Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing *in vivo* and *in vitro*

Mikko Muona, A Sesilja Aranko, Vytas Raulinaitis & Hideo Iwaï

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki, Finland. Correspondence should be addressed to H.I. (hideo.iwai@helsinki.fi).

Segmental isotopic labeling is a powerful labeling technique for reducing nuclear magnetic resonance (NMR) signal overlap, which is associated with larger proteins by incorporating stable isotopes into only one region of a protein for NMR detections. Segmental isotopic labeling can not only reduce complexities of NMR spectra but also retain possibilities to carry out sequential resonance assignments by triple-resonance NMR experiments. We described *in vivo* (i.e., in *Escherichia coli*) and *in vitro* protocols for segmental isotopic labeling of multi-domain and fusion proteins via protein *trans*-splicing (PTS) using split DnaE intein without any refolding steps or α-thioester modification. The advantage of PTS approach is that it can be carried out *in vivo* by time-delayed dual-expression system with two controllable promoters. A segmentally isotope-labeled protein can be expressed in *Escherichia coli* within 1 d once required vectors are constructed. The total preparation time of a segmentally labeled sample can be as short as 7–13 d depending on the protocol used.

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is an indispensable tool to determine protein structures and to study interactions and dynamics of proteins. Because of slower molecular tumbling of larger proteins, however, NMR signals of larger proteins become broader, thereby reducing signal-to-noise ratios. Moreover, overlapping of NMR signals of larger proteins also makes it harder and time-consuming to interpret NMR spectra. Thus, broadening and overlapping of NMR signals are the underlying difficulties of NMR studies of larger proteins. Transverse-relaxation optimized NMR spectroscopy has alleviated the line-broadening effect of larger proteins[1,2]. The increased number of atoms in large proteins (>25–30 kDa) intrinsically increases the chances of signal overlap, which could be difficult to be resolved even with extravagant ultra-high field magnets. Reducing the number of signals is one way to cope with the problem related to larger proteins[3]. Traditionally, selective amino-acid labeling has been exploited for reducing the number of signals and for identifying residue-types by feeding one or more labeled amino-acids into organisms expressing the proteins of interest. However, selective amino-acid labeling creates additional problems such as the loss of possibilities to perform sequence-specific resonance assignments because it is no longer simple to relate signals originating from neighboring residues for the sequential assignments. It is also necessary to prepare a number of samples with different amino-acid combinations to obtain sequence-specific assignments[4]. In contrast, sequential resonance assignments could be obtained with a single fully [13]C- and ^{15}N-labeled sample by many elegant triple-resonance experiments, which enables to connect neighboring residues through heteronuclear J-couplings[5]. NMR analysis with uniformly [15]N- and ^{13}C-labeled samples becomes increasingly difficult and time-consuming when the proteins become bigger and/or contain recurring sequences. The most promising approach for reducing the complexity of NMR spectra is to exclusively incorporate NMR-active stable isotopes such as ^{13}C and ^{15}N into only one region of a protein, which is often called 'segmental labeling'[6,7]. The isotope-labeled region is still

fully [13]C- and ^{15}N-labeled, therefore permitting the use of triple-resonance experiments without disturbances from the unlabeled regions. Segmental labeling also does not require as many samples as selective amino-acid labeling for obtaining sequence-specific assignments because triple-resonance experiments can be applied in the same way as for fully [13]C- and ^{15}N-labeled samples. Segmental labeling could, therefore, be very useful for studying a domain or a region of a large protein. However, the preparation of segmental isotope-labeled samples has been technically demanding and often results in relatively poor yields because of the procedures previously used for protein ligation[6–10]. These major obstacles have prevented the widespread application of segmental labeling techniques. In the past years, we have developed a novel protocol for segmental isotopic labeling overcoming these obstacles[11,12]. The protocol described in this article makes use of protein *trans*-splicing (PTS) by a naturally split intein that catalyzes protein ligation without any refolding steps[11,13,14] or C-terminal thioester modification and an N-terminal cysteine required for other segmental labeling approaches, such as expressed protein ligation (EPL) and native chemical ligation (NCL)[7,15,16]. PTS is an auto-catalytic process requiring only the formation of an active form of the split intein to induce protein ligation (**Fig. 1**)[13,14]. Therefore, PTS can be carried out not only *in vitro* but also *in vivo*. This makes it possible to simplify the segmental labeling procedure by using the time-delayed dual-expression system, as demonstrated previously (**Fig. 2**)[11,12]. The problem associated with low yields and the sequence dependences at the splicing junctions has been recently alleviated by the robust activity of the naturally split dnaE intein from *Nostoc punctiforme* (*Npu*)[17,18].

Segmental isotopic labeling by naturally split DnaE inteins both *in vitro* and *in vivo* is a routine protocol in our laboratory[11,12]. However, the protocol described here might not be suitable for segmental isotopic labeling of a self-contained globular domain because proteins usually become insoluble when a globular domain is split into two pieces and, at least, one of the precursor proteins

Figure 1 | Canonical reaction steps of protein *trans*-splicing (PTS). Two separate precursor proteins are encoded in two separate genes. Upon association of the two precursor chains containing the two split-intein fragments (Int_N and Int_C), the four concerted reactions of (1) N–S acyl shift, (2) *trans*-esterification, (3) Asn cyclization and (4) S–N acyl shift will take place, resulting in the ligation of the two flanking sequences (Ext_N and Ext_C) with a peptide bond.
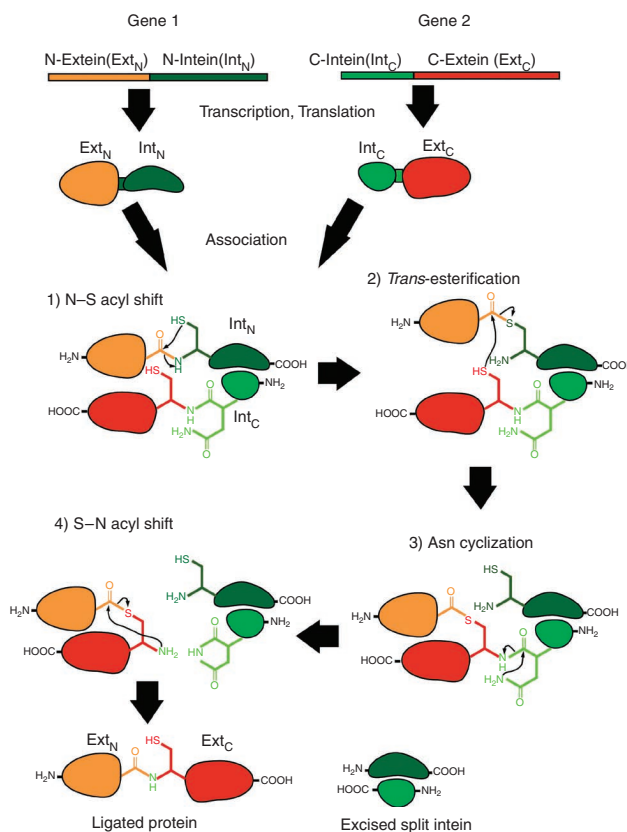
has to be soluble for this protocol. Thus, the solubility of precursor proteins fused with split intein fragments could be a limiting factor for this protocol, although refolding of the insoluble precursors can be considered as an alternative approach. The solubility issue of recombinant proteins is not a specific problem to our ligation system by PTS but a general issue of recombinant expression of foreign proteins in *Escherichia coli*[19]. One of the advantages of our protocol is that one can conveniently evaluate the feasibility of protein ligation by PTS without any tedious purification steps when the two required plasmids were constructed. Moreover, it requires only one additional step of medium replacement during protein expression compared with a conventional protein expression. Therefore, the protocol described here is technically less demanding than the approaches previously reported[6,7,9,10,16]. The segmental isotopic labeling by PTS could be very useful for quick assignments of NMR signals to segmentally labeled regions and for suppression of undesired NMR signals originating from repetitive protein sequences, flexible regions of proteins and fusion tags, such as solubility enhancement tags[12,19,20]. Segmental isotopic labeling by PTS could also be used for in-cell NMR spectroscopy to detect a region of interest of a protein in living systems.[21] Thus, the simple and robust segmental labeling protocol described here could be widely applied in protein NMR spectroscopy.

Experimental design

The following protocol is a procedure to prepare segmentally isotope-labeled samples for multi-domain and fusion proteins, excluding proteins composed of a single self-contained domain that would most likely require refolding steps. The first step of segmental isotopic labeling is to test protein ligation of the two split fragments from a protein of interest (POI). When *in vivo* ligation

by PTS cannot be detected in the SDS-PAGE analysis of the control experiments, it might be unrealistic to obtain sufficient amounts of segmentally isotope-labeled samples for NMR studies both *in vivo* and *in vitro*. The key to produce segmentally labeled samples is to establish protein ligation by PTS with a high yield. This protocol utilizes the dual-expression system developed in our laboratory to test protein ligation *in vivo*, in which N- and C-terminal precursors are cloned into two separate plasmids that can be induced by two different inducers (i.e., L-arabinose and isopropyl β-D-1-thiogalactopyranoside (IPTG)) (**Fig. 3**). One of the crucial steps is to design the two precursor proteins fused with the two split fragments of POI (see Design of plasmids). The comparison between the individual expressions of the two precursors by the individual inducers (L-arabinose or IPTG respectively) and the dual expression of the two precursors by the simultaneous induction with the two inducers should visualize the effects of the dual expression. When PTS is initiated by the interaction between the two precursors, additional bands in the SDS-PAGE are
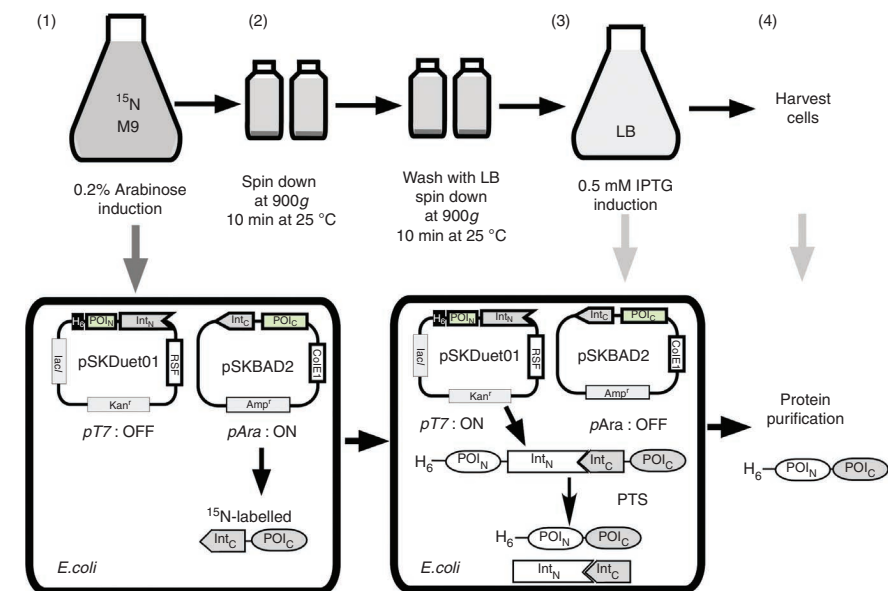

Figure 2 | Outline of the *in vivo* segmental isotope-labeling procedure described in the text. (1) Induction of the first precursor in the labeled medium by L-arabinose. (2) Replacement and washing of the cell culture with unlabeled (Luria–Bertani (LB)) medium, which turn off the expression of the first precursor. (3) Induction of the second precursor in unlabeled medium. (4) Protein purification.

Figure 3 | Plasmid maps of the vectors used for protein ligation by protein *trans*-splicing (PTS). (**a**) pSKDuet01 for the N-terminal precursor. It contains Int_N of *Npu*DnaE intein under T7 promoter, *RSF* origin (*RSF*), kanamycin resistant gene (Kan[r]) and lac repressor gene (lacI). The DNA sequence of *Bam*HI site for the junction between the N-terminal fragment of the protein of interest (POI_N) and Int_N is shown in bold. −1 indicates the preceding residue of the N-terminal intein fragment (Int_N). (**b**) pSKBAD2 for the C-terminal precursor. This vector contains Int_C fragment of *Npu*DnaE intein under arabinose promoter, *ColE1* origin (*ColE1*) and ampicillin resistant gene (Amp[r]). The cloning site of *Kpn*I is highlighted in bold. The sequence corresponding to amino-acid sequence of 'CFN' at the front of *Kpn*I site is inserted to improve the ligation efficiency. The cysteine residue indicated in italic (+1) is indispensable for protein-splicing reaction but can be replaced with serine (see text). The change to serine could reduce the ligation yield. +1 and +2 indicate the first and second residue of the C-terminal extein (POI_C), which could affect the PTS activity. (**c**) pMMRSF17 for the N-terminal precursor containing a fusion protein (Smt3) at the N-terminus, which can be digested away by Ulp1 (ref. 11). (**d**) pHYRSF1-02 for the C-terminal precursor containing the identical gene in pSKBAD2, but with an N-terminal His-tag. The expression is controlled under T7 promoter. This vector can be used for *in vitro* protocol. The plasmids described here are available from Addgene (http://www.addgene.org/).
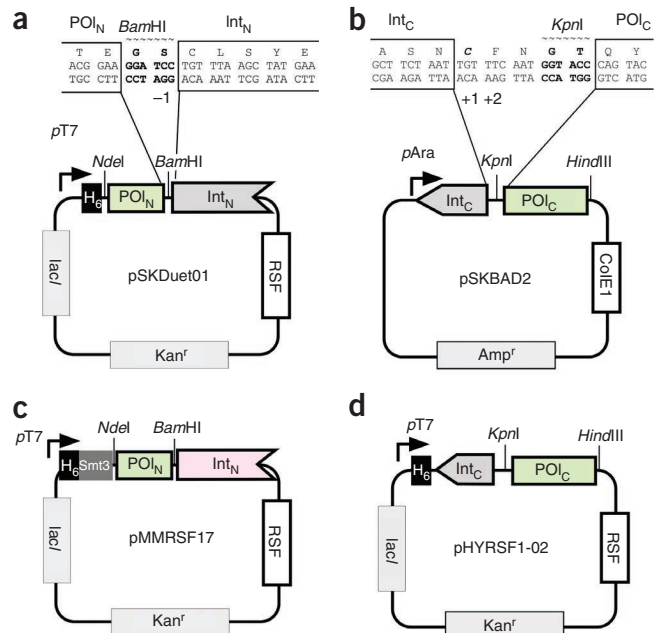
expected to appear. These bands are expected to be different from the two precursor proteins indicating spliced products (**Figs. 4** and **5a**). In the best-case scenario, only one new dominant band corresponding to the ligated product will emerge upon protein ligation, as all the precursors will disappear immediately after protein splicing. The inducer concentrations and growth temperatures might be optimized to maximize the yield of the ligated product. Cleaved Int_N might be visible at around 15 kDa (or other sizes, depending on the split intein used) (**Fig. 4**). When the dual expression is a mere overlay of the two single expressions by the individual inducers, PTS is not initiated (**Fig. 4b**, lane c; **Fig. 5a**, lane c). In this case, it is most likely that either both the precursors or one of the precursors is insoluble (see TROUBLESHOOTING). Because the interaction between Int_N and Int_C is required for PTS and *trans*-cleavages, PTS cannot take place when the precursors are insoluble and inaccessible for the association. It is advisable to check the solubility of both N- and C-terminal precursors. It is noteworthy that the attachment of Int_N or Int_C could decrease the solubility of the target fragments. If both precursor proteins are insoluble, it might be necessary to change the split site of the POI and/or attach a solubility enhancement protein tag such as maltose-binding protein, B1 domain of the immunoglobulin-binding protein G (GB1) or yeast Smt3 protein[11,19,20]. If only one of the precursor proteins is insoluble, protein ligation might still be achievable by expressing the soluble precursor before the other insoluble precursor. In this case, the timing of the inductions and the amounts of the inducers might have profound

influences on the yield of the ligated product (**Fig. 5**). If protein ligation cannot be detected by inducing the two precursors at the same time, it is advisable to probe a few different inducer concentrations with different orders to maximize the amount of ligated products (see Expression test, **Fig. 5**). Another problem associated with PTS can be the competing side-reactions. Instead of PTS reaction, *trans*-cleavage reactions either at the N- and/or C-terminal junctions might take place, thereby significantly reducing the ligation yield (**Fig. 5a**, lane e)[22]. This should be checked by identifying cleaved Int_N and split POI fragments in the control experiments (**Fig. 5a**, lane e). Another consideration for designing the experiment is the incorporation of a purification tag. For *in vivo* ligation, we typically add a purification tag (His-tag) to only one of the precursor proteins (e.g., N-terminal precursor). The use of only one purification tag in one of the precursors facilitates efficient removal of the other precursor bearing no tag during the first purification via the purification tag and could simplify the protein purification. However, we prepare an additional variant of the precursor with a tag (e.g., an N-terminal His-tag at the front of C-terminal split intein) for *in vitro* protein ligation (**Fig. 3d**). *In vitro* protein ligation is particularly useful to analyze limiting steps in protein *trans*-splicing reaction[22]. Once protein ligation by PTS is established, segmental isotopic labeling can be straightforward, as it is usually simple to scale-up the production. The major difference between a standard protein-expression procedure and the segmental isotope-labeling procedure is merely one additional step to replace the inducer and
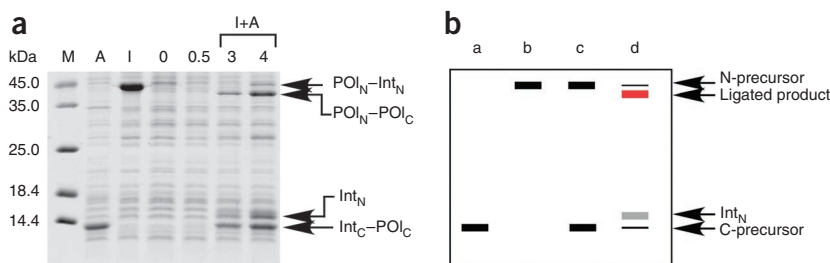

Figure 4 | Demonstration of *in vivo* protein ligation by the dual expression. (**a**) Comparison of the single expressions of the individual precursors (lanes A and I) and the dual expression (lanes 3 and 4). Lane A: induction by L-arabinose; lane I: induction by IPTG. 'I+A' indicates the dual induction with L-arabinose and IPTG. Lane 0: before induction; lane 0.5: 0.5 h after the induction with L-arabinose; lane 3: 3 h after the dual induction with L-arabinose and IPTG; and lane 4: 4 h after the dual induction. (**b**) Schematic illustration of the possible patterns in SDS-PAGE upon dual expression.

Figure 5 | The effect of the induction order and inducer concentrations on protein ligation of A and R domains from AlgE4 epimerase (A-R). (**a**) Schematic illustration of all possible patterns in SDS-PAGE upon the dual expression. (**b**) Time-course of the *in vivo* protein ligation when the N-terminal precursor (A-Int$_N$) was induced first with IPTG (0.5 mM) and followed by the expression of the C-terminal precursor (Int$_C$-R) with various concentrations of L-arabinose (0.02, 0.05, 0.1 and 0.2%). (**c**) Time-course of the *in vivo* protein ligation when the N-terminal precursor (A-Int$_N$) induced with IPTG after the induction of the C-terminal precursor by L-arabinose. The times after the first induction are indicated at the top of the lanes. The positions of the ligated product (A-R) and the two precursors (A-Int$_N$ and Int$_C$-R) are indicated by arrows. A-R indicates the ligated product of AlgE4 which consists of the A and R domains[24].

the culture media containing the isotopes, although the replacement and sequential expression could lower the ligation yield compared with the simultaneous dual expression. Downstream purification processes of the ligated protein should be similar to those of the full-length proteins and has to be established separately before the segmental isotopic labeling. Typically, we introduce an N-terminal His-tag in the N-terminal precursor proteins. Thereby, the elution from IMAC would contain the N-terminal precursor (and the C-terminal precursor protein if both precursors interact but splice slowly) and the ligated product. The ligated product can be further purified (e.g., by ion-exchange chromatography) after the cleavage of the purification tag to obtain a segmentally isotope-labeled sample of high purity.

Design of plasmids

The most important consideration in designing plasmids for protein ligation by PTS is to define appropriate sequences near the splicing junctions, which could significantly affect the final yield. In other words, it is critical to decide an appropriate location to divide the POI into two parts (POI$_N$ and POI$_C$) and the sequences connecting the split fragments and the intein sequences. Therefore, it has to be possible to identify a linker region connecting domains from the primary sequence alignments or from known structures. Once the split site is determined, the cloning strategy should be determined. One approach is to use the restriction sites in the plasmids depicted in **Figure 3**. In this protocol, we use a dual-expression system that uses two compatible plasmids with two different inducible promoters (T7 and arabinose promoters). The first plasmid (pSKDuet01) bearing kanamycin resistance and *RSF* origin is under the control of T7 promoter, which can be induced by IPTG (**Fig. 3a**)[12,17]. The second plasmid (pSKBAD2) contains ampicillin resistance gene, *ColE1* origin, and an arabinose promoter, which is inducible by L-arabinose (**Fig. 3b**)[12,17,23]. In these plasmids, *Bam*HI and *Kpn*I can be used for cloning of the split protein fragments at the splicing junctions. The amino acid sequence of 'GS' will be inserted at the front of Int$_N$ if *Bam*HI in pSKDuet01 is used. The amino acid sequence of 'CFNGT' will be inserted between Int$_C$ and POI$_C$ when the gene is cloned between the *Kpn*I and *Hind*III sites in pSKBAD2. Because cloning with *Bam*HI and *Kpn*I sites in those plasmids will result in an insertion of a primary sequence 'GSCFNGT' between POI$_N$ and POI$_C$, seven residues from the original linker should be removed to keep the same length as the original full-length protein. When changes in the linker sequence are unfavorable for its biological activity, one might consider either to convert this sequence closer
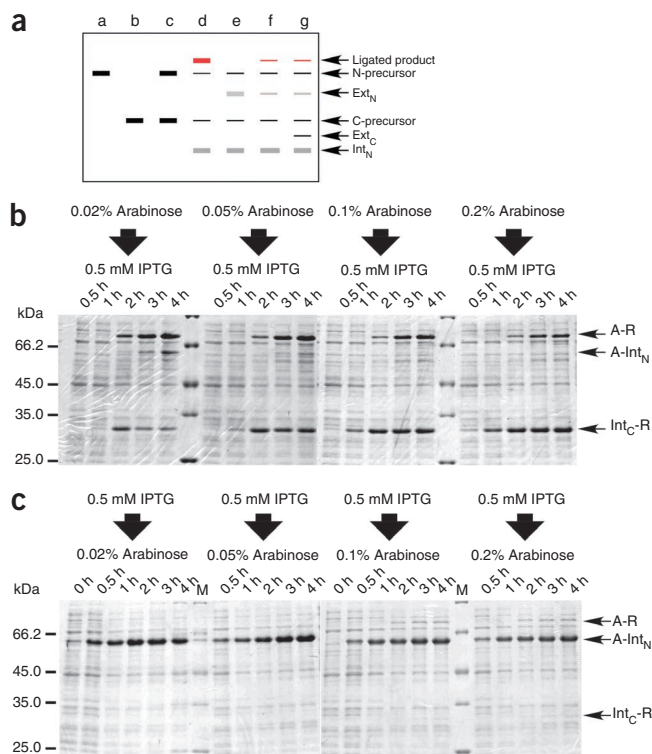
to the original linker sequence by site-directed mutagenesis or to create the desired junction sequence by assembly PCR[24]. However, the first residue (+1) of C-extein (POI$_C$) must be either Cys or Ser, which is indispensable for *trans*-esterification during protein splicing reaction (**Figs. 1** and **3b**)[14]. In addition, it is noteworthy that the second residue (+2) of C-extein (it corresponds to Phe when pSKBAD2 is directly used) has strong influence on the ligation yield[17]. On the other hand, the last residue of N-extein, which is located at the N-terminal junction and often called −1 position, seems to have little influence on the splicing activity of *Npu*DnaE intein because we tested Ser, Asn, Tyr, Phe, His, Asp, Glu, Lys and Gln at this position and they yielded well with a model system (unpublished data). For *in vitro* protein ligation, an N-terminal His-tag can be incorporated at the front of the C-terminal precursor for IMAC purification by cloning the gene of interest into the plasmid pHYRSF1-02 with the same restriction sites as pSKBAD2 (**Fig. 3d**). When the protein ligation by PTS is not observed (see TROUBLESHOOTING), it might be worthwhile to try other inteins and/or a fusion system as depicted in **Figure 3c** (refs. 11,12).

Timing of inductions and media exchange

For *in vivo* segmental isotopic labeling described in this protocol, the timing of inductions and medium exchange might be crucial steps to obtain high yields and high selectivity of isotopic labeling (**Fig. 2**). Although the expression of the precursor protein controlled by arabinose promoter can be effectively turned off after the removal of L-arabinose from the culture medium because of its tight regulation[23], T7 promoter cannot be sufficiently shut off when the induction by IPTG is used. Therefore, the first induction is recommended to be under the arabinose promoter and followed by the second induction with IPTG for T7 promoter to suppress cross-labeling due to the leaky expression.

BOX 1 | SAMPLE PREPARATION FOR SDS-PAGE ANALYSIS

To prevent possible protein degradation of the sample, it is recommended to prepare SDS-PAGE samples immediately.

1. Measure the OD_{600} of each 0.5 ml sample with NanoDrop UV spectrometer and pellet the cells for 1 min at 14,000g at room temperature. Discard the supernatant.

2. Add 1× SDS-PAGE sample buffer to the pellet. The amount of buffer used for each sample should be based on the OD_{600} and should adjust the cell density to the equivalent of a final OD_{600}=5.0, which keeps the cell concentrations in the sample buffer uniform among all the collected SDS-PAGE samples.

3. Vortex the tube to resuspend the cells in SDS-PAGE sample buffer for 20 s and store the samples at −20 °C.

It is important to minimize the time required for the exchange of the inducer and medium to obtain higher yields and highly selective isotope incorporation. The other consideration is the solubility of the two precursors. If one of the precursors is not highly soluble, then the soluble precursor needs to be induced before the insoluble precursor because the insoluble precursor will not be accessible for the interaction to induce PTS. The vectors encoding the two precursors should be constructed considering these factors, depending on the region of interest for segmental labeling. In all the cases, the protein ligation should be tested and optimized in small-scale expression tests by changing timing of inductions, growth temperatures and concentrations of inducers for maximizing ligation products in SDS-PAGE analysis (**Fig. 5**).

Overview of the protocol
In vivo segmental isotopic labeling is described in the PROCEDURE section and the method for preparing SDS-PAGE samples is found in **Box 1**. The *in vitro* approach is described in **Box 2**.

BOX 2 | *IN VITRO* SEGMENTAL ISOTOPIC LABELING

Plasmid transformation for the expression of individual precursor proteins ● TIMING 1 d

1. Prepare the two plasmids, each bearing one of the two precursor proteins, for the dual expression (see Design of plasmids).
2. To transform the cells, transfer 0.5 µl (50–100 ng) of each of the two constructed plasmids into two separate microcentrifuge tubes kept on ice and add 50 µl of competent *Escherichia coli* ER2566 cells to each tube.
3. Incubate the tubes on ice for 30 min.
4. Heat shock the cells for 45 s at 42 °C on a heating block and return them onto ice for 30 s.
5. Add 450 µl of pre-warmed Super optimal-broth (SOB) medium to each tube with the transformed cells on ice.
6. Place the tubes on a bench top shaker for 1 h, at 37 °C and 300 r.p.m.
7. Plate 100 µl of the tubes onto two separate ampicillin (100 µg ml⁻¹) or kanamycin (25 µg ml⁻¹) LB-agar plates.

Expression of individual protein segments ● TIMING 2 d

8. The procedures for individual precursor protein expression in unlabeled (option A) and labeled (option B) media are very similar and can be carried out simultaneously.

(A) Expression of unlabeled precursor protein
  (i) Pick a colony bearing the plasmid, which encodes the precursor protein selected to be unlabeled. Use the colonies to inoculate 50 ml of Luria–Bertani (LB) with an appropriate antibiotic for the plasmid (either kanamycin, 25 µg ml⁻¹ or ampicillin, 100 µg ml⁻¹).
  (ii) Incubate the cell culture overnight at 25–28 °C, 250 r.p.m.
  (iii) Next morning, transfer 10 ml of the overnight culture to inoculate each of the three 2 liter flasks containing 500 ml of LB medium with an appropriate antibiotic for the cells containing the plasmid (either kanamycin, 25 µg ml⁻¹ or ampicillin, 100 µg ml⁻¹).
  (iv) Incubate the cell culture at 37 °C, 250 r.p.m. Periodically check OD_{600} of the cell culture (every 30–60 min) by taking 2 µl of the cell culture and measuring the optical density with NanoDrop UV spectrometer.
  (v) When OD_{600} reaches 0.5–0.7, add the promoter-specific inducer (either at a final concentration of 0.5 mM IPTG or 0.1% (wt/vol) L-arabinose). Take a sample for SDS-PAGE analysis as described in **Box 1**.
  (vi) Continue incubating for 2–5 h (at 37 °C, 250 r.p.m.). During incubation, take a few samples for SDS-PAGE every 1–2 h as described in **Box 1**.
It should be noted that the incubation temperature may need to be adjusted depending on the protein of interest.
? TROUBLESHOOTING
  (vii) Add 100 µl of antifoam to each flask and transfer the culture suspension into two 1,000 ml centrifuge tubes.
  (viii) Harvest the cells by centrifugation for 10 min, at 9,000g, 4 °C.
  (ix) Discard the supernatant and resuspend the pellets with 20–30 ml of lysis and binding buffer.
■ PAUSE POINT The suspended cells can be stored at −80 °C in a 50 ml Falcon tube until protein purification.
(B) Expression of labeled precursor protein
  (i) Pick a colony bearing the plasmid encoding the precursor protein to be labeled. Use the colony to inoculate 5 ml of LB medium with an appropriate antibiotic (either ampicillin, 100 µg ml⁻¹ or kanamycin, 25 µg ml⁻¹). Incubate for 2–3 h, at 37 °C and 250 r.p.m.
  (ii) Transfer the LB medium culture into 50 ml of ¹⁵N-labeled M9 medium containing the appropriate antibiotic (either ampicillin, 100 µg ml⁻¹ or kanamycin, 25 µg ml⁻¹) that these cells are resistant to. Incubate overnight, at 25–28 °C, 250 r.p.m.
  (iii) The next day, transfer 15–25 ml of the overnight M9 culture to inoculate each of three 2-liter flasks containing 500 ml of M9 medium with the antibiotic (either ampicillin, 100 µg ml⁻¹ or kanamycin, 25 µg ml⁻¹).

(continued)

BOX 2 | (CONTINUED)

(iv) Incubate the cells at 37 °C, 250 r.p.m. Periodically check the cell growth by measuring OD_{600} (every 30–60 min) with NanoDrop UV spectrometer by taking 2 μl from the cell culture.

(v) When OD_{600} reaches 0.5–0.7, add the promoter-specific inducer at a final concentration of either 0.1% (wt/vol) L-arabinose or 0.5 mM IPTG.

(vi) Continue incubating (37 °C, 250 r.p.m.). During the incubation, collect a small amount of the cells and prepare SDS-PAGE samples (see **Box 1**).

It should be noted that the incubation temperature may need to be altered depending on the protein of interest.

? TROUBLESHOOTING

(vii) After the total incubation time (2–5 h), divide and transfer the culture suspension into two 1,000 ml centrifuge tubes.

(viii) Harvest the cells by centrifugation at 9,000*g*, 4 °C, for 10 min.

(ix) Discard the supernatant and resuspend the pellets carefully with 20–30 ml of lysis buffer.

■ **PAUSE POINT** The suspended cells can be stored at –80 °C in a 50 ml Falcon tube until protein purification.

Individual protein-precursor purification and dialysis ● **TIMING** 2–3 d

9. Purify each labeled and unlabeled protein under native conditions using appropriate procedures for the proteins of interest. For instance, IMAC can be used if His-tags are incorporated into the precursor proteins. Follow the instruction provided by the supplier such as HisTrap HP column (GE Healthcare, UK).

10. Dialyze protein fractions from Step 9 (total volume of ~20 ml) containing the N- or C-precursors against 3–4 litres of the ligation buffer overnight (4 °C with stirring). Note: Please avoid phosphate buffers which might deteriorate the reducing capacity of tris(2-carboxyethyl)phosphine hydrochloride solution (TCEP)[25,26].

11. Spin down dialyzed protein solution in a 50 ml Falcon tube using Eppendorf centrifuge (10 min, 4,000 r.p.m., 4 °C) and transfer supernatant into a new tube. Visually check if there is significant precipitation after dialysis.

? TROUBLESHOOTING

Protein ligation ● **TIMING** 1–2 d

12. Determine the concentrations of the N- and C-precursor proteins by UV spectrometer using theoretical extinction coefficients at 280 nm that are calculated from the protein sequences (e.g., enter the protein sequence into the web tool, such as http://ca.expasy.org/tools/protparam.html to calculate the theoretical extinction). Concentrate the proteins if needed.

13. Mix the two precursor proteins with 1:1 molar ratio at a final concentration of 20–100 μM in a total volume of 10–50 ml (a smaller volume is preferable for the subsequent purification steps).

14. Add TCEP at a final concentration of 0.5 mM.

It should be noted that before proceeding to a full-scale ligation, it may be useful to carry out a small-scale pilot ligation experiment. It may help to optimize the ligation conditions (temperature, duration).

Ensure that there is no phosphate ion in the ligation mixture. Phosphate ions might deteriorate the reducing ability of TCEP[25,26].

15. In the course of ligation reaction (i.e., after 5, 10 and 30 min and 1, 2, 4 and 12 h after mixing the two precursors), take small aliquots (e.g., 10 μl) of the reaction mixture and prepare SDS-PAGE samples by mixing them with 10 μl of 1× SDS buffer. Store the samples at –20 °C until analysis (the sample can be stored for a few days). Before loading the samples on the gel, heat them for 5 min at 95 °C. Load 10 μl of each sample onto 18% SDS-PAGE gel.

16. Using SDS-PAGE, evaluate the progress and completion of the protein ligation. Refer to **Figure 5a** for the interpretation of results.

? TROUBLESHOOTING

Segmentally labeled protein: dialysis and purification ● **TIMING** 1–2 d

17. Dialyze the entire reaction mixture against the buffer of your choice for further purification (e.g., 10 mM sodium phosphate buffer for ion-exchange chromatography).

18. Purify the ligated protein from the unprocessed precursors and spliced-intein fragments by available methods of one's choice such as ion-exchange chromatography and gel filtration.

MATERIALS

REAGENTS

Chemicals

- $FeCl_2 \cdot 4H_2O$ (Aldrich-Sigma Chemical, cat. no. 44939)
- H_3BO_3 (Aldrich-Sigma Chemical, cat. no. B6768)
- $CoCl_2 \cdot 6H_2O$ (Aldrich-Sigma Chemical, cat. no. 12914)
- $CuCl_2 \cdot H_2O$ (Aldrich-Sigma Chemical, cat. no. C3279)
- $ZnCl_2$ (Fluka Chemie GmbH, cat. no. 96468)
- $Na_2MoO_4 \cdot 2H_2O$ (Fluka Chemie GmbH, cat. no. 71756)
- $MnCl_2 \cdot 4H_2O$ (Aldrich-Sigma Chemical, cat. no. 221279)
- Glacial acetic acid (Aldrich-Sigma Chemical, cat. no. A9967)
- Glycerol (Aldrich-Sigma Chemical, cat. no. G5516)
- Bromophenol blue (Aldrich-Sigma Chemical, cat. no. B8026)
- $^{15}NH_4Cl$ (Aldrich-Sigma Chemical, cat. no. 299251)

- 2-Mercaptoethanol (Aldrich-Sigma Chemical, cat. no. M3148)
- 40% Acrylamide/bis solution (Aldrich-Sigma Chemical, cat. no. A6050)
- Agar (Hispanlab S. A., cat. no. H1800)
- Ampicillin (MP Biomedicals, cat. no. 190148)
- Ammonium persulfate (Bio-Rad Laboratories, cat. no. 161-0700)
- Antifoam A (Aldrich-Sigma Chemical, cat. no. 10794) **! CAUTION** Antifoam A is an irritant to eyes. Wear protective safety glasses.
- Butyl alcohol (butanol, Aldrich-Sigma Chemical, cat. no. 71-36-3)
- $CaCl_2 \cdot 2H_2O$ (Merck, cat. no. 102382)
- D-(+)-Glucose monohydrate (Fluka Chemie GmbH, cat. no. 49160)
- EDTA (J.T. Baker, cat. no. 1073)
- HCl (Aldrich-Sigma Chemical, cat. no. 30721) **! CAUTION** HCl is corrosive. Wear protective clothing, gloves and safety glasses.

- Imidazole (Aldrich-Sigma Chemical, cat. no. I0250)
- IPTG (VWR BDH, cat. no. 437144N)
- Kanamycin (Aldrich-Sigma Chemical, cat. no. K-4000)
- KCl (Aldrich-Sigma Chemical, cat. no. 31248)
- KH_2PO_4 (Aldrich-Sigma Chemical, cat. no. 60218)
- L-(+)-Arabinose (Fluka Chemie GmbH, cat. no. 10845)
- $MgCl_2 \cdot 6H_2O$ (Merck, cat. no. 105833)
- $MgSO_4 \cdot 7H_2O$ (Merck, cat. no. 105886)
- Minimum essential medium (MEM) vitamin solution (100×) (Aldrich-Sigma Chemical, cat. no. M6895)
- $Na_2HPO_4 \cdot 2H_2O$ (Fluka Chemie GmbH, cat. no. 71633)
- NaCl (Aldrich-Sigma Chemical, cat. no. 31434)
- NaH_2PO_4 (Aldrich-Sigma Chemical, cat. no. S8282)
- NaOH (J.T. Baker, cat. no. 0402) ❗ **CAUTION** NaOH is corrosive. Wear protective clothing, gloves and safety glasses.
- Sodium dodecyl sulfate (SDS; Invitrogen, cat. no. 15553-027)
- Sterile deionized water
- Tetramethylethylenediamine (TEMED; Fluka Chemie GmbH, cat. no. 87689)
- Tris (MP Biomedicals, cat. no. 819638)
- Tris(2-carboxyethyl)phosphine hydrochloride (TCEP; Aldrich-Sigma Chemical, cat. no. 646547)
- Tryptone (Hispanlab S. A., cat. no. H1612)
- Yeast extract (Hispanlab S. A., cat. no. H1702)

Vectors and bacterial strains
- pSKDuet01 (plasmid #12172, Addgene, http://www.addgene.org/)
- pSKBAD2 (also called pSKBAD02)[8] (plasmid #15335, Addgene)
- pHYRSF1-02 (available from Addgene for academic researchers)
- pMMRSF17 (plasmid #20178, Addgene)
- pMMSFR1-16 (plasmid #20177, Addgene)
- pMMBAD16 (plasmid #20176, Addgene)
- *Escherichia coli* ER2566 (T7 express competent *E. coli*, New England Biolabs, cat. no. C2566H)

Enzymes
- Restriction endonucleases: *Bam*HI, *Nde*I, *Kpn*I and *Hind*III (MBI Fermentas, cat. no. R0136S, R0540S, R0111S, R0122S and R0180S, respectively)
- T4-DNA ligase (MBI Fermentas, cat. no. El0011)

Molecular weight markers
- GeneRule 1 kb Plus (MBI Fermentas, cat. no. SM1331)
- Unstained protein molecular-weight marker (MBI Fermentas, cat. no. SM0431)

Kits
- QIAquick gel extraction kit (Qiagen, cat. no. 28704)
- QIAprep spin miniprep kit (Qiagen, cat. no. 27104)
- GeneJET plasmid miniprep kit (MBI Fermentas, cat. no. K0503)

Buffers and solutions
- 10× Electrode buffer for SDS-PAGE (10× EB; see REAGENT SETUP)
- 10× Phosphate-buffered saline buffer (10× PBS; see REAGENT SETUP)
- EDTA (0.5 M (pH 8); see REAGENT SETUP)
- Elution buffers (see REAGENT SETUP)
- LB medium (Luria–Bertani broth; see REAGENT SETUP)
- SOB medium (Super optimal broth; see REAGENT SETUP)
- Lysis buffers (1× and 8×; see REAGENT SETUP)
- Running buffers for agarose gel electrophoresis (50× and 1× TAE buffers; see REAGENT SETUP)
- Sample buffer (2× SDS reducing buffer; see REAGENT SETUP)
- SDS-PAGE resolving buffer (1.5 M Tris-HCl, 0.4% (wt/vol) SDS, pH 8.8; see REAGENT SETUP)
- SDS-PAGE stacking buffer (0.5 M Tris-HCl, 0.4% (wt/vol) SDS, pH 6.8; see REAGENT SETUP)
- Sodium phosphate buffer (0.5 M (pH 8); see REAGENT SETUP)
- TE buffer (see REAGENT SETUP)
- Tris-HCl buffers (1 M (pH 8) and 0.5 M (pH 6.8); see REAGENT SETUP)
- Wash buffer (see REAGENT SETUP)

EQUIPMENT
- Sorvall Evolution RC centrifuge (Thermo Scientific, cat. no. 728211)
- Beckman Coulter Avanti J-20 XP centrifuge (Beckman Coulter, cat. no. 368640)
- Eppendorf table-top centrifuge 5804R (Eppendorf AG, cat. no. 5805 000.017)
- Microcentrifuge (Eppendorf MiniSpin, Eppendorf AG, cat. no. 5452 000.018)
- SLA-3000 Lite Rotor (Thermo Scientific, cat. no. 07149)
- JLA-8.1000 rotor (Beckman Coulter, cat. no. 363688)

- Eppendorf swing bucket rotor with four Falcon buckets (Eppendorf AG, cat. no. 5804 719.000)
- 1,000 ml tubes for JLA-8.1000 rotor (Beckman Coulter, cat. no. 363676)
- 500 ml tubes for SLA-3000 rotor (Thermo Scientific, cat. no. 07149)
- 500 ml centrifuge bottles with screw caps (Nalgene, cat. no. 3120–9500)
- 50 ml BD Falcon conical tubes (BD Falcon, cat. no. 352070)
- 15 ml conical tubes (Genesee Scientific, cat. no. 21–103)
- 1.7 ml microcentrifuge tubes (Midwest Scientific, cat. no. AVSS1700)
- Incubator shaker with cooling system (Kühner SM, Adolf Kühner AG, cat. no. 1701)
- Shaking incubator (VWR, cat. no. 35962–093)
- Eppendorf Thermomixer Comfort R (Aldrich-Sigma Chemical, cat. no. Z605271)
- Vortex shaker (Vortex Genie 2, Scientific Industries, cat. no. SI-0256)
- Bench-top shaking platform (Eppendorf AG, cat. no. 14005–830)
- UV Spectrophotometer (NanoDrop 1000, NanoDrop Technologies, cat. no. ND-1000)
- 2,000 ml culture flasks with baffles (Nalgene, cat. no. 4110–2000)
- 250 ml culture flasks with baffles (Nalgene, cat. no. 4110–0250)
- 2-liter glass Erlenmeyer flasks (Kimble-Chase, cat. no. 265002000)
- 4-liter beakers (Kimble-Chase, cat. no. 140054000)
- 20.0 ml syringes (BD Medical, cat. no. 301031)
- 0.22 μm syringe filters (Nalgene, cat. no. 190-2520)
- Sterile 100 mm × 25 mm petri dishes (Midwest Scientific, cat. no. TPP 93100)
- Surgical blades no. 21 (Henry Schein, cat. no. 100–3535)
- 10 ml Poly-prep chromatography columns (Bio-Rad, cat. no. 731–1550)
- SpectraPor 6,000–8,000 molecular weight cut-off dialysis tubing (Spectrum Laboratories, cat. no. 131300)
- Analytical balance (Mettler Toledo, cat. no. XP105DR)
- Ultra-clear cellophane (Research Products International, cat. no. 1090)
- Mini-PROTEAN3 cell system (Bio-Rad Laboratories, cat. no. 165–3323)

REAGENT SETUP
0.5 M EDTA solution (pH 8.0) For 1 liter, dissolve 186.12 g of EDTA in 800 ml of deionized H_2O. Adjust pH to 8.0 with NaOH. Add water to make a final volume of 1 liter and autoclave. EDTA solution can be stored at room temperature for 1 year.
0.5 M Sodium phosphate buffer (pH 8.0) Dissolve appropriate amount of sodium phosphate (monobasic or dibasic) in 800 ml of deionized H_2O. Adjust pH to 8.0 with NaOH. Add water to make a final volume of 1 liter. Sodium phosphate solution can be stored at room temperature for 1 year.
0.5 M Tris-HCl buffer (pH 6.8) For 1 liter, dissolve 60.55 g of Tris in 800 ml of deionized H_2O. Adjust pH to 6.8 by adding HCl. Add water to make a final volume of 1 liter and autoclave. Buffer can be stored at room temperature for 1 year.
Electrode buffer for SDS-PAGE (10×) For 1 liter, take 30 g of Tris, 144 g of glycine and 10 g of SDS. Dissolve the reagents in deionized water for a final volume of 1 liter. Do not sterilize. Dilute 10× electrode buffer tenfold before use. Buffer can be stored at room temperature for 1 year.
Elution buffer (1×) Take 100 ml of 0.5 M sodium phosphate buffer (pH 8.0) and add 60 ml of 5 M NaCl and 500 ml of 0.5 M imidazole, for their final concentrations of 50 mM, 300 mM and 250 mM, respectively. Check the pH and adjust it to 8.0 if necessary with NaOH. Add deionized water to make a final volume of 1 liter and filter through 0.22 μm syringe filters. Elution buffer can be stored at room temperature for 1 year.
LB-agar Prepare LB-agar by weighing 13.5 g of agar, 9 g of tryptone, 4.5 g of yeast extract and 4.5 g of NaCl. Dissolve the constituents by heating in 900 ml of deionized water and autoclave. Store at 4 °C for up to 1 month.
LB-agar plates Heat LB-agar to melt it and let it cool down to 55 °C. Add antibiotics as required (for final concentrations of 100 μg ml^{-1} for ampicillin and 25 μg ml^{-1} for kanamycin). Pour LB-agar into petri dishes (~25 ml per plate) and store at 4 °C for up to 2 weeks.
LB-medium For 1 liter, weigh 10 g of tryptone, 5 g of yeast extract and 5 g of NaCl. Dissolve all the reagents in 800 ml of deionized water and adjust the pH to 7.4 by adding NaOH. Then add deionized water for a final volume of 1 liter and autoclave. LB-medium can be stored at 4 °C for 1 month.
Ligation buffer (10 mM Tris-HCl, 500 mM NaCl and 1 mM EDTA, pH 7.0) For 1 liter, dissolve 1.21 g of Tris, 29.25 g of NaCl in 800 ml of distilled H_2O and add 2 ml of 0.5 M EDTA solution (pH 8.0). Adjust pH to 7.0 with

HCl. Add deionized water to make a final volume of 1 liter and filter through 0.22 μm syringe filters. Ligation buffer can be stored at room temperature for 1 year.

Lysis and binding buffer (1×) Take 100 ml of 0.5 M sodium phosphate buffer (pH 8.0) and add 60 ml of 5 M NaCl and 20 ml of 0.5 M imidazole, for their final concentrations of 50 mM, 300 mM and 10 mM, respectively. Check the pH and adjust it to 8.0 if necessary with NaOH. Add deionized water to make a final volume of 1 liter and filter through 0.22 μm syringe filters. Lysis and binding buffer can be stored at room temperature for 1 year.

¹⁵N-labeled M9-medium For 1 liter, use 7.2 g of $Na_2HPO_4·H_2O$, 3.0 g of KH_2PO_4, 0.5 g of NaCl, 1.0 g of ¹⁵NH_4Cl and 2.0 g D-glucose monohydrate. Add 1.5 ml of 1 M $MgSO_4$, 3.0 ml of MEM vitamin solution, 2.0 ml of Q-solution and 0.15 ml 1 M $CaCl_2$. Dissolve all the reagents in 800 ml of deionized water for a final volume of 1 liter. Work under sterile conditions. Do not autoclave. Adjust the pH to 7.4 with NaOH. M9-medium can be stored at 4 °C for 3 weeks.

PBS buffer (10×) For 1 liter, use 80 g of NaCl, 2 g of KCl, 14 g of $Na_2HPO_4·2H_2O$ and 2 g of KH_2PO_4. Dissolve the reagents in deionized water for a final volume of 1 liter and autoclave. Buffer can be stored at room temperature for 1 year. Dilute 10× PBS tenfold before use.

Q-solution stock solution For 1 liter, weigh 5 g of $FeCl_2·4H_2O$, 184 mg of H_3BO_3, 64 mg of $CoCl_2·6H_2O$, 18 mg of $CuCl_2·H_2O$, 4 mg of $ZnCl_2$, 605 mg of $Na_2MoO_4·2H_2O$, 40 mg of $MnCl_2·4H_2O$ and use 8 ml 5 M HCl. Dissolve all the reagents in deionized water and bring to a final volume of 1 liter. Q-solution can be stored at room temperature for 1 year.

Running buffer for agarose gel electrophoresis (50× TAE buffer) For 1 liter, take 242 g of Tris, 57.1 ml of glacial acetic acid and 100 ml of 0.5 M EDTA solution (pH 8.0). Dissolve all the reagents in deionized water for a final volume of 1 liter and autoclave. 50× TAE buffer can be stored at room temperature for 1 year. Dilute the 50× TAE buffer 50-fold with distilled water before use.

SDS-PAGE sample buffer (2×, pH 6.8) To prepare 20 ml of solution, use 5.0 ml of 0.5 M Tris-HCl, 8 ml of 10% (wt/vol) SDS, 4 ml of glycerol, 1 ml of mercaptoethanol and 0.4 ml of 0.5% (wt/vol) bromophenol blue. Dissolve all the reagents in deionized water for a final volume of 20 ml. Dilute in deionized water to 1× before use. Sample buffer can be stored at room temperature for 1 year.

SDS-PAGE resolving buffer For 1 liter, weigh 181.5 g of Tris and 4 g of SDS. Dissolve reagents in 800 ml of deionized water and adjust pH to 8.8 by adding HCl. Then bring the final volume to 1 liter. Store at room temperature for up to 1 year.

SDS-PAGE stacking buffer For 1 liter, weigh 60.5 g of Tris and 4 g of SDS. Dissolve reagents in 800 ml of deionized water and adjust pH to 6.8 by adding HCl. Then bring the final volume to 1 liter. Store at room temperature for up to 1 year.

SOB medium For 1 liter, use 20 g of tryptone, 5 g of yeast extract, 0.58 g of NaCl, 0.19 g of KCl, 2.03 g of $MgCl_2·6H_2O$ and 2.46 g of $MgSO_4·7H_2O$. Dissolve all reagents in 800 ml of deionized water, adjust pH to 7.4 by adding NaOH and then add water for a final volume of 1 liter and autoclave. SOB medium can be stored at 4 °C for 1 month.

Wash buffer (1×) Take 100 ml of 0.5 M sodium phosphate buffer (pH 8.0) and add 60 ml of 5 M NaCl and 50 ml of 0.5 M imidazole, for their final concentrations of 50 mM, 300 mM and 25 mM, respectively. Check pH and adjust it to 8.0 if necessary with NaOH. Add deionized water to make a final volume of 1 liter and filter through 0.22 μm syringe filters. Wash buffer can be stored at room temperature for 1 year.

SDS-PAGE gels (18%) For the preparation of stacking/concentrating and resolving gels, use respective buffers prepared beforehand (see above). To prepare four resolving gels mix 8.1 ml of acrylamide/bisacrylamide (37:1 ratio), 4.5 ml of resolving buffer and 6.9 ml of deionized H_2O. Finally add 0.1 ml of 10% (wt/vol) ammonium persulfate and 0.015 ml of TEMED.

Load the mixture immediately between gel plates and cover with butanol to level the top of it and remove any bubbles/foam. Wait for over 30 min, until the gel is polymerized. Then remove butanol completely by tipping the gel upside down and soaking it into filter paper or paper towel.

To prepare four stacking gels take 0.8 ml of acrylamide/bisacrylamide (37:1 ratio), 1.5 ml of stacking buffer and 3.7 ml of deionized H_2O. Finally add 0.02 ml 10% (wt/vol) ammonium persulfate and 0.015 ml TEMED. Immediately pour the mixture between the gel plates on the top of resolving gel. Fill to the top and then insert gel-well combs.

When the stacking gel is polymerized, remove the combs. SDS-gels can be used immediately or wrapped with wet and clean paper towel and kept in plastic bags at 4 °C for up to 2 weeks.

PROCEDURE
In vivo segmental isotopic labeling

Dual-plasmid transformation for protein ligation ● **TIMING** 1 d

1| Prepare the two plasmids, each bearing one of the two precursor proteins for the dual expression (see Design of plasmids).

2| To transform the cells, transfer 0.5 μl (50–100 ng) of each of the two constructed plasmids into a microcentrifuge tube kept on ice and add 50 μl of competent *E. coli* ER2566 cells. (The chemically prepared competent cells can be purchased from NEB as 'T7 Express Competent').

3| Incubate the tube on ice for 30 min.

4| Heat shock cells for 45 s at 42 °C on a heating block and return them onto ice for 30 s.

5| Add 450 μl of pre-warmed (room temperature) SOB medium to the transformed cells on ice.

6| Place the tube on a bench top shaker for 1 h, set for 37 °C and at 300 r.p.m.

7| Plate 200 μl of the cell suspension from the tube onto an LB-agar plate supplemented with the appropriate antibiotics (kanamycin and ampicillin, in this case) and incubate at 37 °C overnight (12–14 h).

It should be noted that before plating the cell suspension onto agar plates, it may be necessary to spin down the cells to obtain dual-transformed cells with the two plasmids. For this purpose, spin cell suspension (after Step 6) gently down

(<1,000g, 1 min at room temperature) and remove upto 400 µl of the supernatant. Gently resuspend the cells in the remaining supernatant and then transfer the suspension onto an agar plate.

Control expression test ● TIMING 1 d

8| Pick colonies from the plate incubated overnight and inoculate 5 ml of LB medium in a 14 ml culture tube containing both kanamycin (25 µg ml^{-1}) and ampicillin (100 µg ml^{-1}).
It is to be noted that more than one colony can be used for inoculation to shorten the growth time of the cell culture.

9| Culture the cells in a shaking incubator (37 °C, 250 r.p.m., until the OD$_{600}$ reaches a value of 0.5–0.7).
It should be noted that it may be useful to adjust the incubation temperature, depending on the POI.

10| Monitor the optical density of the cell culture at 600 nm (OD$_{600}$) every 30–60 min by taking 2 µl of the cell culture and measuring OD$_{600}$ with NanoDrop UV spectrometer.

11| When the OD$_{600}$ reaches a value between 0.5–0.7 take the first 0.5 ml sample from the culture tubes and transfer it into a 1.7 ml microcentrifuge tube. Use this aliquot for SDS-PAGE analysis (see sample preparation in **Box 1**).
▲ CRITICAL STEP To prevent possible protein degradation of the sample, it is recommended to prepare SDS-PAGE samples immediately.

12| Once the OD$_{600}$ reaches a value between 0.5–0.7, induce the primary protein expression with a final concentration of 0.04% (wt/vol) L-arabinose.
▲ CRITICAL STEP Owing to the nature of arabinose and T7 promoters, it is recommended to induce first with arabinose, followed by the induction with IPTG for the second precursor protein. It should be noted that the optimal order of induction and concentration of each inducer, however, may need to be determined empirically by maximizing the amount of the ligated product on the SDS-gel. It is recommended to test the following final concentrations of inducers: 0.02, 0.04, 0.1 and 0.2% (wt/vol) of L-arabinose and 0.2 mM and 0.5 mM of IPTG.

13| To monitor the expression of the first precursor protein, take a 0.5 ml sample from the culture tube after 30 min. Prepare the SDS-PAGE sample as described in **Box 1**.

14| Induce the second precursor protein with a final concentration of 0.5 mM IPTG after 30 min of the first induction.

15| Take a few samples for SDS-PAGE analysis (once every hour) during 3–5 h from the first induction. Prepare SDS-PAGE samples as described in **Box 1**.

16| Heat the samples collected from Steps 11–15 for 5 min, at 95 °C. Load 5 µl of each sample onto 18% SDS-PAGE gel.

17| Analyze the results from SDS-PAGE to evaluate the successful protein ligation (see **Figs. 4** and **5a** for the interpretation of results). If the result from Steps 1–17 was satisfactory, move on to a large-scale production for segmental isotopic labeling.
? TROUBLESHOOTING

Large-scale segmental isotopic labeling ● TIMING 2–3 d

18| Repeat Steps 1–7 if necessary. These steps may not have to be repeated, if there are colonies available from the original transformation after Step 8 and if those colonies are not older than 1 week. (The plate with transformed colonies from Step 7 can be stored at 4 °C for upto 1 week.) If these steps can be omitted, then timing for segmentally labeled protein expression (Steps 19–40) is reduced to 2 d.

19| Pick 1–10 colonies from the plate and use them to inoculate 5 ml of LB medium in a 14 ml culture tube containing both the appropriate antibiotics (25 µg ml^{-1} of kanamycin and 100 µg ml^{-1} of ampicillin).

20| Culture the cells in a shaking incubator for 2–3 h (37 °C, 250 r.p.m.).

21| Transfer 2–3 ml of the culture into a 250 ml flask with baffles containing 50 ml of M9 medium and the appropriate antibiotics (25 µg ml^{-1} of kanamycin and 100 µg ml^{-1} of ampicillin).

22| Culture the cells overnight in a shaking incubator (37 °C, 250 r.p.m.).

23| Next morning, transfer 25 ml of the overnight culture into two 2,000 ml flasks with baffles, each containing 750 ml of pre-warmed (37 °C) ^{15}N-labeled M9 medium supplemented with the appropriate antibiotics (25 µg ml^{-1} of kanamycin and 100 µg ml^{-1} of ampicillin) and ^{15}N-labeled ammonium chloride as the sole nitrogen source.

24| Culture the cells in a shaking incubator (37 °C, 250 r.p.m., until the OD_{600} reaches a value of 0.5–0.7).

25| Monitor the OD_{600} every 30–60 min as described in Step 10.

26| When the OD_{600} reaches a value between 0.5–0.7, add the appropriate amount of the first inducer (typically a final concentration of 0.1% (wt/vol) L-arabinose, unless different optimal concentration or order of inductions is favored).
   It is to be noted that different concentrations of IPTG usually work equally well both on small and on large scale. However, as D-glucose could inhibit the induction by L-arabinose, it is recommended to use higher (upto twofold) concentration of L-arabinose, than the one that has been established as efficient and optimal in a small-scale experiment.

27| Culture the cells for 3 h in a shaking incubator (37 °C, 250 r.p.m.).
   It should be noted that the cultivation temperature and length may need to be altered depending on the POI. Optimal conditions can be determined empirically.

28| To check the induction of protein expression by SDS-PAGE analysis, take and prepare samples from the culture every 30–60 min (as described in **Box 1**).

29| During the first induction period, pre-warm the second induction medium (1.5 liter of LB medium for incubation, plus 400–500 ml more for a washing step and resuspension) at 37 °C. Add the second inducer and antibiotic (i.e., IPTG, 0.5 mM, unless determined otherwise, and kanamycin, 25 µg ml^{-1}) to the medium in advance.

30| Divide and transfer the pre-warmed LB medium containing the second inducer and antibiotics prepared in Step 29 into three 2,000 ml flasks (500 ml per each). Keep flasks and their content at 37 °C until the second induction.

31| Prepare four sterile 500 ml centrifuge tubes for centrifuging the cells (to minimize the time required for replacing the medium).

32| After the first induction incubation, divide and transfer the M9 culture into four 500 ml centrifuge tubes as fast as possible.

33| Centrifuge the cells for 10 min at 900*g*, 25 °C.

34| Gently decant the supernatant and add 50–100 ml of the pre-warmed LB medium (prepared in Step 29) to each bottle and resuspend the pellet by gently tapping the centrifuge tubes.

35| Centrifuge the cells once more for 10 min at 900*g*, 25 °C.

36| Gently decant the supernatant and add ~50 ml of the pre-warmed LB medium containing the inducer and the antibiotics to each centrifuge tube. Resuspend the pellets gently by tapping the centrifuge tubes. Divide and transfer the cell suspensions from the centrifuge tubes into the flasks prepared in Step 30.
   ▲ **CRITICAL STEP** Keep the total time for Steps 32–36 within 30–40 min. Add the inducer in the washing medium.

37| Culture the cells for 3 h in a shaking incubator (37 °C, 250 r.p.m.). Continue taking and preparing SDS-PAGE samples as described in Step 11 and **Box 1**.
   It is to be noted that the cultivation temperature and length may need to be adjusted depending on the POI. An optimal condition can be determined empirically.
   ? TROUBLESHOOTING

38| Add 100 µl of antifoam to each flask and transfer the medium into two 1,000 ml centrifuge tubes for centrifugation.

39| Centrifuge the cells for 10 min at 9,000*g*, 4 °C.

40| Discard the supernatant and resuspend the pellets with 20–30 ml of lysis and binding buffer.
   ■ **PAUSE POINT** The suspended cells can be stored at –80 °C in a 50 ml Falcon tube for further purification.

41| Purify the protein according to the appropriate procedure preferred for the specific ligated product (e.g., IMAC, ion-exchange chromatography and/or gel filtration).

● **TIMING**
In vivo **segmental isotopic labeling: 7–8 d**
Step 1, Plasmid construction: 3 d
Steps 2–7, Dual-plasmid transformation: 1 d
Steps 8–17, Control expression test: 1 d
Steps 18–41, Large-scale segmental isotopic labeling: 2–3 d

Box 2 *In vitro* **segmental isotopic labeling: 10–13 d**
Step 1, Plasmid construction: 3 d
Steps 2–7, Separate plasmid transformation: 1 d
Step 8 (option A or B), Expression of individual protein segments: 2 d
Steps 9–11, Individual protein-precursor purification and dialysis: 2–3 d
Steps 12–16, Protein precursor ligation: 1–2 d
Steps 17–18, Segmentally labeled protein dialysis and purification: 1–2 d

? **TROUBLESHOOTING**
Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
17	No protein ligation is observed in the control expression test	Precursor fragments are insoluble	Change the temperature for protein expression. For example, lowering the temperature might increase the amounts of soluble precursor proteins
			Check the solubility of the individual precursor proteins. If they are insoluble, consider making a fusion protein for improving the solubility. At least the first protein induced has to be soluble in the cells[12]
			Try co-expression of chaperons for improving proper folding of the proteins (e.g., the chaperone plasmid set from Takara Bio)[27]
			Check the order of the inducers and their final concentrations, which could affect the ligation efficiency significantly
			Consider replacing the split intein with other inteins[11,22]
			Purify the insoluble precursor under denaturing conditions and carry out protein ligation *in vitro* by refolding
		Cleavage reactions are dominating the protein-ligation reaction	Check the amount of the spliced Int_N fragment (~15 kDa for NpuDnaE-Int_N)
			Change the primary sequences near the splicing junctions or the split site of the protein of interest
			Try other split inteins[11,22]
37	Protein ligation worked well in the control experiment, but a poor yield was obtained from the segmental isotopic labeling	The precursor proteins (particularly for the primarily induced protein) are unstable *in vivo*	Optimize the duration for the expression of the individual precursor proteins Consider adding a fusion tag to increase protein stability *in vivo* (**Fig. 3c**)[19,20]
			Carry out protein ligation *in vitro*. See **Box 2**
		Imbalance of the molar ratio between the N- and C-terminal precursors	Optimize the inducer concentrations using the M9 medium that is used for the large-scale expression

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
	Lower selectivity of the isotope incorporation (high cross labeling)	Leaky expression during the first expression	Use a plasmid such as pSKDuet01 containing *lacI* gene to suppress leaky expression of T7 promoter
			Consider using pLysS or pLysE plasmids to suppress leaky expression by T7 RNA polymerase[28,29]
		Imperfect removal of the first inducer	Make sure that the first inducer is removed completely during the washing step (consider an additional washing step)[11]
			Make sure to add the second inducer in the washing buffer for immediate protein induction upon replacing the medium
			Carry out protein ligation *in vitro*. Follow the protocol of *in vitro* ligation (**Box 2**)
Box 2			
8A(vi) and 8B(vi)	Precursor protein is not expressed upon induction		Improve the expression by changing vectors, promoters, and so on. Try tRNA-supplemented host strains such as Rosetta (Novagen)[30]
11	Precursor protein precipitates during dialysis		Add solubility enhancement tag[19,20]
16	No protein ligation is observed		Use other split inteins or change the split sites[11,22,31]

ANTICIPATED RESULTS
Expression test
In the control expressions, the protein ligation of the two split fragments by PTS has to be confirmed before proceeding to segmental isotopic labeling. **Figures 4** and **5** show the examples of c-CrkII adaptor protein and AlgE4 epimerase[11,32]. If protein ligation of the target fragments is successful, SDS-PAGE analysis should identify additional bands that emerged by PTS, i.e., the ligated product (POI_N–POI_C in **Fig. 4a** and **A-R** in **Fig. 5**) and the split-intein fragment (Int_N in **Fig. 4a**). The ligated product can be easily identified if POI_C is bigger than Int_N because a new additional band will appear with an increased apparent molecular weight (for instance, in the case of AlgE4 epimerase, A-R in **Fig. 5**). In **Figures 4b** and **5a**, possible scenarios of the protein-ligation reaction are illustrated. In the worst-case scenario, the dual expression results in a mere overlay of the two separate expressions by individual inducers (lane c in **Fig. 4b**, lane c in **Fig. 5a**). In the best-case scenario, the ligated product and cleaved Int_N would be the main product upon dual expression (lanes d in **Figures 4b** and **5a**). Because it is not trivial to adjust the expression levels of both the precursors equally, there is usually an excess of one of the precursor proteins. In an ideal case, a ligation product band will appear dominantly in the SDS-PAGE analysis. When N-terminal cleavage is the major reaction, cleaved products (Int_N and Ext_N) will appear (lane e in **Fig. 5a**). If the cleaved Int_N fragment was observed but no ligated product could be identified, it would suggest that cleavage reactions are dominating the PTS reaction. In this case, it is advisable to introduce some changes at the splicing junctions (see also TROUBLESHOOTING)[17]. In some cases, there could be a mixture of N-cleavage and ligation reactions (lane f in **Fig. 5a**). It is also possible that all N-, C-cleavage and ligation reactions occur at the same time, thereby producing many new bands in SDS-PAGE analysis (lane g in **Fig. 5a**). It might be advisable to test different orders and concentrations of the two inducers when one cannot observe any difference upon the dual expression (lane a–c in **Fig. 5a**). One example of the ligation between A and R domains in AlgE4 epimerase is shown in **Fig. 5b** and **c**. When the N-terminal precursor was first induced by 0.5 mM IPTG and followed by the second induction with L-arabinose, the ligated product was hardly observable (A-R in **Fig. 5c**). In contrast, when the C-terminal precursor was first induced by the addition of arabinose and followed by the second induction with 0.5 mM IPTG, the ligated product was predominantly observed with an optimal concentration between 0.02% and 0.05% arabinose (indicated by A-R in **Fig. 5a**). This is because T7 promoter dominates the arabinose promoter. The solubility of the precursor proteins could also affect the ligation and lead to a similar result. Thus, the order of the two inducers could be a critical factor to achieve high yields of the ligated products. Moreover, it is noteworthy that protein ligation by PTS with naturally split inteins is dependent not only on the splicing junction sequences[17] but also on the target fragments.[22] At present, all the factors influencing protein ligation by PTS still remain elusive[17,22]. Obtaining a good yield of the protein ligation in control expression tests is therefore essential for segmental isotopic labeling.

Segmental isotopic labeling
The yield obtained from the large-scale segmental isotopic labeling *in vivo* is usually lower than that of the control expression test because of the removal of the first inducer during the second expression. Once *in vivo* protein ligation is properly optimized,

the typical yields of the ligated protein after purification can be 2–50 mg l⁻¹, depending on the proteins and purification procedures involved. From segmentally isotope-labeled samples, significant suppression of NMR signals is expected for the region expressed in unlabeled medium, which can be immediately used for the assignments of individual peaks to the region expressed in the labeled medium (**Fig. 6**). There might be a small portion (<5%) of residual signals that can be observed because of isotopic scrambling during the second expression[11,12]. If the residual signals are greater than 5%, this might be because of the leaky expression during the first or second expression (see TROUBLESHOOTING). When this cross labeling caused by the *in vivo* preparation deteriorates NMR experiments, such as isotope-filtering experiments, *in vitro* approach might be more suitable. We also noticed that these residual signals become non-negligible when NMR signals originating from very flexible parts of proteins are significantly stronger than the rest of the protein. It is also noteworthy that one should not expect a better quality of NMR spectra from the segmentally isotope-labeled samples than the original spectra but rather a reduced number of signals. Segmental isotopic labeling is a method to reduce only overlap of NMR signals and cannot improve the intrinsic quality of NMR spectra such as line-widths. The *in vitro* method may result in lower yields and requires several additional purification steps. However, the advantage of the *in vitro* approach is that the bottlenecks of protein ligation can be analyzed by testing and monitoring the reactions with different conditions[22]. More importantly, cross labeling between the N- and C-terminal fragments can be prevented.

Here we described *in vivo* and *in vitro* procedures for the productions of segmentally isotope-labeled proteins by making use of PTS with naturally split DnaE intein, which does not require any refolding steps. These procedures can generally be used for segmental isotopic labeling of multi-domain proteins composed of self-contained globular domains[11] and fusion proteins[12]. Recently, we also demonstrated that these procedures can be extended to segmental isotopic labeling of a central domain in a multi-domain protein, opening a new avenue to label an arbitral region in multi-domain proteins with the same protocol described in this article[33]. We anticipate that our simple and robust protocols will increasingly become an important tool for NMR studies of challenging proteins, such as large multi-domain proteins, proteins with low solubility and proteins with repeating sequences.
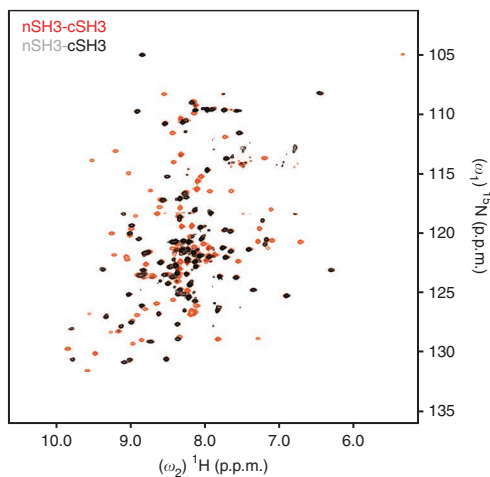

Figure 6 | HSQC spectra from one example. An overlay of two dimensional [¹⁵N, ¹H]-HSQC spectra of fully ¹⁵N-labeled and segmentally ¹⁵N-labeled protein of c-CRKII adaptor protein (131–304) prepared by the *in vivo* segmental labeling procedure[11]. The spectrum from the fully ¹⁵N-labeled sample is colored in red. The spectrum of the C-terminally segmental labeled sample is colored in black, demonstrating the instant assignments of the peaks to the C-terminal domain.

1. Riek, R., Pervushin, K. & Wüthrich, K. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends Biochem. Sci.* **25**, 462–468 (2000).
2. Wider, G. & Wüthrich, K. NMR spectroscopy of large molecules and multi-molecular assemblies in solution. *Curr. Opin. Struct. Biol.* **9**, 594–601 (1999).
3. Ohki, S. & Kainosho, M. Stable isotope-labeling methods for protein NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **53**, 208–226 (2008).
4. Kainosho, M. & Tsuji, T. Assignment of the three methionyl carbonyl carbon resonances in *Streptomyces* subtilisin inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. *Biochemistry* **21**, 6273–6279 (1982).
5. Bax, A. Multidimensional nuclear-magnetic-resonance methods for protein studies. *Curr. Opin. Struct. Biol.* **4**, 738–744 (1994).
6. Yamazaki, T. *et al.* Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* **120**, 5591–5592 (1998).
7. Xu, R., Ayers, B., Cowburn, D. & Muir, T.W. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. USA* **96**, 388–393 (1999).
8. Iwai, H. & Züger, S. Protein ligation: applications in NMR studies of proteins. *Biotechnol. Genet. Eng. Rev.* **24**, 129–146 (2007).
9. Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. & Kyogoku, Y. Improved segmental isotope labeling of proteins and application to a larger protein. *J. Biomol. NMR* **14**, 105–114 (1999).
10. Yagi, H., Tsujimoto, T., Yamazaki, T., Yoshida, M. & Akutsu, H. Conformational change of H+-ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 16632–16638 (2004).
11. Muona, M., Aranko, A.S. & Iwai, H. Segmental isotopic labelling of a multi-domain protein by protein ligation using protein *trans*-splicing. *Chembiochem* **9**, 2958–2961 (2008).
12. Züger, S. & Iwai, H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat. Biotech.* **23**, 736–740 (2005).
13. Wu, H., Hu, Z.M. & Liu, X.Q. Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226–9231 (1998).
14. Paulus, H. Protein splicing and related forms of protein autoprocessing. *Ann. Rev. Biochem.* **69**, 447–496 (2000).
15. Evans, T.C. Jr., Benner, J. & Xu, M.Q. Semisynthesis of cytotoxic proteins using a modified protein-splicing element. *Protein Sci.* **7**, 2256–64 (1998).
16. Skrisovska, L. & Allain, F.H. Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. *J. Mol. Biol.* **375**, 151–164 (2008).

17. Iwai, H., Züger, S., Jin, J. & Tam, P.H. Highly efficient protein *trans*-splicing by a naturally occurring split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* **580**, 1853–1858 (2006).

18. Zettler, J., Schütz, V. & Mootz, H.D. The naturally split *Npu*DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).

19. Nallamsetty, S. & Waugh, D.S. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **45**, 175–182 (2006).

20. Zhou, P., Lugovskoy, A.A. & Wagner, G. A solubility enhancement tag (SET) for NMR studies of poorly behaving proteins. *J. Biomol. NMR* **20**, 11–14 (2001).

21. Serber, Z. & Dötsch, V. In-cell NMR spectroscopy. *Biochemistry* **40**, 14317–14323 (2001).

22. Aranko, A.S., Züger, S., Buchinger, E. & Iwaï, H. *In vivo* and *in vitro* protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS ONE* **4**, e5185 (2009).

23. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

24. Dillon, P.J. & Rosen, C.A. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques* **9**, 298–300 (1990).

25. Han, J.C. & Han, G.Y. A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal. Biochem.* **220**, 5–10 (1994).

26. Getz, E.B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P.R. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* **273**, 73–80 (1999).

27. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**, 1694–1699 (1998).

28. Studier, F.W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**, 37–44 (1991).

29. Zhang, X. & Studier, F.W. Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *J. Mol. Biol.* **269**, 10–27 (1997).

30. Brinkmann, U., Mattes, R.E. & Buckel, P. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the DnaY gene product. *Gene* **85**, 109–114 (1989).

31. Oeemig, J.S., Aranko, A.S., Djupsjöbacka, J., Heinämäki, K. & Iwaï, H. Solution structure of DnaE intein from *Nostoc punctiforme*: structural basis for the design of a new split intein suitable for site-specific chemical modification. *FEBS Lett.* **583**, 1451–1456 (2009).

32. Aachmann, F.L., Svanem, B.I., Güntert, P., Petersen, S.B., Valla, S. & Wimmer, R. NMR structure of the R-module: a parallel beta-roll subunit from an *Azotobacter vinelandii* mannuronan C-5 epimerase. *J. Biol. Chem.* **281**, 7350–6 (2006).

33. Busche, A.E.L., Aranko, A.S., Talebzadeh-Farooji, M., Bernhard, F., Dötsch, V. & Iwaï, H. Segmental isotopic labelling of a central domain in a multi-domain protein by the use of only one robust DnaE intein. *Angew. Chem. Int. Ed.* **48**, 6128–6131 (2009).