

Supporting Information

Photo-triggered fluorescent labelling of recombinant proteins in live cells

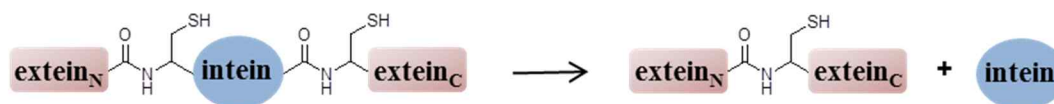
Deokho Jung¹, Kohei Sato², Kyoungmi Min¹, Akira Shigenaga², Juyeon Jung³, Akira Otaka^{2*}, and Youngeun Kwon^{1*}

¹ Department of Biomedical Engineering, Dongguk University-Seoul, Pildong 3-ga, Seoul, Korea

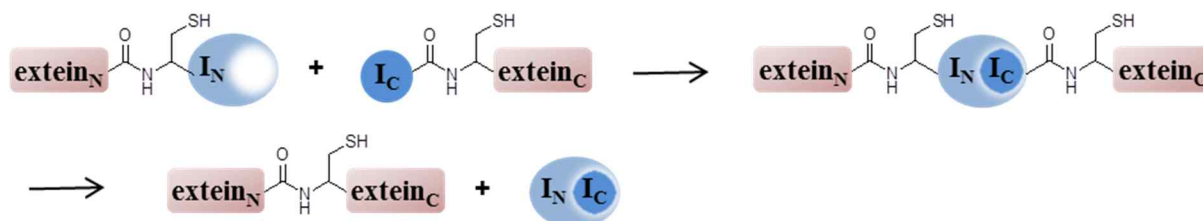
² Institute of Health Biosciences and Graduate School of Pharmaceutical Sciences, The University of Tokushima, 1-78-1 Shomachi, Tokushima, Japan

³ BioNano Health Guard Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong-gu, Daejeon, Korea

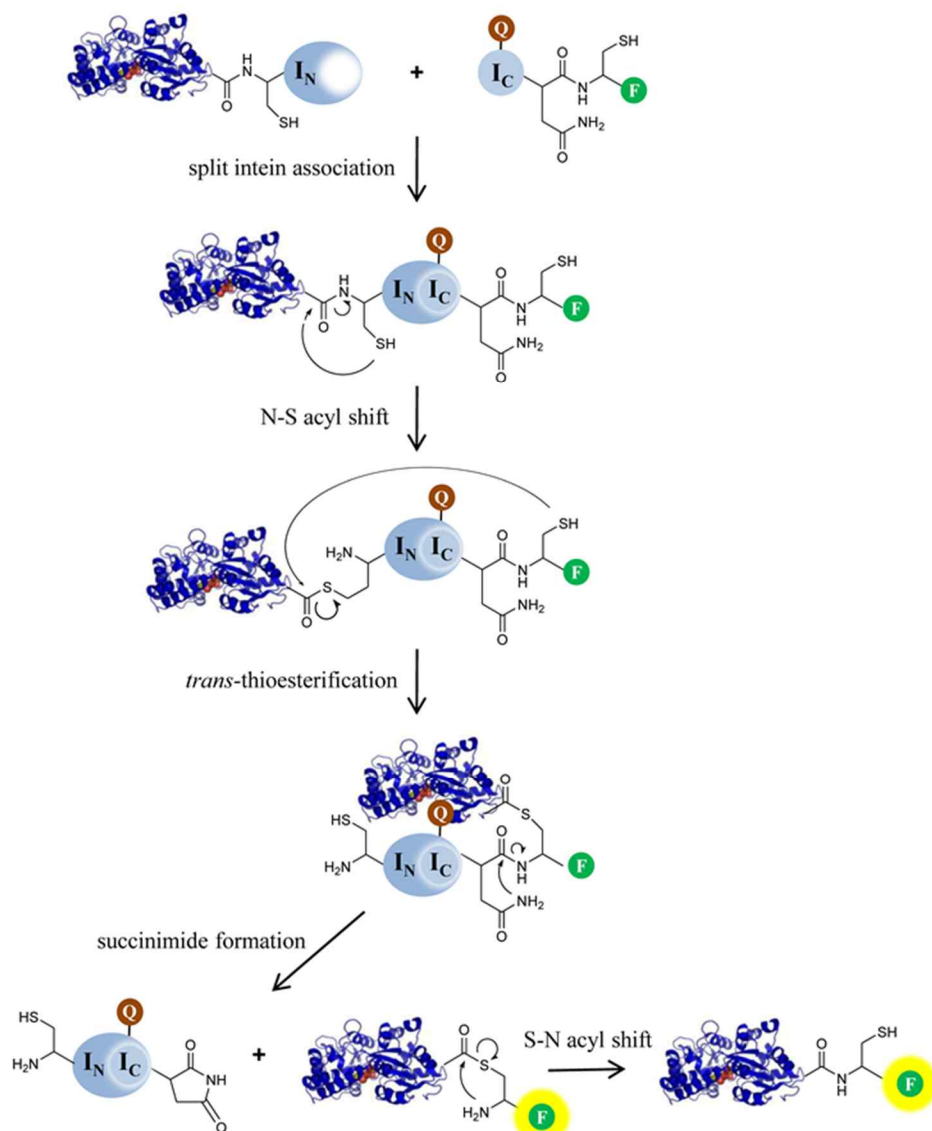
(A) Protein Splicing



(B) Protein Trans-Splicing

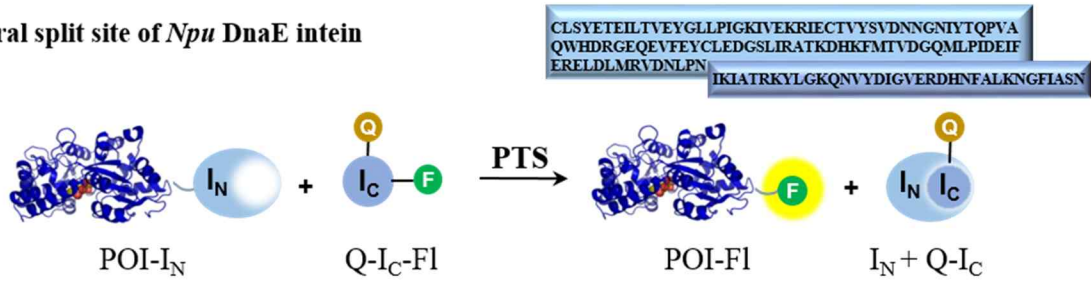


Scheme S1. Mechanisms of intein mediated protein splicing and split-intein mediated protein trans-splicing (PTS). (A) In the protein splicing reaction, inteins splice themselves off while conjugating N- and C-exteins through an amide bond. (B) A pair of split-inteins binds to each other to form a functionally active domain that can mediate the PTS reaction.

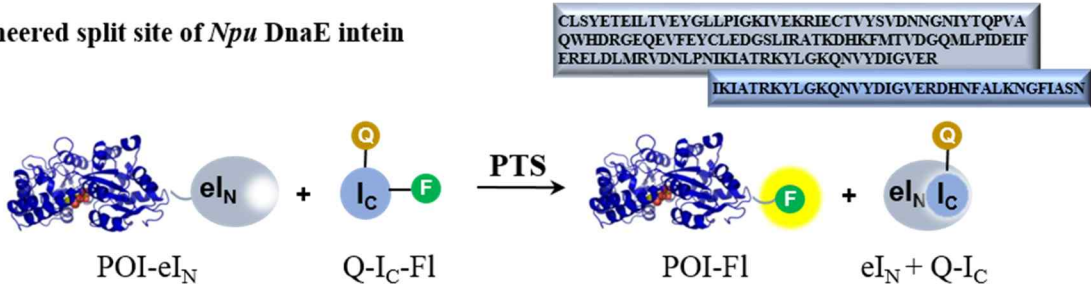


Scheme S2. Mechanism of PTS based fluorescent labeling Two split fragments of an intein, I_N and I_C , fold together to form a functionally active domain. The trans-splicing process begins with an N to S acyl shift to generate a thioester bond, which is labile to nucleophilic attacks. The thioester bond is attacked by the SH-group of the first Cys residue of C-extein resulting in *trans*-thioesterification. The following cyclization of the conserved Asn residue at the C-terminus of the intein releases the intein domain with a concomitant conjugation of exteins by a thioester bond. The final S to N acyl shift generates a seamless conjugated product.

• Natural split site of *Npu* DnaE intein



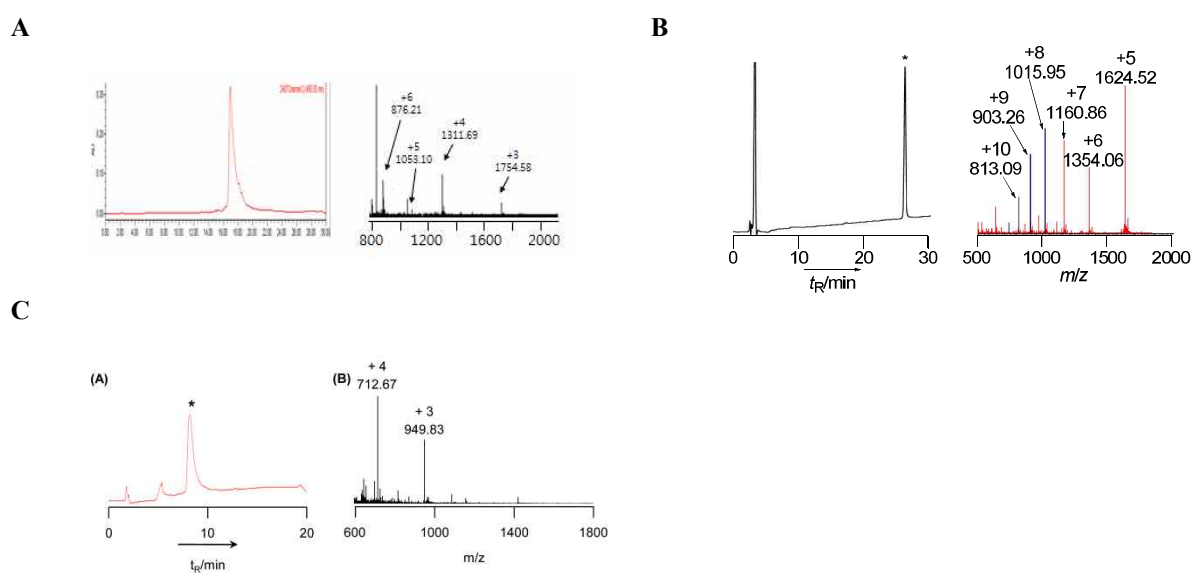
• Engineered split site of *Npu* DnaE intein



Scheme 3. PTS can be mediated by a pair of naturally split-inteins or a mixed pair of split-inteins.

Table S1. List of peptides used for experiments

Label	Sequence	M.W	
		Expected	Found
Q-Ic-F1 1	DabcyI-MIKIATRKYLGKQNVYDIGVERDHNFKLN GFIASNCFNK(FI)-NH ₂	5255.5	5257.4
Q-iso-Ic-CPP-F1 2	Dabsyl-MIKIATRKYLGKQNVYDIGVERDHNFKLN G F I A -iso S (H-R _s G-Spr(<i>o</i> NB)- N C F N K (F I)GGDYK DDDKG-NH ₂	8121.1	8120.5
Pep-1 5	H-KETWWETWWTEWSQPKKKRKV-NH ₂	2848.2	2846.7

**Figure S1.** Analytical HPLC and ES-TOF analysis data of purified Q-Ic-F1 1 (A), Q-iso-Ic-CPP-F1 2 (B), Pep-1 5 (C).

Materials

All amino acid derivatives and resins were purchased from NovaBiochem (Merck, Darmstadt, Germany), Watanabe Chemical Industries (Hiroshima, Japan), and AnaSpec (Freemont, CA, USA). General chemicals were of the best grade available and were supplied by Sigma-Aldrich (St. Louis, MO, USA), Kanto Chemicals (Tokyo, Japan), and Fisher-Scientific (Pittsburgh, PA, USA). Restriction enzymes were purchased from Elpis-Biotech (Daejeon, Korea) and New England Biolabs (Ipswich, MA, USA). The transfection reagent was purchased from Polyplus-transfection SA (Strasbourg, France). Dulbecco's modified Eagle's medium (DMEM) and fetal bovine serum (FBS) were purchased from PAA Laboratories Inc. (Pasching, Austria). Analytical HPLC was performed on a Waters 2796 Bioseparations Module instrument with a Waters 2487 Dual λ absorbance detector (Waters Corp., Bedford, MA, USA), Waters e2695 Separations Module with a Waters 2489 UV/visible detector (Waters Corp., Bedford, MA, USA) or a Lachrom Elite with a UV detector L-2400 (Hitachi High-Tech, Tokyo, Japan) using Vydac C8 and C18 columns (5 μ m, 4.6 \times 150 mm) (Vydac, Hesperia, CA, USA) or a Cosmosil 5C₁₈-AR-II analytical column (Nacalai Tesque, 4.6 \times 250 mm, flow rate 1.0 mL/min) at a flow rate of 1 ml/min. All runs used linear gradients of 0.1% aqueous formic acid (solvent A) vs. 90% acetonitrile (MeCN) in H₂O with 0.1% formic acid (solvent B) or 0.1% aqueous TFA (solvent C) vs. MeCN with 0.1% TFA (solvent D). Purified peptides were analyzed by electrospray mass spectrometry (ES-MS) using an Agilent 6460 Triple Quadrupole MS (Agilent Technologies, Santa Clara, CA, USA). Calculated masses were obtained using ChemDraw 10 (Cambridgesoft, Waltham, MA, USA). Protein samples were analyzed on 12% polyacrylamide gels and stained with Coomassie Brilliant Blue R or imaged using illuminator to detect fluorescent signal. Digitalized gel images were quantified using NIH Image-J software (ver. 1.45s, <http://rsb.info.nih.gov/ij/>). Confocal fluorescence images were obtained with an Eclipse Ti (Nikon Instruments, Tokyo, Japan) and solution fluorescent intensity was measured using Infinite 200 pro (Tecan, Grodig, Austria). Fluorescein fluorescence was excited with the 488 nm light and the emission was captured using 515/530 nm filter.

Solid Phase Peptide Synthesis

General procedure for peptide synthesis Peptides were synthesized on a 0.5 mmol scale on Rink Amide AM resin LL according to a general Fmoc solid phase peptide synthesis (SPPS)

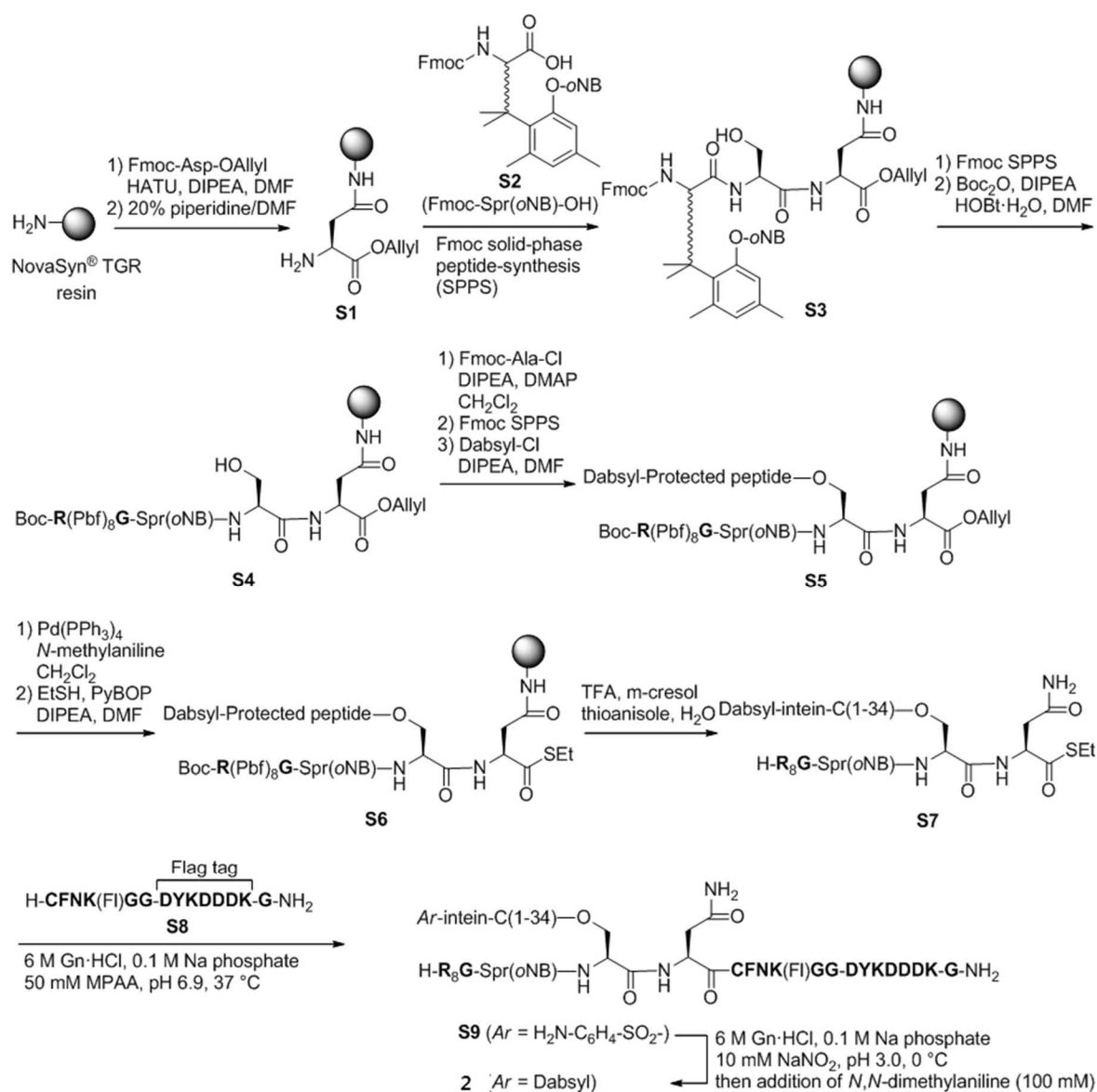
protocol. The *O*-Benzotriazole-*N,N,N',N'*-tetramethyluroniumhexafluorophosphate (HBTU) was used for activation of protected amino acids. Fmoc-protected amino acids with acid labile side chain protecting groups were used for the synthesis except for ϵ -Lys residue which were incorporated as Fmoc-Lys(Dde)-OH for orthogonal deprotection of ϵ -amine functional group. Dde was removed by treating with 2% hydrazine in DMF for 3 min four times after the chain assembly. Then the DMF solution of FITC (4 equiv.) and DIPEA (8 equiv.) was added to the resin and the coupling reaction was carried out for 3 h in dark. Assembled peptides were cleaved from the resins and the side chain functional groups were deprotected by treating with cleavage cocktail, i.e., TFA: EtSH: H₂O: triisopropylsilane (TIS) = 90:5:2.5:2.5 v/v), for 3 h at room temperature. The resin was filtered, and the peptides were precipitated in cold Et₂O. The crude materials were dissolved in a minimal amount of H₂O: MeCN (2:1 v/v) and purified by HPLC.

Preparation of *O*-iso-Ic-CPP-FI 2 A general scheme for the synthesis of iso-peptide is shown in Fig. S2.

Peptide elongation: On NovaSyn[®] TGR resin (Rink amide type: 0.20 mmol amine/g, 1.8 g, 0.36 mmol) was coupled Fmoc-Asp-OAllyl (2 equiv.) with the aid of *O*-(7-azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU, 2 equiv.) and *N,N*-diisopropylethylamine (DIPEA, 2 equiv.) in DMF for 2 h followed by Fmoc removal by 20% (v/v) piperidine/DMF to give Asp-OAllyl-incorporated resin S1. The resulting resin S1 was treated with Fmoc-Ser-OH (4 equiv.), *N,N'*-diisopropylcarbodiimide (DIPCDI, 4.1 equiv.) and HOBT·H₂O (4 equiv.) followed by Fmoc removal to yield the side-chain-unprotected Ser-linked resin. Then, Fmoc-protected Spr(oNB) S2 (2 equiv. racemic materials) was attached to the resin by the use of HATU (2 equiv.) and DIPEA (2 equiv.) in DMF for 2 h to give the peptide resin S3. On this resin, standard Fmoc SPPS (coupling: 4.0 or 6.0 equiv. each of amino acid using DIPCDI (4.1 or 6.1 equiv.) and HOBT·H₂O (4.0 or 6.0 equiv.) in DMF (2 h); Fmoc removal: 20% (v/v) piperidine/DMF (10 min)) were performed for elongation of the peptide chain corresponding to the cell penetrating peptide (CPP) consisting of octa-Arg (R₈) sequence. An N-terminal amino group of the resulting resin was protected with Boc group by the use of Boc₂O (4 equiv.), HOBT·H₂O (4 equiv.) and DIPEA (4 equiv.) in DMF for 2 h to give Boc-protected peptide resin S4.

Acylation of Ser side chain: A mixture of Fmoc-Ala-Cl¹ (10 equiv.), DIPEA (22 equiv.) and

DMAP (0.2 equiv.) in CH₂Cl₂ were added to the Boc-protected peptide resin S4 (0.36 mmol). The reaction mixture was shaken overnight. Reaction progress was monitored by HPLC and ESI-MS analyses of the sample obtained by pilot deprotection. On the completed resin, standard Fmoc SPPS and subsequent incorporation of 4-dimethylaminoazobenzene-4'-sulfanyl (Dabsyl) group by the treatment with Dabsyl chloride (6.0 equiv.) and DIPEA (6.0 equiv.) in DMF were performed to give Dabsyl-incorporated protected resin S5.



Protected peptide:

MIK(Boc)IAT(fBu)R(Pbf)K(Boc)Y(fBu)LGK(Boc)Q(Trt)N(Trt)VY(fBu)D(OtBu)IGVE(OtBu)R(Pbf)D(OtBu)H(Trt)N(Trt)FALK(Boc)N(Trt)GFIA

Figure S2. A general scheme for the synthesis of Q-iso-Ic-CPP-FI 2.

On-resin-thioesterification:² To the peptide resin S5 (0.020 mmol) were added Pd(PPh₃)₄ (3.0 equiv.) and N-methylaniline (63 μl, 30 equiv.) in CH₂Cl₂ (2 ml) to remove the C-terminal allyl

group. After being shaken for 1 h in the reaction mixture, the resin was thoroughly washed with DMF in the presence of PPh₃ to yield the C-terminal carboxylic acid resin. To the resulting resin were added PyBOP (15 equiv.), DIPEA (15 equiv.) and EtSH (90 equiv.) in DMF with additional shaking for 2 h at room temperature to afford the peptide thioester-linked resin **S6**.

Global deprotection and purification: The resulting completed resin **S6** was treated with TFA:*m*-cresol:thioanisole:H₂O (85:5:5:5 (v/v), 50 μl/1 mg resin) for 2 h. Then, the resin was filtrated off. To the resulting filtrate was added cooled Et₂O to give precipitate. The formed precipitate was collected by centrifugation and thoroughly washed with Et₂O to afford a crude iso-peptide **S7** as a diastereomeric mixture (derived from the use of racemic Spr(*o*NB)). The crude iso-peptide **S7** was purified by semi-preparative HPLC (HPLC conditions: linear gradient of solvent D in solvent C over 30 min, 32% to 35%). Analytical HPLC conditions: linear gradient of solvent D in solvent C, 30% to 40% over 30 min, retention time = 20.7 min or 21.2 min, respectively for each diastereomer, MS (ESI-TOF) *m/z* expected (average isotopic mass) 6118.1, found 6117.4 and 6117.3 (diastereomeric mixture).

Preparation of fluorophore and Flag tag-containing extein peptide: For the preparation of fluorophore and Flag tag-containing extein peptide **S8**, NovaSyn[®] TGR resin (Rink amide type: 0.25 mmol amine/g) was used. On this resin, requisite protected peptide chain was elongated by using standard Fmoc-based protocols. At Lys4 position, Fmoc-Lys(ivDde)-OH was used for the chemoselective incorporation of the fluorophore. After chain elongation, the N-terminal Fmoc group was replaced with Boc group. After treatment of the resin (0.025 mmol) with 2% (v/v) hydrazine monohydrate in DMF to deprotect the ivDde group for 2 days, the resulting amino group of Lys4 side chain was reacted with fluorescein isothiocyanate (FITC, 2.4 equiv.) and DIPEA (2.5 equiv.) in DMF (200 μl) overnight to yield a fluorophore-containing protected peptide derivative **S8**. Global deprotection was performed in a manner similar to that of iso-peptide **S7** (TFA:*m*-cresol:thioanisole:H₂O (80:5:5:5:2.5:2.5 (v/v), 50 μl/1 mg resin). The crude extein peptide **S8** was purified by preparative HPLC (HPLC conditions: linear gradient of solvent D in solvent C over 30 min, 20% to 30%). Analytical HPLC conditions: linear gradient of solvent D in solvent C, 10% to 40% over 30 min, retention time = 23.6 min, MS (ESI-TOF) *m/z* calcd ([M + 2H]²⁺) 1032.9, found 1032.9.

Preparation of Q-iso-Ic-CPP-FI 2:³ The diastereomeric mixture of **S7** (0.23 μmol) and **S8**

l of ligation buffer (6 M guanidine·HCl (Gdn·HCl), 0.1 M Na phosphate, 50 mM 4-mercaptobenzenoic acid (MPAA), pH 6.9). After 12 h at 37 °C, the NCL reaction³ gave unexpected 4-aminophenylsulfonyl derivative **S9** resulting from reductive cleavage of diazo moiety in the Dabsyl group. After being purified by HPLC, the resulting aniline peptide **S9** was reconverted to the desired Dabsyl peptide Q-iso-Ic-CPP-FI **2** through diazo coupling with *N,N*-dimethylaniline. To a solution of **S9** in 6 M Gdn·HCl containing 0.1 M sodium phosphate buffer (pH 3.0) was added aqueous solution of NaNO₂ (500 mM, final concentration 10 mM) followed by incubation on ice for 20 min. *N,N*-dimethylaniline was added to the reaction mixture (final concentration 100 mM), then the mixture was incubated on ice for 1 h. The product was purified by semi-preparative HPLC (HPLC conditions: linear gradient of solvent D in solvent C over 30 min, 32% to 35%) to give Q-iso-Ic-CPP-FI **2** as an inseparable diastereomeric mixture (0.1 mg, 0.01 μmol, 4%) (Fig. S1, S3). Analytical HPLC conditions: linear gradient of solvent D in solvent C, 25% to 40% over 30 min, retention time = 25.8 min; MS (ESI-TOF) *m/z* calcd (average isotopic mass) 8121.1, found 8120.5.

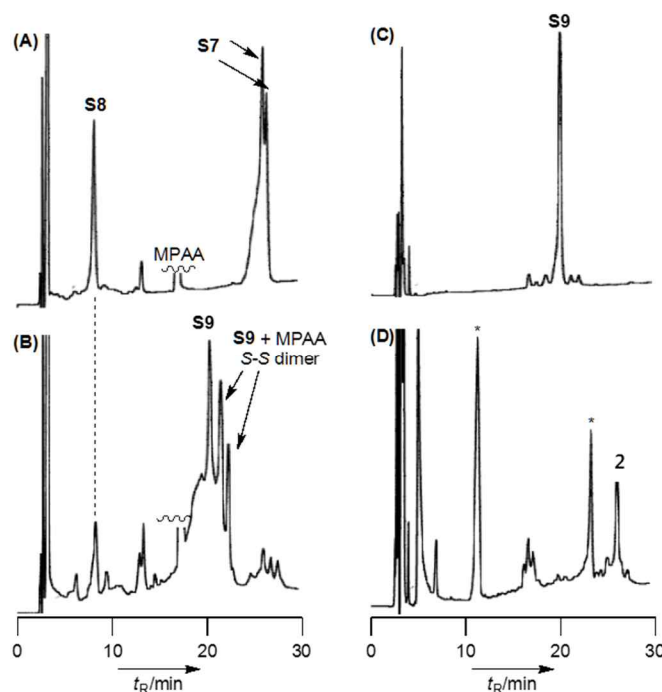


Figure S3. HPLC monitoring of NCL between **S7** and **S8** (A and B) followed by diazo coupling of **S9** with *N,N*-dimethylaniline (C and D). (A) NCL ($t = 0$ h). (B) NCL ($t = 12$ h). (C) diazo coupling ($t = 0$ h) without NaNO₂ and *N,N*-dimethylaniline. (D) diazo coupling ($t = 1$ h). HPLC conditions: linear gradient of solvent D in solvent C, 25% to 40% over 30 min *: non-peptidyl compounds.

Construction of the expression plasmids

All DNA constructs were prepared using general cloning strategy and sequences were identified by DNA sequencing.

Preparation of fusion protein MBP-I_N 3 (pMKM012): The I_N coding sequence was amplified by PCR using the total DNA from *Nostoc punctiforme* (strain ATCC 29133/PCC 73102) as a template using a pair of oligo-nucleotide primers (5'-ACG GAT CCT GTT TAA GCT ATG AAA CGG AAA TAT TG-3' and 5'-ATG AAG CTT AAT TCG GCA AAT TAT CAA CCC G-3', restriction sites are indicated by underlines). The resulting PCR products were digested and inserted into the pMAL-C2 vector using *Bam*H I and *Hind* III restriction sites to create pMKM012 plasmid, which encodes MBP-I_N 3.

Preparation of assembled whole *Npu* DnaE intein (pJDH012): The I_N coding sequence (amino acid residues 1-98) was amplified by PCR using the total DNA from *Nostoc punctiforme* (strain ATCC 29133/PCC 73102) as a template using a pair of oligo-nucleotide primers (5'-GGT CGC CAT ATG TGT TTA AGC TAT-3' and 5'-CAA ATT GTC GAC CCG CAT-3', restriction sites are indicated by underlines). The resulting PCR products were digested and inserted into the pET28a vector using *Nde* I and *Sal* I restriction sites to create pJDH011 plasmid, which encodes His6-I_N (1-98). Sequences of I_N 97-98, GGTGAT (VD) were mutated by *Sal* I restriction site, GTCGAC (VD). Then the I_C coding sequence (amino acid residues 97-137 (including I_N amino acid residues 97-102)) was amplified by PCR using the total DNA from *Nostoc punctiforme* (strain ATCC 29133/PCC 73102) as a template using a pair of oligo-nucleotide primers (5'-TAA GTC GAC AAT TTG CCG AAT ATC AAA ATA GCC-3' and 5'-TCC CGT CTC GAG ATT AGA AGC TAT-3'). The resulting PCR products were digested and inserted into the pJDH011 vector using *Sal* I and *Xho* I restriction sites to create pJDH012 plasmid, which encodes His6-whole *Npu* intein.

Preparation of fusion protein, MBP-eI_N 4 (pJDH025): The eI_N coding sequence was amplified by PCR using the pJDH012 plasmid as template using a pair of oligo-nucleotide primers (5'-GAT ATA GAA TTC TGT TTA AGC TAT GAA AC-3' and 5'-GCA AA AAG CTT TTA GCG CTC AAC-3'). Resulting PCR product was digested and inserted into the pMAL-c2 vector using *Eco*R I and *Hind* III restriction sites to create pJDH015 plasmid, which encodes MBP-eI_N 4. The MBP-eI_N 4 coding sequence was amplified by PCR using the pJDH015

plasmid as template using a pair of oligo-nucleotide primers (5'-GGA TCC ATG AAA ATC GAA GAA GGT AA-3' and 5'-CTC GAG TTC GCG CTC AAC TCC A-3'). Resulting PCR product was digested and inserted into the pcDNA3 mammalian expression vector (Invitrogen, Carlsbad, CA, USA) using *BamH* I and *Xho* I restriction sites to create mammalian vector pJDH025 plasmid, which encodes MBP-e_{I_N} 4.

Analysis of *in vitro* protein trans-splicing kinetics

Purified recombinant MBP-I_N 3 (2 mM) and MBP-e_{I_N} 4 (2 mM) were added to synthetic Q-Ic-FI 1 (10 mM), individually. The PTS was carried out in trans-splicing buffer (50 mM Tris·HCl, 350 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol, pH 7.0) and the formation of the PTS product was monitored by using SDS-PAGE analysis followed by Coomassie brilliant blue staining or fluorescent imaging. The image density of reactant and product bands was measured using ImageJ program and the yield was calculated using the following equation. The experiments were repeated three times to obtain error bars.

$$\frac{\frac{\text{Band density of TS product}}{M W \text{ of TS product}}}{\frac{\text{Band density of TS product}}{M W \text{ of TS product}} + \frac{\text{Band density of N - inte fusion}}{M W \text{ of N - inte fusion}}} \quad \text{Eq. 1}$$

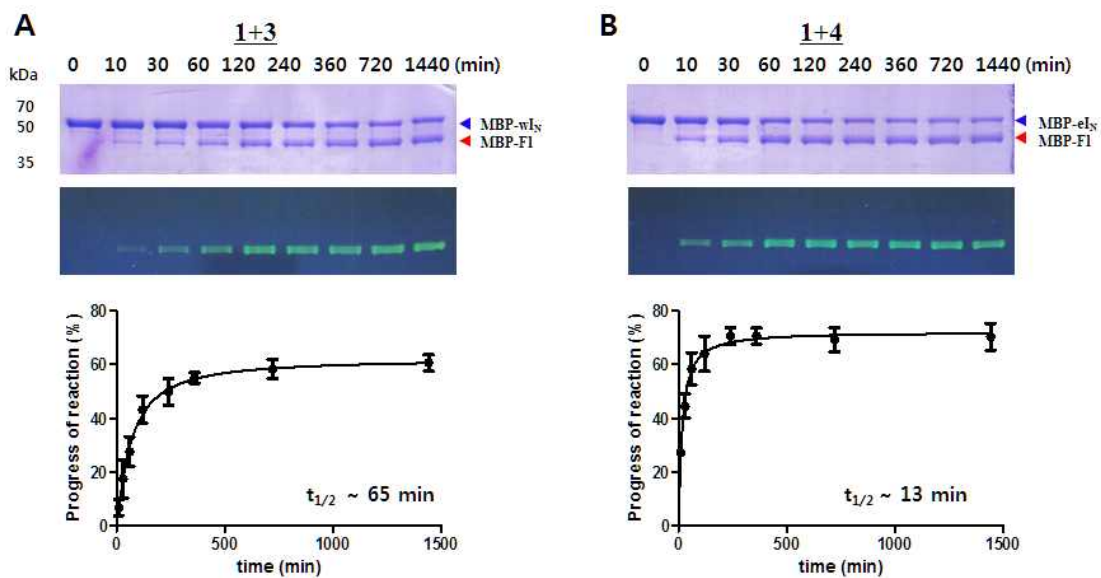


Figure S4. *In vitro* protein trans-splicing reaction of 1+4 and 1+2 were analyzed using SDS-PAGE. The progress of reaction (fraction of product) was plotted versus reaction time for 24 hours. The observed $t_{1/2}$'s were 65 min for 1+3 reaction and 13 min for 1+4 reaction in the absence of chemical reductants.

Photoactivation of iso-peptide and *in vitro* protein trans-splicing

The interaction of Q-iso-I_C-CPP-Fl **2** and MBP-eI_N **4** was monitored with or without irradiation, *in vitro* (Fig. S4). Purified Q-iso-I_C-CPP-Fl **2** and MBP-eI_N **4** were reconstituted in trans-splicing buffer (50 mM Tris·HCl, 350 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol, pH 7.0) to a final concentration of 10 μM and 2 μM, respectively. The PTS was initially inhibited with the use of the photocaged peptide, Q-iso-I_C-CPP-Fl **2**, and then activated by removing the photocage by irradiation with 365 nm wavelength light for 10 min using a hand-held UV lamp (Fig. S5). The activation of fluorescent signal was monitored and converted to the percent yield of the PTS reaction (Table S1).

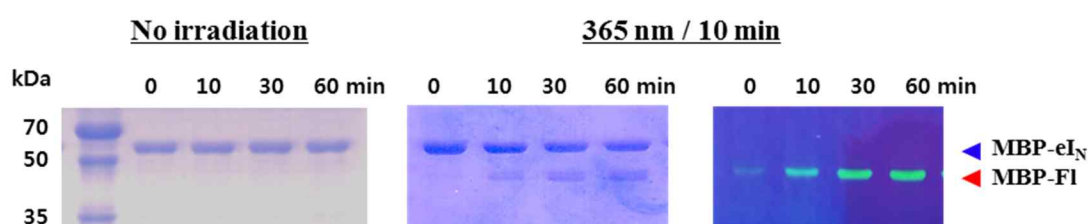


Figure S5. SDS-PAGE analyses of PTS between Q-iso-I_C-CPP-Fl **2** and MBP-eI_N **4** with and without irradiation. The PTS between Q-iso-I_C-CPP-Fl **2** and MBP-eI_N **4** at 10 and 2 μM, respectively, were performed in splicing buffer (1 mM EDTA, 50 mM Tris·HCl, 350 mM NaCl, 10% (v/v) glycerol, pH 7.0) at room temperature. The PTS was successfully inhibited when photo-caged iso-I_C was used (no irradiation) but the PTS activity was restored upon the removal of photo-cage (365 nm/10 min) and restoration of original I_C backbone (final yield ~60% at 24 h).

Species	Relative fluorescent intensity (au)	Calculated PTS yield (%)
I _C -Fl	100	
Q-iso-I _C -CPP-Fl 2	4	
2+4 irradiated for 10 min	7	3
2+4 10 min after irradiation	15	11
2+4 30 min after irradiation	28	25
2+4 60 min after irradiation	34	31

Table S1. The activation of fluorescent signal was measured and converted to PTS yield.

Fluorescent labeling of target proteins in live cells

HeLa cells were cultured in DMEM supplemented with 10% (v/v) FBS, and 1% (v/v) antibiotics (100 μg/ml streptomycin, 20 U/ml penicillin) at 37°C in a humidified atmosphere

of 5% CO₂ and 95% air. HeLa cells were grown to 60–70% confluency and then transiently transfected with pJDH025 or an empty vector by using a transfection reagent, Polyplus transfection (Strasbourg, France), according to manufacturer's protocol. Both batches of HeLa cells were individually treated with a mixture of Q-I_C-FI 1 (2 μM) and cell penetrating peptide, Pep-1 5 (KETWWETWWTEWSQPKKKRKV, 40 μM) for 1 h, individually. The cells were washed twice with phosphate-buffered saline (PBS) to remove the peptides remaining in the culture medium and were then investigated using confocal fluorescence microscopy. (Fig. S6)

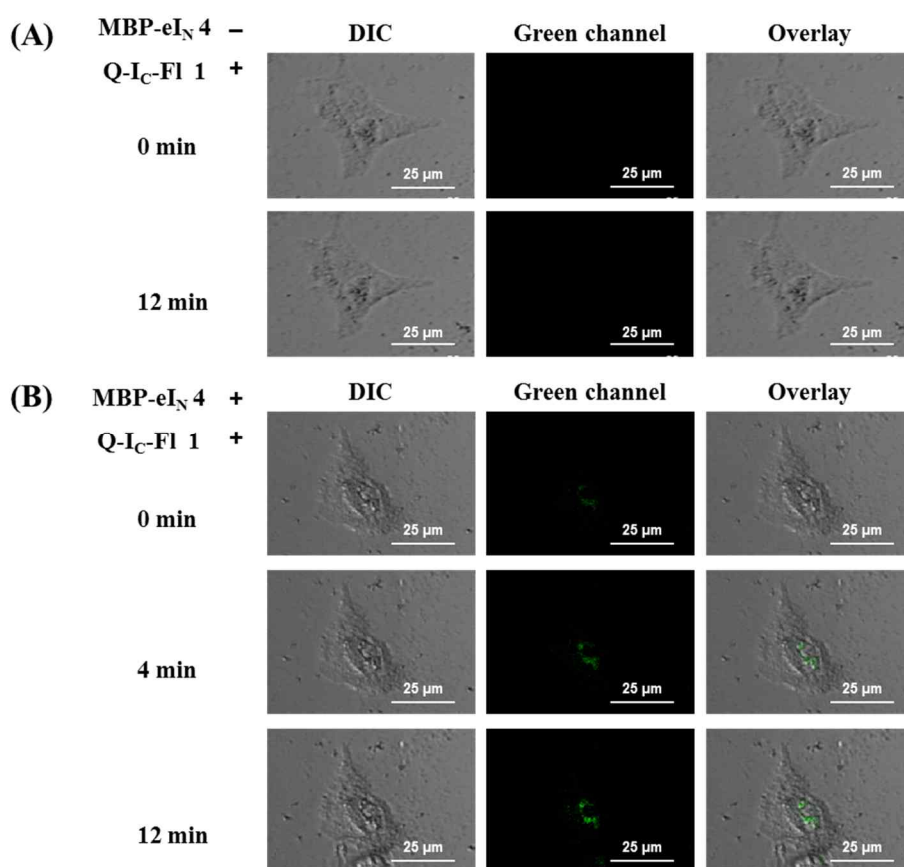


Figure S6. Specific labelling of target proteins in live cells based on the split-intein mediated protein trans-splicing reaction using a Q-I_C-FI 1. A control batch of HeLa cells (A) and a batch of HeLa cells expressing MBP-eIN 4(B) were treated with a mixture of Q-I_C-FI 1 (2 μM) and Pep-1 5 (40 μM) for 1 h, respectively. While no significant fluorescent signal was observed from the control experiment, MBP-eIN 4 expressing HeLa cells showed activation of fluorescent signal as the PTS progressed to label the MBP with fluorescein.

Photo-activation of Q-iso-I_C-CPP-FI 2 in live cells

Photo activation of fluorescent labeling of target proteins was carried out by using simultaneous stimulation and imaging program provided with Eclipse Ti (Nikon Instruments, Tokyo, Japan)

confocal fluorescent microscopy. The cells expressing MBP-eI_N **4** were cultured on a confocal dish and treated with the Q-iso-I_C-CPP-FI **2** (2 μM). The dish was placed in the live cell imaging chamber (Nikon Instruments, Tokyo, Japan) mounted on the microscope. The cells were stimulated for 60 sec by using 403 nm laser with maximum laser power at 2 sec/frame scan rate. The fluorescent image was captured in 30 min after stimulation. Images were processed and quantified using the NIS-Elements Viewer software program.

Western-blot analysis of PTS in live cells

Cells were washed twice with PBS, and the lysates were prepared in RIPA buffer (50 mM Tris·HCl, 150 mM NaCl buffer at pH 8.0 containing 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycholate and 1% (w/v) octylphenoxypolyethoxyethanol) in the presence of 1 mM PMSF. The lysates were then analyzed by SDS-PAGE and transferred to PVDF membranes using standard Western blotting procedures. The trans-splicing product was detected by antibodies recognizing either a MBP or a Flag tag.

References for supporting information

- (1) Beyermann, M.; Bienert, M.; Niedrich, H.; Carpino, L. A.; Sadat-Aalae, D. *J. Org. Chem.* **1990**, *55*, 721.
- (2) Ficht, S.; Payne, R. J.; Guy, R. T.; Wong, C. H. *Chem. Eur. J.* **2008**, *14*, 3620.
- (3) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.