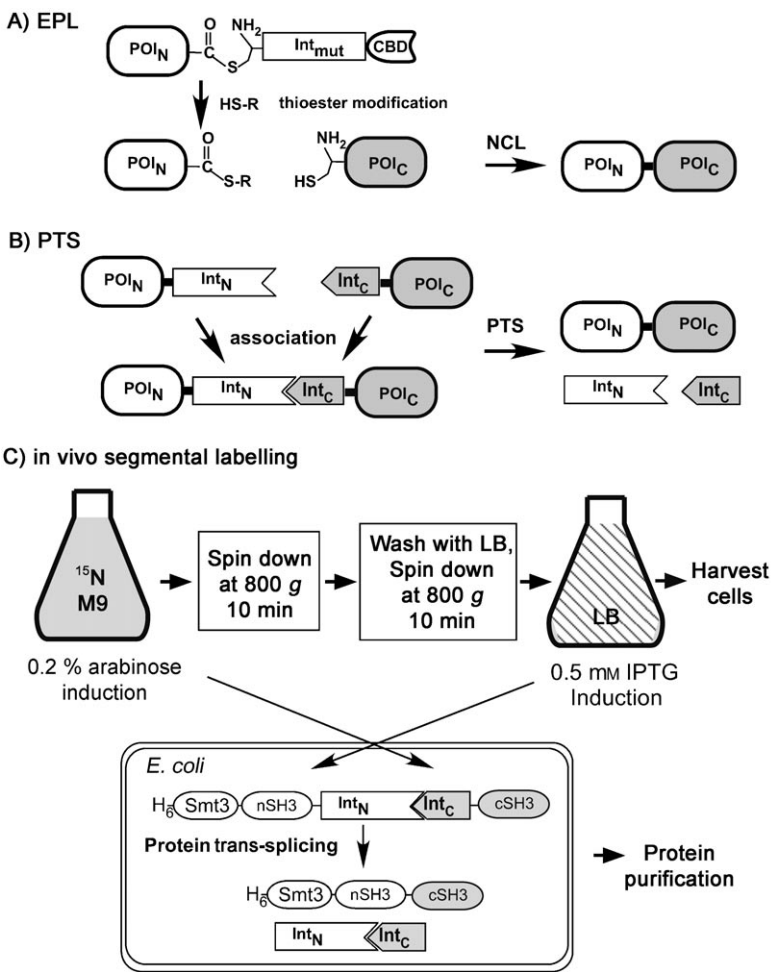
DOI: 10.1002/cbic.200800604

Segmental Isotopic Labelling of a Multidomain Protein by Protein Ligation by Protein Trans-Splicing

Mikko Muona, A. Sesilja Aranko, and Hideo Iwai*[a]

Segmental isotopic labelling is a powerful method for the incorporation of stable isotopes into particular regions within proteins for NMR detection, thereby reducing the complexity of NMR spectra and offering the potential to perform sequential assignments.[1–3] Here we have demonstrated segmental isotopic labelling of a domain in a multidomain protein both in vivo and in vitro through protein ligation by protein trans-splicing. This robust protein trans-splicing approach could open possibilities for studying particular domains in intact proteins without dissection into smaller globular domains.

Recent advances in optimization of transverse-relaxation in NMR spectroscopy have opened avenues for study of larger molecules (close to 1 MDa).[4] However, sequential resonance assignments in large proteins remain time-consuming and challenging because of the increased number of signals and signal overlapping.[4] Segmental isotopic labelling is one promising approach among numerous isotope-labelling techniques,[5] because, unlike in the case of selective amino acid labelling, segmentally isotope-labelled samples can be directly analysed by triple-resonance NMR techniques developed for sequential resonance assignments. Segmentally isotope-labelled proteins have been prepared either by expressed protein ligation (EPL), which makes use of native chemical ligation (NCL),[2,6] or by protein trans-splicing (PTS), through the use of artificially split protein splicing domains (inteins).[1,7] EPL requires the preparation of an α-thioester group from a thiol reagent and an N-terminal cysteine residue by proteolysis in vitro (Scheme 1 A), which demands considerable preparation efforts,[2,8] although an easier approach has recently been proposed.[6] In protein splicing, an intein catalyses protein ligation of two polypeptide fragments fused to the N- and C-terminal ends of an intein.[9] Protein splicing could take place in trans, when an intein is split into two fragments (Scheme 1 B).[9,10] Segmental isotopic labelling through protein trans-splicing with artificially split inteins requires no additional thiol reagent nor cofactor, but denaturation and renaturation steps are necessary before protein-splicing activity can be restored.[1,7,11]



Scheme 1. Protein ligation by A) expressed protein ligation and B) protein trans-splicing. C) Outline of the in vivo procedure for segmental isotopic labelling used in this article.

Unlike artificially split inteins, naturally split inteins do not require any denaturation and renaturation steps for protein splicing.[10] Therefore, these have been suggested as potentially useful for segmental isotopic labelling of multidomain proteins.[12] Protein trans-splicing with naturally split inteins has advantages over EPL because protein ligation can be performed not only in vitro but also in vivo, making it possible to achieve segmental isotopic labelling in vivo.[13] Despite its many potential applications, it has never been used for segmental isotopic labelling of multidomain proteins except for a fusion tag for enhancing protein solubility. This is because the protein-splicing activity of the split inteins could be negatively affected even when naturally split inteins were fused with the

[a] M. Muona, A. S. Aranko, Dr. H. Iwai
Research Program in Structural Biology and Biophysics
Institute of Biotechnology, University of Helsinki
P.O. Box 65, Helsinki 00014 (Finland)
Fax: (+358) 9-191-59541
E-mail: hideo.iwai@helsinki.fi

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

target domains.[14] Applications have therefore been limited to "well-behaved" model proteins such as GB1.[13,14] Moreover, the selectivity of in vivo segmental isotopic labelling for segmentally labelled C-terminal domains was low, diminishing the usefulness of segmental isotopic labelling.[13]

In this article we describe segmental isotopic labelling of a multidomain protein containing a multiple copy of one of the most abundant domains of the Src homology 3 (SH3) domain by protein ligation with an engineered naturally split intein from *Nostoc punctiforme* (*Npu*).[14] The selectivity of isotopic labelling by the in vivo approach is also enhanced by more than fivefold.

The SH3 domain is one of the most abundant domains, found in many organisms and usually a part of larger proteins as a module mediating protein–protein interactions.[15] Segmental isotopic labelling of proteins containing multiple copies of the SH3 domain would be of particular importance for study of their functional and structural roles in their full-length contexts by NMR spectroscopy. The in vivo method with naturally split inteins would be a logical and important step towards segmental isotopic labelling of such a multidomain protein, because it is a simple and versatile procedure requiring no preparation step prior to protein ligation of isotope-labelled fragments, circumventing several purifications of the individual protein fragments.[13] However, naturally split DnaE inteins either from *Synechocystis* sp. strain PCC6803 (*Ssp*DnaE) or from *Nostoc punctiforme* (*Npu*DnaE) intein were unable to ligate two SH3 domains efficiently either in vivo or in vitro because of the dominant cleavage reaction (data not shown). We therefore decided to use an engineered split *Npu*DnaE intein in which the split site was moved towards the C terminus by 20 residues and which still retained highly efficient protein trans-splicing activity in vivo.[16] Additionally, the N-terminally His-tagged ubiquitin-like protein Smt3 was attached as a fusion protein tag for convenient protein purification.

The N and C fragments of the split intein were fused to the N-terminal SH3 domain (nSH3) and the C-terminal domain (cSH3), respectively, of c-CRKII adaptor domain.[17] The N and C precursor proteins containing the SH3 domains were cloned into the two compatible plasmids under the control of two different promoters: that is, T7 and arabinose promoters.[13] The two precursor proteins were induced by addition of L-arabinose (0.2%), followed by induction with isopropyl-β-D-thiogalactopyranoside (IPTG; 0.5 mM) for 4 h. After the two sequential inductions, the two precursor proteins dominantly produced the ligated product containing the two SH3 domains resulting from protein trans-splicing (Figure 1). The protein ligation of the two SH3 domains was confirmed by ESI mass spectrometry after removal of the fusion tag of Smt3 protein (the molecular mass of nSH3-cSH3 was found to be 19945.0 Da, in accordance with the expected value). About 21 mg of the ligated nSH3-cSH3 was obtained from two litres of ^{15}N-labelled M9 medium. Protein ligation performed in vivo can thus indeed produce sufficient amounts of the ligated multidomain protein for NMR studies, thanks to the high efficiency of the engineered split *Npu*DnaE intein.

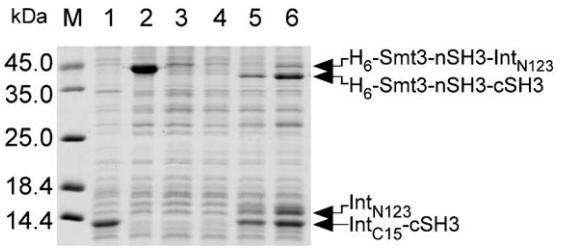


Figure 1. SDS-PAGE analysis of in vivo protein ligation of the two SH3 domains with the engineered split *Npu*DnaE intein. Lane 1: induction only with L-arabinose (0.2%). Lane 2: induction with IPTG (0.5 mM). Lane 3: before induction. Lane 4: after 0.5 h of L-arabinose induction. Lane 5: 1 h after the second induction with IPTG (0.5 mM). Lane 6: after 2.5 h of IPTG induction.

We next applied in vivo segmental isotopic labelling, in which the C-terminal domain was first induced by L-arabinose in ^{15}N-labelled M9 medium, followed by the replacement of the labelled medium with Luria–Bertani medium and the inducer for the expression of the N-terminal precursor (Scheme 1C). We used LB medium as both unlabelled medium and washing medium in order to improve the selectivity of segmental labelling.[13] The biggest disadvantage of in vivo labelling is isotopic scrambling due to recycling and conversion of labelled amino acids.

In the case of in vivo segmental isotopic labelling, the domain induced later was particularly contaminated with the isotope initially incorporated into the domain primarily expressed in the isotope-enriched medium. It has previously been reported that the residual signals of the second induced domain contained as much as 25% of the original signals.[13] This could be sufficient to disturb certain NMR experiments such as isotope-filtering experiments. Figure 2 demonstrates segmental isotopic labelling of nSH3-cSH3 through the use of the engineered split *Npu*DnaE intein with the new in vivo procedure. In the new procedure, LB medium containing the second inducer (IPTG) was used both for the production of the unlabelled domain and for a short wash in order to suppress isotopic scrambling (Scheme 1C). M9 medium (1.5 L) for the primary expression and LB medium (1.8 L) for the secondary expression were used for in vivo segmental isotopic labelling. After removal of the fusion tag by limited proteolytic digestion and ion metal chelating chromatography and subsequent ion-exchange chromatography, the segmental isotope-labelled protein (8.2 mg) was successfully purified. Figure 2A shows the superposition of the two [^{15}N,^{1}H]-TROSY spectra for comparison, demonstrating that the spectrum of the segmentally isotope-labelled nSH3-[^{15}N]-cSH3 lacks the signals of nSH3 observed in the spectrum of the ligated fully labelled [^{15}N]-nSH3-[^{15}N]-cSH3. Unlike in the previous report, a high degree of selective labelling can be recognized when the slices of the two spectra taken at the resonance of residue 289 in c-CRK II adaptor protein are compared (Figure 2B). The signals from the N-terminal domain are below the noise level (<5%), although the segmentally labelled sample was prepared in the cells (Figure 2B). The selectivity of the in vivo procedure can thus be enhanced by more than fivefold, which should allow us to use the sample for most NMR experiments. Highly selective

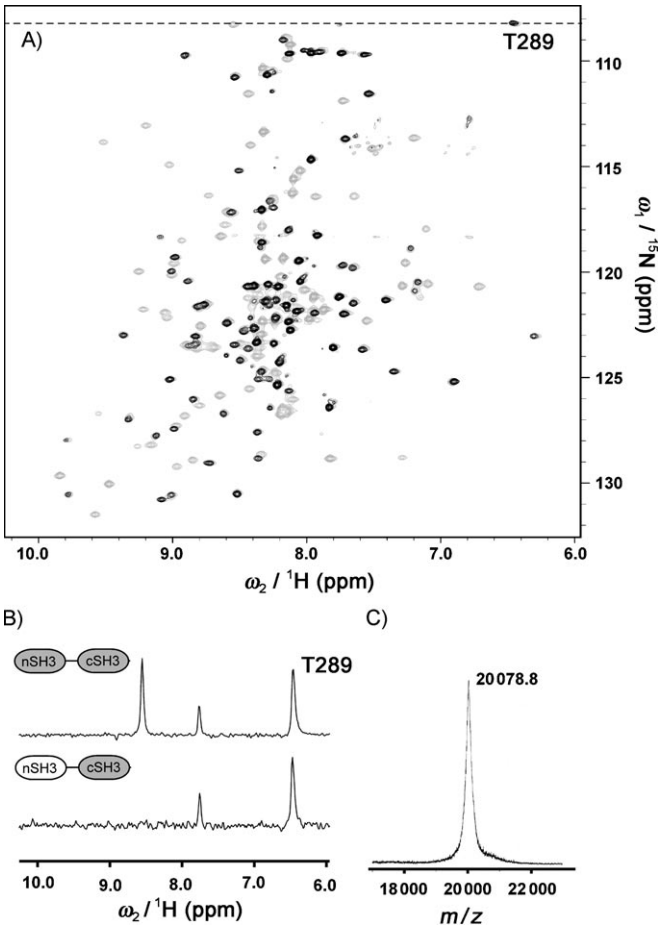


Figure 2. A) Superposition of the [^{15}N,^{1}H]-TROSY spectra of fully ^{15}N-labelled ligated nSH3-cSH3(131–304) (grey) and nSH3-[^{15}N]-cSH3(213–304) prepared by in vivo segmental labelling (black). The peak from residue 289 of c-CRK II is labelled. B) Comparison of the slices taken at the positions indicated by the broken line from the spectra obtained from the fully labelled sample (upper) and from the segmentally labelled sample (lower). C) MALDI-TOF mass analysis of the segmentally isotope-labelled nSH3-[^{15}N]-cSH3(213–304). The averaged mass of the unlabelled sample is 19945.0 Da. cSH3(213–304) has the chemical formula $C_{474}H_{733}N_{133}O_{143}S_2$.

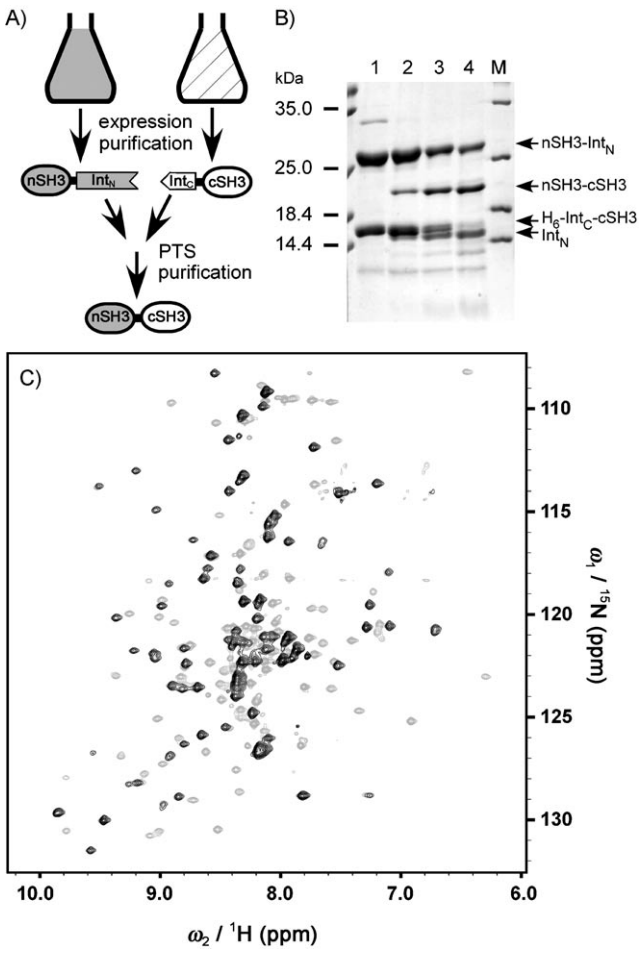


Figure 3. A) Scheme for segmental isotopic labelling by in vitro protein ligation. B) Time course of the in vitro protein ligation of nSH3 and cSH3 with *Npu*DnaE-Int$_N$/Int$_C$. Lane 1: immediately after mixing of nSH3-Int$_N$ and H$_6$-Int$_C$-cSH3, lane 2: 1 h after mixing, lane 3: 4 h after mixing, lane 4: after overnight incubation, M: molecular weight markers. C) Superposition of [^{15}N,^{1}H]-TROSY spectra of segmentally ^{15}N-labelled [^{15}N]-nSH3(131–205)-cSH3 (black) and fully labelled nSH3-cSH3(131–304) (grey).

segmental isotopic labelling was additionally confirmed by MALDI-TOF mass spectrometry as a 133 Da increase in the molecular weight due to the 133 nitrogen atoms in cSH3 (Figure 2C). It is thus clearly demonstrated that in vivo segmental isotopic labelling can be applied to biologically relevant multidomain proteins with high yields and sufficient selectivity of isotopic incorporation into the domain of interest.

Furthermore, only one additional medium replacement step during the protein expression (in comparison with standard protein expression) is required during the preparation, making it a simple and versatile approach. Prerequisites for this in vivo approach might be solubility and stability of individual precursors in *E. coli* cells and efficient trans-splicing activity achieved by the engineered split intein. Protein trans-splicing is not limited to in vivo conditions, but can also be induced in vitro, although individual fragments have to be produced and purified separately, as demonstrated in Figure 3. For protein ligation in vitro, the N-terminal precursor was prepared from two litres of M9 medium containing $^{15}NH_4Cl$ as sole nitrogen source and purified through the use of the N-terminal His-tag. The C-terminal precursor protein bearing a His-tag at the front of Int$_C$ was prepared separately from two litres of unlabelled LB medium. The two purified fragments were dialysed against Tris (10 mM), NaCl (500 mM), pH 7 and mixed in vitro in the presence of tris(2-carboxyethyl)phosphine hydrochloride (TCEP; 0.5 mM) in order to initiate protein trans-splicing. The reaction mixture was incubated overnight at 4 °C and further purified by use of a Histrap column and anion-exchange chromatography to remove the spliced intein and the residual precursor proteins. The [^{15}N,^{1}H]-TROSY spectrum of the N-terminally segmental isotope-labelled [^{15}N]-nSH3-cSH3 is shown in Figure 3C, indicating that only nSH3(131–205) is indeed ^{15}N-labelled. About 9 mg of the segmentally isotope-labelled sample was purified from two litres of each medium, so the yield was slightly lower than that of the preparation obtained by the in vivo approach.

Here, we have demonstrated that segmental isotopic labelling of a multidomain protein bearing a multiple copy of the SH3 domain could be efficiently achieved by protein trans-splicing either in vivo or in vitro. The in vivo approach may require fewer preparation steps than the in vitro approach, because it does not need individual expression and purification of the two precursor proteins and additional purification steps after the protein ligation in vitro (Figure 3 A). Segmental isotopic labelling by protein trans-splicing should be easily extendable to protein ligation of three fragments for isotopic labelling of a central domain with an additional split intein[7] or with EPL after the creation of an N-terminal cysteine residue. The in vivo approach by protein trans-splicing with naturally split inteins might be more advantageous than the in vitro approach because it does not increase the number of purification steps even if more fragments are to be ligated and because high selectivity of segmental labelling can be achieved as demonstrated here. Segmental isotopic labelling could be combined with other labelling methods such as methyl labelling and fractional labelling.[18,19] Thus, segmental isotopic labelling is very likely to contribute to NMR studies of large multidomain proteins because many proteins contain more than one modular domain that can be defined as structural and functional entities such as the SH3 domain. It also allows us to study large multidomain proteins by NMR without dissecting the protein into the smallest possible domains. Further discoveries and engineering of protein splicing domains suitable for every situation of the linker sequences with robust splicing activities could make segmental isotopic labelling an indispensable approach to the elucidation of structure–function relationships of intact multidomain proteins by NMR spectroscopy.

Acknowledgements

We thank Dr. P. Forrer (Molecular Partners, AG) for providing us with the plasmid of pAT044 and the Protein Chemistry Research Group at the Institute of Biotechnology for mass spectrometry analysis. This work is supported by grants from the Academy of Finland (118385), the Sigrid Jusélius Foundation, and the Biocentrum Helsinki.

Keywords: inteins · NMR spectroscopy · protein design · protein splicing · segmental labelling

[1] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki , N. Ito, Y. Ishino, H. Nakamura, *J. Am. Chem. Soc.* **1998**, *120*, 5591–5592.
[2] R. Xu, B. Ayers, D. Cowburn, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388–393.
[3] a) H. Iwai, S. Züger, *Biotechnol. Genet. Eng. Rev.* **2007**, *24*, 129–146; b) C. Fernández, G. Wider, *Curr. Opin. Struct. Biol.* **2003**, *13*, 570–580.
[4] a) K. Wüthrich, *Nat. Struct. Biol.* **1998**, *5 (Suppl.)*, 492–495; b) J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, *Nature* **2002**, *418*, 207–211.
[5] S. Ohki, M. Kainosho, *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 208–226.
[6] L. Skrisovska, F. H. T. Allain, *J. Mol. Biol.* **2008**, *375*, 151–164.
[7] T. Otomo, N. Ito, Y. Kyogoku, T. Yamazaki, *Biochemistry* **1999**, *38*, 16040–16044.
[8] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705–6710.
[9] H. Paulus, *Annu. Rev. Biochem.* **2000**, *69*, 447–496.
[10] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226–9231.
[11] T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki, Y. Kyogoku, *J. Biomol. NMR* **1999**, *14*, 105–114.
[12] D. D. Martin, M. Q. Xu. , T. C. Evans, *Biochemistry* **2001**, *40*, 1393–1402.
[13] S. Züger, H. Iwai, *Nat. Biotechnol.* **2005**, *23*, 736–740.
[14] H. Iwai, S. Züger, J. Jin, P. H. Tam, *FEBS Lett.* **2006**, *580*, 1853–1858.
[15] T. Pawson, *Nature* **1995**, *373*, 573–580.
[16] A. S. Aranko, H. Iwai, unpublished results.
[17] C. T. Reichman, B. J. Mayer, S. Keshav, H. Hanafusa, *Cell Growth Differ.* **1992**, *3*, 451–460.
[18] V. Tugarinov, V. Kanelis, L. E. Kay, *Nat. Protocols* **2006**, *1*, 749–754.
[19] H. Iwai, J. Fiaux, *J. Biomol. NMR* **2007**, *37*, 187–193.

Received: September 8, 2008
Published online on November 21, 2008

