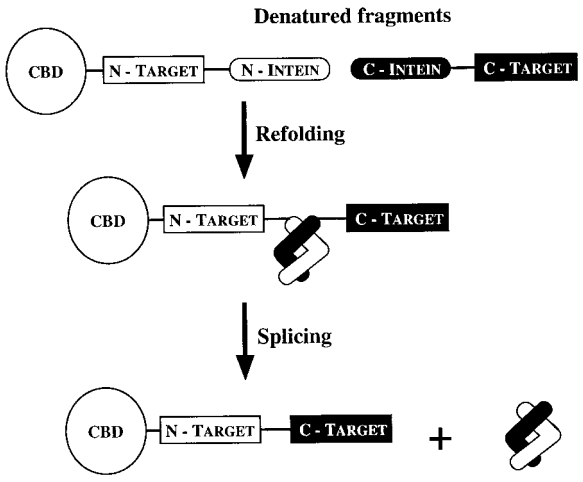

Figure 1. A schematic representation of the peptide splicing reaction for segmental labeling. The fragmented intein in the expressed N- and C-terminal fragments of a precursor is refolded and then spliced. After splicing of the refolded intein, the N- and C-terminal fragments of the target protein are joined by the peptide bond. For labeling of the N-terminal half with ^{15}N, the N-terminal fragment was produced in ^{15}N-rich medium and the C-terminal fragment in unlabeled medium.

gives similar chemical shifts, the overlap problem is not really solved.

Stable isotope labeling of peptide segments in a protein sample has recently been achieved by means of protein splicing using a protein splicing element, intein (Yamazaki et al., 1998), as described in Figure 1. This method makes it possible to observe signals of the amino or carboxyl terminal half along the peptide chain. Unambiguous assignments will be made by conducting triple-resonance experiments on the segmentally labeled protein, which gives fewer NMR signals.

It is known that several genes of prokaryotes and lower eukaryotes contain an in-frame open reading frame, which encodes an internal protein (intein) (Hirata et al., 1990; Kane et al., 1990; Perler et al., 1997). Inserted inteins are cleaved off after translation. Of particular interest is that the preceding and succeeding fragments are ligated, leaving a continuous peptide sequence (called an external protein, extein). For different types of labeling on the N- and C-terminal segments of the extein, the precursor is cut at the middle of the intein sequence, and thus the fragments are separately produced in culture media with different isotopes. Successively, they are mixed in vitro and refolded, which leads to triggering of ligation of the target peptide fragments.

All inteins share four conserved residues around the splicing junction, which are crucial for the splicing reaction: Cys/Ser at the N-terminus of inteins, a His-Asn sequence at the C-terminus of inteins, and Ser/Thr/Cys at the N-terminus of C-terminal exteins (Pietrokovski, 1994; Perler et al., 1997). Since very few amino acid residues in the extein sequence are involved in the splicing reaction, the natural extein can be replaced, in principle, with any target proteins by constructing two expression plasmids. One contains the N-terminal fragment of the target protein and the N-terminal fragment of intein and the other contains the C-terminal fragment of intein and the C-terminal fragment of the target protein. The N- and C-terminal fragments can be independent domains or fragments in a single domain, which is the case we pursue in this article.

In the previous study, the final yield of the spliced protein was not good enough, and only one protein, consisting of 88 amino acids, was examined as a target protein for segmental isotope labeling and NMR observation. In this paper, we describe improvement of the yield of a protein obtained by expression of the N-terminal fragment as a fused protein with the cellulose binding domain (CBD) and the optimized refolding conditions for the C-terminal domain of the *E. coli* RNA polymerase α subunit (αC) (Jeon et al., 1995). Then, we report the successful application of the method to a larger protein, maltose binding protein (MBP), which consists of 370 amino acids.

Materials and methods

Intein

The N-terminal intein (PI-*Pfu*I) of the two inteins in the ribonucleotide reductase from *Pyrococcus furiosus* (Riera et al., 1997) was also used to generate a splicing function as described previously (Yamazaki et al., 1998). PI-*Pfu*I is 454 amino acids long and was fragmented at Lys^{160}-Gly^{161} (counting from the N-terminal cysteine of the intein). Basically, expression vectors for the N-terminal fragments carry the N-terminal sequences of the target proteins followed by a short connector sequence and the N-terminal intein sequence (Cys^{1}-Lys^{160}). Expression vectors for the C-terminal fragments carry the C-terminal intein sequence (Gly^{161}-Asn^{454}) followed by a short connector sequence and the C-terminal sequences of the target proteins.

Construction of plasmids for the N-terminal fragment of αC-intein

A plasmid for the N-terminal fragment of αC-intein was constructed as follows. First, Asn^{294} (counting from the N-terminus of the full α subunit) in the N-terminal fragment of αC on the plasmid used in the previous work (Yamazaki et al., 1998) was deleted using QuickChange (Stratagene). Successively, the DNA sequence which contains the N-terminal gene of αC (Met + Glu^{248}-Pro^{293}) + the Gly-Gly-Gly peptide gene + the N-terminal gene of intein (Cys^{1}-Lys^{160}) was amplified by PCR from the above plasmid and inserted into the LIC (ligation independent cloning) site of pET35b(+) (Novagen), which is a vector for the cellulose binding domain (CBD) fusion system.

Construction of plasmids for the N- and C-terminal fragments of MBP-intein

Four plasmids, pCBDM30IN, pCBDM99IN, pCBDM172IN and pCBDM238IN, for expression of the N-terminal fragments of MBP-intein chimeras with MBP fragments of different lengths were constructed as follows. M30, M99, M172 and M238 in the plasmid names indicate residues 1–30, 1–99, 1–172 and 1–238 of MBP, respectively. The genes encoding the N-terminal fragments of MBP and intein were separately amplified by PCR from the pMAL-p plasmid (New England Biolabs) and PI-*Pfu*I intein gene. The PCR primers contained the overlapped sequences of MBP and intein. Each N-terminal fragment of MBP and intein was then annealed and amplified by PCR again. The 5′ PCR primer contained an *Sca*I site between pET35b(+) and the N-terminus of MBP. The PCR products were inserted into the LIC site of pET35b(+) as described by the manufacturer (Novagen). To delete the S-Tag and Factor Xa recognition sites, the plasmid was treated with *Sca*I, purified on an agarose gel and then self-ligated. Every N-terminal fragment contained a connector sequence of Gly-Gly-Gly between MBP and intein by including the corresponding nucleotide sequence in the PCR primers (see Results).

Four plasmids, pICM31, pICM101, pICM175 and pICM239, for expression of the C-terminal fragments of intein-MBP chimeras, with MBP fragments of different lengths, were constructed in almost the same way as for the plasmids of the N-terminal fragments. M31, M101, M175 and M239 in the plasmid names indicate residues 31–370, 101–370, 175–370 and 239–370 of MBP, respectively. The genes encoding the C-terminal fragments of MBP and intein were joined by PCR. The 5′ and 3′ PCR primers contained *Nde*I and *Bam*HI sites, respectively, and were treated with the restriction enzymes. Thr-Gly peptide (Thr before Gly^{175} and Gly^{101}, none before Thr^{31}) sequences were inserted between the intein and MBP by including them in the PCR primers (see Results). The PCR products were inserted downstream of the T7 promoter in the modified pGEMEX-1 (Promega).

Expression of the fragment polypeptides by E. coli

E. coli BL21(λDE3) was transformed with each plasmid and then grown at 37 °C in LB medium or M9 minimal medium containing 40 μg/ml kanamycin (for the vectors coding the N-terminal fragment) or 50 μg/ml ampicillin (for the C-terminal fragment) up to an OD_{620} of ∼0.5. Cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and harvested at 4 h after the induction. M9 minimal medium contained $^{15}NH_4Cl$ as the sole nitrogen source, ^{13}C-glucose as the sole carbon source in the case of double labeling, and several vitamins: 1 mg/l biotin, 1 mg/l folic acid, 1 mg/l d-pantothenate, 0.1 mg/l riboflavin, and 1 mg/l thiamin.

Splicing reaction and purification of ligated proteins

The experimental procedures for the splicing reaction were similar to those in the previous work (Yamazaki et al., 1998). Cells were resuspended in the lysis buffer (50 mM Tris-HCl (pH 7), 0.5 M NaCl, 1 mM EDTA), and then disrupted by sonication. The insoluble fraction was collected by centrifugation. The precipitate was dissolved in the solubilization buffer (20 mM Tris-HCl (pH 7∼8), 0.1 M NaCl, 1 mM EDTA, 1 mM DTT, 6 M GdnHCl). Solutions of the N- and C-terminal fragments were mixed and dialyzed against various refolding buffers (see Results) at 4 °C for the reassembly of two fragments and refolding. The detailed conditions for this step are given in the Results section. Then, the refolded precursor was heated to 37 °C or 70 °C for the splicing reaction (see Results). The spliced target protein, αC, was purified as described previously, and MBP was purified using an amylose column (New England Biolabs), gel filtration and MonoQ (Pharmacia). During the purification, CBD was removed by thrombin (Pharmacia) digestion. The S-Tag and Factor Xa protease cleavage sequence remained at the N-terminus of αC, and a Gly-Ser-Thr tripeptide remained at the N-terminus of MBP. The purified proteins were concentrated using Centricon (Amicon).

NMR measurements

The solution conditions for αC samples were 0.7 mM protein, 20 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 100% D_2O, pD 6.8 (uncorrected). Those for MBP samples were 0.2∼0.4 mM protein, 1.8 mM maltose, 20 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% D_2O, pH 7. All NMR experiments were performed at 303 K on a 600 MHz spectrometer (Bruker DRX) equipped with a triple-axis pulsed field gradient unit and a triple-resonance probe. All NMR data were processed using the software NMRPipe (Delaglio, 1995), and data analysis was aided by the software Pipp (Garrett et al., 1991).

Three-dimensional (3D) ^{13}C-edited NOESY data were recorded as a matrix of 128 × 16 × 512 complex points in each of t_1 (^{1}H), t_2 (^{13}C), and t_3 (^{1}H) with spectral widths of 7201 Hz, 6036 Hz and 8389 Hz in F_1, F_2 and F_3, respectively. The GARP pulse sequence (Shaka et al., 1985) was used for ^{13}C decoupling during the ^{1}H detection period. The ^{1}H and ^{13}C carriers were centered at 4.7 and 40 ppm, respectively. 3D ^{13}C-edited (F_1), ^{13}C-filtered (F_3) NOESY data were recorded as the same matrix and spectral widths as for the ^{13}C-edited NOESY experiments. During the acquisition period, no ^{13}C decoupling was employed to confirm that the signals were from ^{12}C-attached ^{1}H. 2D ^{15}N-^{1}H HSQC data for the segmentally ^{15}N-labeled MBP were recorded as a matrix of 128 × 512 complex points in each of t_1 (^{15}N) and t_2 (^{1}H) with spectral widths of 3040 Hz and 8992 Hz in F_1 and F_2. The ^{1}H and ^{15}N carriers were centered at 4.7 and 117 ppm, respectively.

Results

Improvement of the expression level of the N-terminal fragment of αC-intein

In the previous study αC was fragmented between Asn^{294} and Leu^{295}, and the N-terminal fragment (Glu^{248}-Asn^{294} + Gly-Gly-Gly + N-terminal intein) and the C-terminal fragment (C-terminal intein + Thr-Gly + Leu^{295}-Glu^{329}) were separately expressed with different isotope labelings and then ligated to produce a continuous peptide (Glu^{248}-Asn^{294} + Gly-Gly-Gly-Thr-Gly + Leu^{295}-Glu^{329}). For the ^{15}N labeling of αC, 1 L and 0.25 L of ^{15}N cultures for the N-terminal and C-terminal fragments, respectively, gave only a 0.05 mM, 0.3 ml protein solution, which was not enough to perform various NMR experiments. This low yield was due to two reasons, as mentioned previously (Yamazaki et al., 1998).

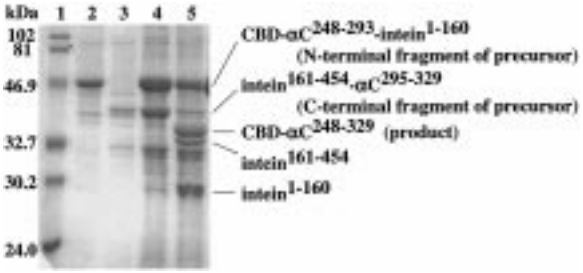

Figure 2. SDS-PAGE analysis of the splicing reaction of αC. The gel was stained with Coomassie brilliant blue. Lane 1, molecular weight markers; lane 2, expression of the N-terminal fragment of αC-intein in *E. coli*; lane 3, expression of the C-terminal fragment of αC-intein in *E. coli*; lane 4, before the splicing reaction; lane 5, after the splicing reaction. The occurrence of the splicing reaction was confirmed by the decrease in the amounts of the two precursor fragments, and the appearance of the ligated protein and excised intein fragments in lane 5.

One was the low expression level of the N-terminal fragment. The N-terminal fragment (Glu^{248}-Asn^{294} + Gly-Gly-Gly + N-terminal intein) was expressed as a soluble protein in *E. coli*, although the C-terminal fragment (C-terminal intein + Thr-Gly + Leu^{295}-Glu^{329}) was expressed as an inclusion body. The expression level of the N-terminal fragment was good in LB medium but poor in M9 minimal medium, in contrast with the case of the C-terminal fragment, of which the expression level was good in both LB and M9 minimal media. In order to improve the yield, we utilized a fusion expression system with a cellulose binding domain (Novagen) at the N-terminus of the N-terminal fragment of the αC-intein precursor. A protein fused with CBD has a tendency to form an inclusion body in cells. But it is suitable in this case because such a long unstructured soluble polypeptide is susceptible to proteases in *E. coli*, and is chemically unstable. Thus, by using the CBD fusion system, the expressed fragment peptides were obtained in the insoluble fraction and the expression level was dramatically improved to about 50 mg per liter of M9 minimal culture (Figure 2). This expression level was high enough for NMR samples.

The other reason was the chemical instability of the unfolded N-terminal fragment. Deamidation of the asparagine side chain at Asn^{294} of αC occurred for 50% of the molecules. Asn-Gly is known to be a fragile sequence (Meinwald et al., 1986), which was artificially introduced by insertion of a Gly residue after Asn^{294}. The additional step of separation of slightly different

products reduced the total yield. In this study, Asn^{294} of αC was deleted. No chemical heterogeneity was detected in the NMR spectra or the patterns of elution from columns during the purification of the spliced protein.

Improvement of the refolding and splicing reaction conditions of αC-intein

The efficiency of refolding of the two αC-intein precursor fragments must be high enough to obtain a high yield of the ligated target protein. In the previous study, the splicing precursor was refolded from a mixture of the two fragments in 6 M GdnHCl by dialysis against a buffer which did not contain a denaturant, but the efficiency was not good and substantial precipitation of the two fragments occurred. In this study, we examined several protocols to determine the optimal refolding and splicing conditions for the αC-intein precursor.

A variety of conditions for the in vitro protein splicing of a purified precursor have already been investigated (Xu et al., 1993; Southworth et al., 1998). It was reported that splicing of the precursor gave optimal efficiency at pH 6–7 and was disabled at 4–16 °C or pH values above 9. Therefore the basic parameters for the refolding were fixed according to Xu et al. (1993) and Southworth et al. (1998): dialysis time, overnight; dialysis temperature, 4 °C; pH 7; and salt concentration, 100 mM NaCl. In order to avoid precipitation due to misrefolding, a detergent, Triton X-100 (1%), or a denaturant, urea (2 or 6 M), was added to the refolding buffer (20 mM Tris-HCl (pH 7), 0.1 M NaCl, 1 mM EDTA, 1 mM DTT). The results of the protocols are summarized in Table 1. No precipitation occurred with 2 or 6 M urea. The splicing reaction was started by heating the mixture solution to 37 °C (or 70 °C), and continued for 8 h (or 2 h). The splicing reaction at 70 °C was more efficient than that at 37 °C in all cases except in the case of the buffer containing 6 M urea. The low efficiency (about 30%) of the splicing reaction at 70 °C in the buffer containing 6 M urea indicated that the fragmented intein did not fold tightly under these conditions. Consequently, the buffer containing 2 M urea and splicing at 70 °C gave the highest yields. However, we worried about chemical modification by urea at a temperature as high as 70 °C, and thus finally we chose the lower temperature of 37 °C with 2 M urea for the production of segmentally isotope labeled αC. The final yield of the segmentally N- or C-terminal ^{13}C/^{15}N-labeled αC was 0.7 mM × 0.3 ml from 500 ml of M9 minimal medium.

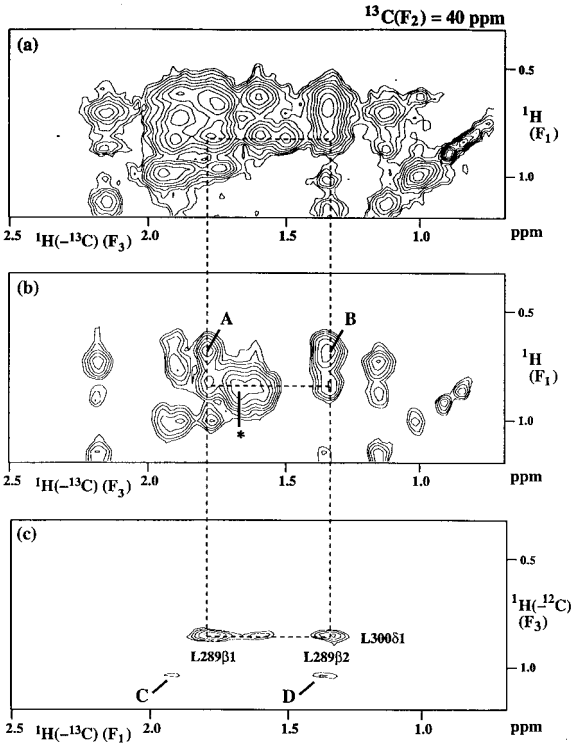

Figure 3. Comparison of slices of the same region of 3D ^{13}C-edited NOESY spectra of (a) the uniformly and (b) the N-terminal segmentally ^{13}C/^{15}N-labeled αC, and (c) a 3D ^{13}C-edited (F_1), ^{13}C-filtered (F_3) NOESY spectrum of the N-terminal segmentally ^{13}C/^{15}N-labeled αC. The horizontal axis of spectrum (c) is F_1 and the vertical axis is F_3. Peak A is the intra-residue NOE between Leu^{289}β1 and Leu^{289}δ1, and peak B is the intra-residue NOE between Leu^{289}β2 and Leu^{289}δ1. Peak C is the inter-segment NOE between Val^{264}β1 and Val^{306}δ2, and peak D is the inter-segment NOE between Val^{264}β2 and Val^{306}δ2. The peak indicated by an asterisk has not been identified. There is no peak in (c) corresponding to peaks A and B, because no intra-segment NOE including an intra-residue one could be obtained for the filtered NOESY spectrum (c). Peak D in (c) exists in (a) but not in (b).

Figure 2 shows the results of SDS-PAGE analysis of the splicing system for αC involving the buffer containing 2 M urea and splicing at 37 °C. On SDS-PAGE, two fragments (CBD-αC$^{248-293}$-intein$^{1-160}$, and intein$^{161-454}$-αC$^{295-329}$) were observed before the splicing reaction, and three new bands of the product (CBD-αC), and the N- and C-terminal fragments of intein appeared after the splicing reaction.

NMR measurement of segmentally ^{13}C/^{15}N-labeled αC

Figure 3 shows the slices of the same region of 3D ^{13}C-edited NOESY spectra of (a) the uniformly and (b) the N-terminal segmentally ^{13}C/^{15}N-labeled αC,

Table 1. Conditions for refolding and the splicing reaction for αC[a]

Refolding conditions[b]	Precipitation[c]	Splicing temperature (°C)	Splicing efficiency[d] (%)
0 M GdnHCl	+	37	80
		70	90
0 M GdnHCl, 1% TritonX-100	+	37	80
		70	90
2 M urea	−	37	70
		70	90
6 M urea	−	37	70
		70	30

[a]N- and C-terminal fragments were mixed in the solubilization buffer containing 6 M GdnHCl, and then dialyzed against the refolding buffer under the indicated conditions at 4 °C overnight. Dialyzed samples were incubated at the indicated temperatures, 37 °C and 70 °C, for 8 h and 2 h, respectively.
[b]Only the different conditions between the refolding buffers are indicated. Other conditions were fixed, i.e. 20 mM Tris-HCl, pH 7, 1 mM DTT, 1 mM EDTA and 0.1 M NaCl.
[c]A large amount of precipitation after the dialysis is indicated by +, and − indicates that no precipitation occurred.
[d]Splicing efficiency was expressed as the ratio of the precursor fragments to the products, as judged on SDS-PAGE analysis.

and (c) a 3D ^{13}C-edited (F_1), ^{13}C-filtered (F_3) NOESY spectrum of the N-terminal segmentally ^{13}C/^{15}N-labeled αC. Clearly, a reduction in the number of peaks in (b) in comparison with that in (a) was observed. The inter-segment NOEs are focused using the filtered NOESY spectrum (c), in which an example of the inter-segment NOE between unlabeled Leu^{300} δ1 and ^{13}C-labeled Leu^{289} β1 or β2 is shown. These spectra would give many unambiguous assignments of NOE signals, which is quite important for precise determination of the protein structure.

Ligation of the fragments of MBP

αC, which was our first target protein, consists of only 88 amino acids. This new segmental labeling method is useful especially for larger proteins, which give very crowded spectra due to a lot of broader signals. In order to confirm that this method is applicable to larger proteins, MBP, which has 370 amino acids, was introduced. Since MBP has an intricate chain topology, it had seemed difficult to refold the denatured separated peptide fragments to the active structure. Four ligation positions (around residue numbers 30, 100, 175 and 238; Figure 4) were examined for MBP, although in the case of αC only one ligation position was examined. The reason why we selected these positions was that they are not involved in any secondary structure and are exposed to the solvent.

The expression strategy for the N- and C-terminal fragments of MBP was almost the same as that for αC. The N-terminal fragments were fused with CBD,

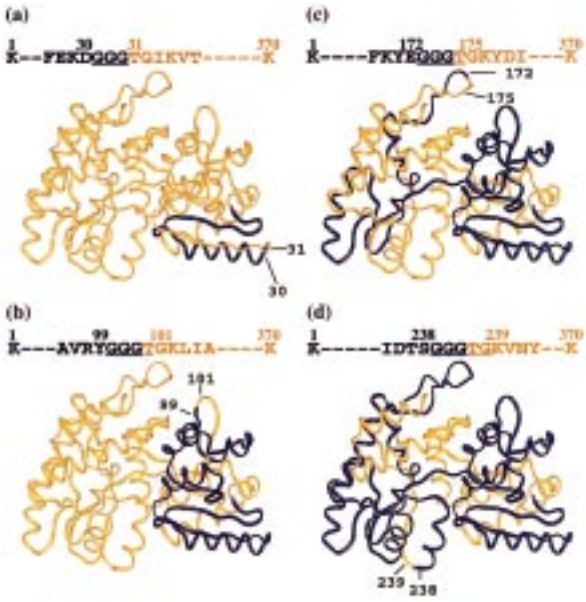

Figure 4. Joint positions in the amino acid sequence and the structure of MBP. Blue indicates the N-terminal segments and orange the C-terminal segments of MBP. The inserted amino acids are indicated by underlines. Gly-Gly-Gly is the last three amino acid sequence of the N-terminal segments and Thr-Gly are the first two amino acid residues of the C-terminal segments.

and the C-terminal fragments were expressed directly. The expression levels of all the N-terminal fragments were high (∼50 mg per liter M9 minimal culture), but the expression levels for the C-terminal fragments became lower as the peptide length increased (∼10 mg

for the product from pICM31 and 30 mg for that from pICM239 per liter M9 minimal culture). All the expressed polypeptides were insoluble, and buffer containing 6 M GdnHCl was, therefore, used to solubilize them. First, we applied the protocol optimized for αC to all the four joints of MBP, i.e. the refolding buffer contained 2 M urea and the splicing reaction was performed at 37 °C. The result indicated that the splicing reactions failed for all the four ligation positions (data not shown).

Improvement of the refolding and splicing conditions for MBP and the production of segmentally isotope labeled MBP

Since the splicing reactions were found to fail for all the four ligation positions, the refolding and splicing conditions for MBP were examined again as shown in Table 2, including two new refolding protocols. The basic parameters for the refolding, i.e. the dialysis temperature (4 °C), and the pH (7) and salt concentration (100 mM NaCl) of the dialysis buffer, were fixed as in the case of αC. The examined splicing temperatures were also 37 °C and 70 °C. These experiments were performed on the two joints, Tyr^{99}-Gly^{101} and Ser^{238}-Lys^{239}. The results were the same for both joints. As summarized in Table 2, both refolding without denaturant and splicing with urea were inefficient.

Then, we introduced a two-step dialysis including an incubation during the refolding with urea (protocol D) and glycerol (protocol E) to improve the refolding efficiency. Protocol D was as follows. The unfolded fragment mixture with 6 M GdnHCl was dialyzed against the buffer containing 2.5 M urea for 24 h, and then dialyzed against the buffer containing no urea overnight to remove the urea. Protocol E was as follows. The same volume of glycerol was added to the unfolded fragment mixture containing 6 M GdnHCl (3 M GdnHCl and 50% glycerol at this step), followed by dialysis against the buffer containing 2.5 M urea and 50% glycerol overnight and then again against the buffer containing 50% glycerol but no urea to eliminate the urea. Since we knew that the splicing reaction did not proceed with urea, the final dialysis was performed to eliminate the denaturant in both protocols D and E. The refolding efficiency was improved in both cases, and protocol E (involving 50% glycerol) showed less precipitation than protocol D (incubation with urea).

In contrast with protocol A (buffer without a denaturant), protocol D showed very good efficiency. This

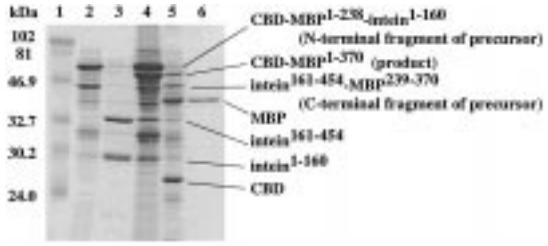

Figure 5. Example of SDS-PAGE analysis of the reaction of MBP (the joint is located at Ser^{238}-Lys^{239}). The gel was stained with Coomassie brilliant blue. Lane 1, molecular weight markers; lane 2, before the splicing reaction; lane 3, the supernatant after the splicing reaction; lane 4, the precipitation after the splicing reaction; lane 5, after thrombin digestion; lane 6, the purified segmentally labeled MBP. Bands of intein fragments were observed for the supernatant after the splicing reaction at 70 °C (lane 3), and the product, CBD-MBP, was in the precipitate (lane 4).

indicates that the incubation with 2.5 M urea for 24 h was very effective for the refolding. It is well known that glycerol is often effective for raising the solubility and stability of proteins, and also effective for refolding (Rudolph and Lilie, 1996; Rariy and Klibanov, 1997). A high concentration of glycerol gave a very good result as to refolding from two fragments in this case. The splicing reaction at 37 °C did not give a good efficiency (about 30%), but at 70 °C it was good (about 80%) with either refolding protocol D or E. We finally selected protocol D for the labeling of the segments of Lys^{1}-Tyr^{99} and Gly^{101}-Lys^{370}, and protocol E for the segments of Lys^{1}-Ser^{238} and Lys^{239}-Lys^{370}. The ligation at the other joints was confirmed using protocol A. The efficiencies of the refolding and splicing reactions for these two joints were almost the same as those for the joints Tyr^{99}-Gly^{101} and Ser^{238}-Lys^{239}, although the other protocols B–E were not examined for these two joints. Protocols D and E are expected to be applicable to raise the efficiencies for the two joints.

The splicing reactions were performed at 70 °C for 2 h. In protocol D, the ligated product (CBD-MBP) was precipitated after the splicing reaction at such a high temperature. In protocol E, the product was in the supernatant after the splicing reaction due to the high concentration of glycerol, but after dialysis for the removal of glycerol, the product was precipitated. Therefore an additional refolding step for the ligation product was necessary in both cases. The precipitated CBD-MBP was dissolved with a buffer (20 mM Tris-HCl (pH 7), 0.1 M NaCl, 1 mM EDTA, 1 mM DTT, 6 M GdnHCl), and dialyzed against the same buffer without GdnHCl for the refolding. An example of SDS-PAGE analysis is shown in Figure 5. As

Table 2. Conditions for refolding and the splicing reaction for MBP[a]

Protocol	Refolding conditions[b]	Precipitation[c]	Splicing temperature (°C)	Splicing efficiency[d] (%)
A	0 M GdnHCl	++	37	30
			70	80
B	2 M urea	−	37	0
			70	10
C	6 M urea	−	37	0
			70	0
D	(step 1) 2.5 M urea	+	37	30
	(step 2) 0 M urea		70	80
E	(step 1) add glycerol to 50%	−	37	30
	(step 2) 2.5 M urea, 50% glycerol		70	80
	(step 3) 0 M urea, 50% glycerol			

[a]N- and C-terminal fragments were mixed in the solubilization buffer containing 6 M GdnHCl, and then dialyzed against the refolding buffer under the indicated conditions at 4 °C overnight. Dialyzed samples were incubated at the indicated temperatures, 37 °C and 70 °C, for 8 h and 2 h, respectively.
[b](A−C) Only the different conditions between the refolding buffers are indicated. Other conditions were fixed, i.e. 20 mM Tris-HCl, pH 7, 1 mM DTT, 1 mM EDTA and 0.1 M NaCl. (D) The first step was dialysis against a buffer containing 2.5 M urea for 24 h and the second dialysis was against the buffer containing no denaturant overnight. (E) The first step was the addition of glycerol to the fragment mixture with 6 M GdnHCl to a final concentration of 50% (v/v) glycerol. The second step comprised dialysis against the buffer containing 2.5 M urea and 50% glycerol overnight, and the third step dialysis against the buffer containing no denaturant overnight.
[c]Large and small amounts of precipitation after the dialysis are indicated by ++, and +, respectively. − indicates that no precipitation occurred.
[d]Splicing efficiency was expressed as the ratio of the precursor fragments to the products, as judged on SDS-PAGE analysis.

in the case of αC, two precursor fragments were observed before the splicing reaction, and three bands corresponding to the intein fragments and the ligated product (CBD-MBP) appeared after the splicing reaction. The two patterns of the joints Tyr^{99}-Gly^{101} and Ser^{238}-Lys^{239} were used for the production of ^{15}N-segmentally labeled MBP. Thus, a total of four patterns of labeling for MBP were made; Lys^{1}-Tyr^{99} + Gly-Gly-Gly, Thr + Gly^{101}-Lys^{370}, Lys^{1}-Ser^{238} + Gly-Gly-Gly, and Thr-Gly + Lys^{239}-Lys^{370}. The final yields of ^{15}N-segmentally labeled MBP were almost the same for all samples. From 800 ml of M9 minimal medium culture for each precursor fragment, a 0.4 mM × 0.3 ml of protein solution was obtained.

NMR measurement of segmentally labeled MBP

2D ^{15}N-^{1}H HSQC spectra of the ^{15}N-segmentally labeled MBP/maltose complex are shown in Figure 6. It is clear that the numbers of peaks in the spectra shown in (b), (c), (d) and (e) were dramatically reduced. Because the reference sample of uniformly labeled MBP had the native sequence and the spliced proteins contained insertions of several amino acids, the chemical shifts of the residues around the joints were slightly shifted. However, the positions of the isolated peaks in the spectra of the uniformly and segmentally labeled samples showed almost perfect agreement. This means that the MBPs produced by this method have the same structure as the wild-type MBP. This really confirms that the segmental-labeling method was successful for a larger protein, MBP.

Discussion

First, we successfully improved the yield of the final product as to the expression level of the N-terminal fragment of the αC-intein precursor in *E. coli* and the refolding and splicing reaction efficiencies. The yields were sufficient for 3D NMR experiments for the structure determination of proteins. Second, we applied the method to a larger protein, MBP. The success of ligation of MBP indicates that the method is applicable to larger proteins.

In the case of MBP, the four selected joint positions do not have a secondary structure and thus seem flexible. The ligation of a position in the secondary structure, like the α-helix or β-sheet of MBP, has not been examined, because we had failed to ligate two fragments of αC on helices (data not shown). At present, we have concluded that the splicing of intein requires structural flexibility of the extein at a

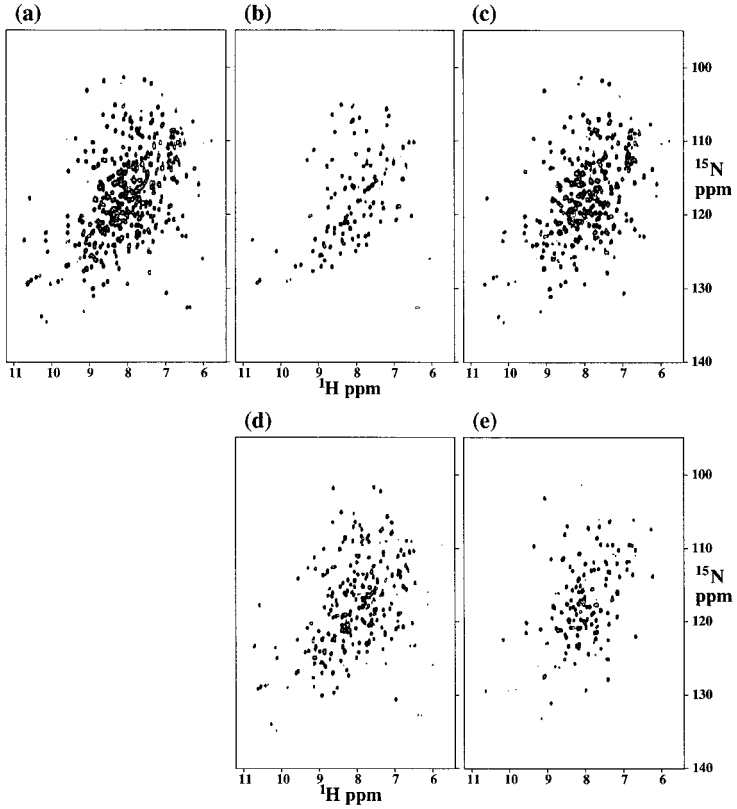

Figure 6. 2D ^{15}N-^{1}H HSQC spectra of (a) the uniformly ^{15}N-labeled wild-type MBP/maltose complex, and those of MBP segmentally labeled with ^{15}N between (b) Lys^{1}-Tyr^{99}; (c) Gly^{101}-Lys^{370}; (d) Lys^{1}-Ser^{238}; and (e) Lys^{239}-Lys^{370}.

joint. Several amino acids, Gly-Gly-Gly-Thr-Gly are inserted at a joint to increase the flexibility. Since the structure of the original extein (ribonucleotide reductase) of PI-*Pfu*I has not been solved, the roles of the several Gly residues are not yet known. The Gly-Gly-Gly-Thr-Gly sequence is a part of the native extein, ribonucleotide reductase of *Pyrococcus furiosus*, but there is no sequence homology at the joints among a variety of exteins except Ser/Thr/Cys at the N-terminus of C-terminal exteins. Our preliminary data indicate that only the Gly-Thr sequence (inserted at around residue number 175 in MBP, which is located in the most flexible region among the four examined positions) is enough for the splicing if the position has structural flexibility (data not shown). For proteins with unknown structures, it is very hard to demonstrate the flexibility of the target position quantitatively, but protease digestion experiments and secondary structure predictions from the primary sequence will give useful information about applicable ligation positions. If mutations by inserting Gly-Gly-Gly-Thr-Gly at the candidate positions do not affect its function and structure stability, they must be appropriate positions for ligation.

A crystal structure study on a GyrA intein precursor revealed the structural basis of the splicing mechanism (Klabunde et al., 1998). According to their proposed mechanism, the peptide bond between the N-terminal residue of the intein and the C-terminal residue of the N-terminal extein showed an energetically unfavorable *cis* conformation. This orientation of the peptide bond allows a nucleophilic attack by Cys (the N-terminal residue of the intein) to the C-terminal residue of the N-terminal extein, which is the first step of the splicing reaction. We consider that not every amino acid can be replaced for the C-terminal residue of the N-terminal extein (Gly for PI-*Pfu*I), because this residue might be involved in the intein structure, which forces the peptide bond to adopt the energetically disadvantageous *cis* conformation.

The refolding and splicing reaction conditions are as important as the design of the joints, as described above. There are some differences in the optimal conditions for the two target proteins, αC and MBP. The

most remarkable one is that the use of urea during the refolding and splicing reaction gave a very good result for αC, but not for MBP. The temperature during the reaction is also important, because αC was spliced at 37 °C as well as 70 °C, but MBP required the temperature of 70 °C for the splicing reaction. These facts suggest that the optimal conditions depend on the character of the target protein. The temperature of 70 °C for the splicing reaction might be severe for some of the target proteins. In the case of our target proteins, αC was soluble and MBP was precipitated at 70 °C. Ligated MBP was recovered from the precipitate by refolding as mentioned above. If this temperature causes an irreversible damage on the target protein, a longer reaction time at a lower temperature must be tried.

The NMR technique using deuterium labeling is applicable to proteins as large as 50 kDa, but actual application is still difficult because of signal overlapping. In particular, the method using a methyl protonated and highly deuterated protein (Gardner and Kay, 1998) and TROSY experiments with a deuterated sample (Pervushin et al., 1997; Salzmann et al., 1998) are powerful for larger proteins. These methods solved the problem of the high relaxation rate but not that of signal overlapping. The combination of them with our method will be the most powerful technique for structure determination of larger proteins.

It is expected that a central segment can be selectively labeled by ligation of three peptide segments using two inteins. If this technique can be established, proteins of several hundred residues can be divided into segments of appropriate sizes, i.e. about 100 amino acid residues. Such a method will be examined in future studies.

Acknowledgements

We wish to thank Dr. Ishino (Biomolecular Engineering Institute) for providing the PI-*Pfu*I intein gene, and Ms. Kondo for the help in the DNA sequencing of plasmids. T.O. is a Research Fellow of the Japan Society for the Promotion of Science. This work was supported by a Grant-in-Aid for a Specially Promoted Area from the Ministry of Education, Science and Culture, Japan.

References

Clore, G.M. and Gronenborn, A.M. (1994) *Methods Enzymol.*, **239**, 349–363.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Garrett, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.

Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, H. and Anraku, Y. (1990) *J. Biol. Chem.*, **265**, 6726–6733.

Jeon, Y., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A. and Kyogoku, Y. (1995) *Science*, **270**, 1495–1497.

Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. and Stevens, T.H. (1990) *Science*, **250**, 651–657.

Kigawa, T., Muto, Y. and Yokoyama, S. (1995) *J. Biomol. NMR*, **6**, 129–134.

Klabunde, T., Sharma, S., Telenti, A., Jacobs Jr., W.R. and Sacchettini, J.C. (1998) *Nat. Struct. Biol.*, **5**, 31–36.

Meinwald, Y.C., Stimson, E.R. and Scheraga, A. (1986) *Int. J. Pept. Protein Res.*, **28**, 79–84.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) *Methods Enzymol.*, **177**, 44–73.

Perler, F.B., Olsen, G.J. and Adam, E. (1997) *Nucleic Acids Res.*, **25**, 1087–1093.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.

Pietrokovski, S. (1994) *Protein Sci.*, **3**, 2340–2350.

Rariy, R.V. and Klibanov, A.M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 13520–13523.

Redfield, A.G., McIntosh, L.P. and Dahlquist, F.W. (1989) *Arch. Biol. Med. Exp.*, **2**, 129–137.

Riera, J., Robb, F.T., Weiss, R. and Fontecave, M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 475–478.

Rudolph, R. and Lilie, H. (1996) *FASEB J.*, **10**, 49–56.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. and Wüthrich, K. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 13585–13590.

Shaka, A., Barker, P.B. and Freeman, R. (1985) *J. Magn. Reson.*, **64**, 547–552.

Southworth, M.W., Adam, E., Panne, D., Byer, R., Kautz, R. and Perler, F.B. (1998) *EMBO J.*, **17**, 918–926.

Xu, M., Southworth, M.W., Mersha, F.B., Hornstra, L.J. and Perler, F.B. (1993) *Cell*, **75**, 1371–1377.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.

Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. and Nakamura, H. (1998) *J. Am. Chem. Soc.*, **120**, 5591–5592.