Supplementary Material (ESI) for Chemical Communications

Supplementary Information

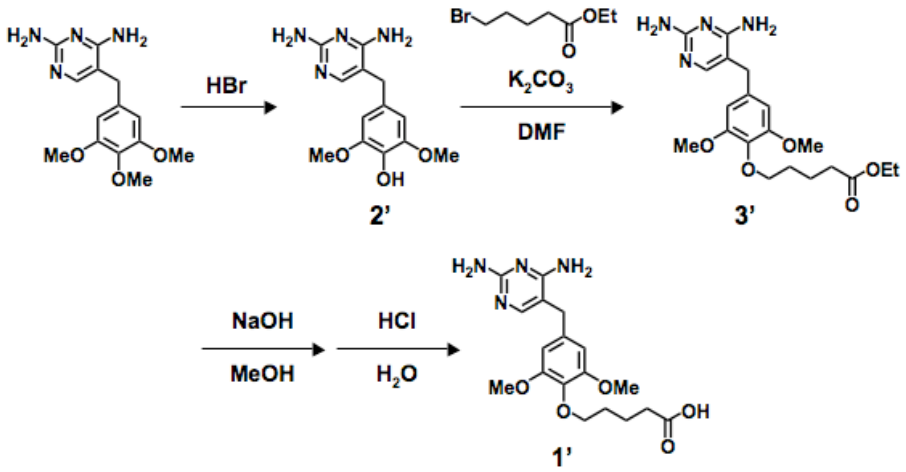
Construction of a small-molecule-integrated semisynthetic split intein for *in vivo* protein ligation

Tomomi Ando, Shinya Tsukiji, Tsutomu Tanaka, and Teruyuki Nagamune*

Department of Chemistry and Biotechnology, Department of Bioengineering, School of Engineering, and Center for NanoBio Integration, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

E-mail: nagamune@bioeng.t.u-tokyo.ac.jp

Synthesis of trimethoprim derivatives



Scheme S1. Synthesis of 4'-carboxy-substituted trimethoprim **1'**

Synthesis of **2'**

Trimethoprim (25.0 g, 86.11 mmol; Sigma) was dissolved in 48 % HBr (300 mL; Aldrich) at room temperature. The mixture was then stirred at 100 ºC for 30 min and quenched by careful addition of 50 % NaOH aq. (60 mL). The reaction was cooled to room temperature and placed at 4 ºC overnight to form crystals. The crystals were filtered and washed with ice-cold water (ca. 100 mL). After the crystals were dissolved in boiling water (200 mL), the solution was neutralized to ca. pH 7 with 28% aqueous ammonia and placed at 4 ºC overnight for recrystalization. White crystals were filtered, washed with ice-cold water (ca. 100 mL), and dried under vacuum over P_2O_5, affording compound **2'** (15.8 g, 66 %).

^{1}H-NMR (*d*-DMSO, 500 MHz): δ 8.23 (brs, 1H), 7.81 (brs, 2H), 7.37 (s, 1H), 7.32 (brs, 2H), 6.56 (s, 2H), 3.72 (s, 6H), 3.54 (s, 2H). brs = broad singlet.

Synthesis of **3'**

To a solution of **2'** (5.0 g, 18.1 mmol) in anhydrous DMF (25 mL) was added potassium carbonate (5.0 g, 36.1 mmol) and ethyl 5-bromovalerate (4.3 mL, 27.1 mmol). The mixture was stirred at 40 ºC for 6 h. After filtration, the filtrate was concentrated under vacuum and diluted with EtOAc (100 mL). The organic layer was washed with saturated $NaHCO_3$ aq. (50 mL × 2) and brine (50 mL), and dried over anhydrous Na_2SO_4. The crude residue was purified by column chromatography on silica gel

(eluent; $CHCl_3$ to 20 : 1 $CHCl_3$: MeOH) to give compound **3'** (4.6 g, 63 %). ^{1}H-NMR (*d*-DMSO, 500 MHz): δ 7.51 (s, 1H), 6.54 (s, 2H), 6.07 (s, 2H), 5.68 (s, 2H), 4.05 (q, 2H), 3.77 (t, 2H), 3.70 (s, 6H), 3.52 (s, 2H), 2.34 (t, 2H), 1.69 (m, 2H), 1.60 (m, 2H), 1.18 (t, 3H)
MALDI-TOF-MS (Dithranol): calcd for $[M+H]^+$ = 405.47; obsd 405.25.

Synthesis of **1'**

To a solution of **3'** (5.68 g, 14.0 mmol) in MeOH (30 mL) was added 5.0 M NaOH aq. (8.4 mL, 42.0 mmol). The mixture was stirred at room temperature for 1 h. After evaporation, the crude residue was dissolved in water (100 mL) and acidified to ca. pH 4 with conc. HCl to form a white precipitate. The precipitate was filtered, washed with ice-cold water (ca. 100 mL), and dried under vacuum over P_2O_5, affording compound **1'** (3.43 g, 65 %).
^{1}H-NMR (*d*-DMSO, 500 MHz): δ 7.50 (s, 1H), 6.55 (s, 2H), 6.32 (s, 2H), 5.93 (s, 2H), 3.78 (t, 2H), 3.71 (s, 6H), 3.53 (s, 2H), 2.26 (t, 2H), 1.64 (m, 4H)
MALDI-TOF-MS (CHCA): calcd for $[M+H]^+$ = 377.41; obsd 377.35.

Peptide synthesis

The peptides were synthesized manually on a Rink Amide resin by standard Fmoc-based solid-phase peptide synthesis protocol. Fmoc-Leu-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Gly-OH, Fmoc-Cys(Trt)-OH, Fmoc-Ile-OH, Fmoc-Asp(O*t*Bu)-OH, Fmoc-Ala-OH, Fmoc-Lys(Mtt)-OH, and Fmoc-amide-dPEG$_4$ Acid (Quanta BioDesign) were used as building blocks. Fmoc deprotection was performed with 20 % piperidine in DMF, and coupling reactions were performed with a mixture of Fmoc-amino acid, diisopropylcarbodiimide (DIC), and HOBt in DMF. In the case of peptide I_N-**2**, Fmoc-Lys(Mtt)-OH was first introduced to a Rink Amide resin and the Mtt group was selectively deprotected by treatment with CH_2Cl_2 containing 2 % TFA and 5 % triisopropylsilane. **1'** was coupled to the side chain of the lysine with a mixture of **1'**, HBTU, HOBt, and *N,N*-diisopropylethylamine in DMF. Following chain assembly, global deprotection and cleavage from the resin was performed with TFA containing 5 % ethanedithiol and 2.5 % H_2O. The crude peptide products were precipitated by Et_2O and purified by reversed-phase HPLC using

a semi-preparative YMC-Pack ODS-A C18 column with a linear gradient of 0.1 % aqueous TFA and acetonitrile containing 0.1 % TFA. The peptides were identified by MALDI-TOF-MS (CHCA): I_N-**1** calcd for $[M+H]^+$ = 1847.14, obsd 1847.5; I_N-**2** calcd for $[M+H]^+$ = 3322.86, obsd 3323.8.

Construction of expression plasmids

The expression plasmid for I_C-**3**, pBAD-DHFR-DnaB$_C$-mRFP-H, was constructed as follows. A DNA fragment encoding His-tag (5'-GCG GTA CCG CGG CCG CTT CTT CTG GTC TGG TGC CAC GCG GTT CTC ACC ATC ATC ATC ATC ATT AAG AGC TCG C-3' and its complementary strand) was inserted into the KpnI/SacI sites of pBAD-DHFR[S1] to yield pBAD-H. The gene encoding *e*DHFR was obtained by PCR from pBAD-DHFR[S1] using the 5'-primer (5'-CAT ACC CAT GGG AAT CAG TCT GAT TGC GGC GTT AGC GG-3') and the 3'-primer (5'-CGG TAC CAG ATC TCC CCC GCC GCT CCA GAA-3') and subcloned into the NcoI/KpnI sites of the pBAD-H to yield pBAD-DHFR-H. The gene encoding DnaB$_C$ (corresponding to amino acids 12-154) was obtained by PCR from pTWIN1 (New England Biolabs) using the 5'-primer (5'-CGG GGG AGA TCT GAA AAC CTG TAT TTT CAG AGC ACA GGA AAA AGA GTT TCT ATT AAA GAT TTG-3') and the 3'-primer (5'-CCA GAA GAA GCG GCC GCG GTA CCA TCC TGT TCG ATG GAG TTG TGT ACA ATG ATG TCA TTC GCG ACA AAG TTA TGT GG-3') and subcloned into the BglII/KpnI sites of the pBAD-DHFR-H to yield pBAD-DHFR-DnaB$_C$-H. Subsequently, the gene encoding mRFP was obtained by PCR from pDsRed-Monomer-N1 (Clontech) using the 5' primer (5'-GCG GTA CCA TGG ACA ACA CCG AGG ACG TC-3') and the 3' primer (5'-AGC TGC GGC CGC CTG GGA GCC GGA GTG GCG GGC CTC GGC G-3') and subcloned into KpnI/NotI sites of the pBAD-DHFR-DnaB$_C$-H to yield pBAD-DHFR-DnaB$_C$-mRFP-H. In *E. coli*, the plasmid expresses the protein construct [*e*DHFR]-GRSENLYFQ-[DnaB$_C$(12-154)]-SIEQDGT-[mRFP]-AAASSGLVPRGS-[His-tag(H6)].

The expression plasmid for the splicing-deficient mutant I_C-**4**, pBAD-DHFR-DnaB$_C$(S155A)-mRFP-H, was constructed by introducing mutations into the pBAD-DHFR-DnaB$_C$-mRFP-H with a QuickChange Site-Directed Mutagenesis Kit (Stratagene) using the 5' primer (5'-TCA TTG TAC ACA ACG CCA TCG AAC AGG

ATG-3'; mutated site are underlined) and the 3' primer (5'-CAT CCT GTT CGA TGG CGT TGT GTA CAA TGA-3'). In *E. coli*, the plasmid expresses the protein construct [*e*DHFR]-GRSENLYFQ-[DnaB$_C$(12-154)]-AIEQDGT-[mRFP]-AAASSGLVPRGS-[His-tag(H6)].

The expression plasmid for I$_C$-**5**, pTrc-DnaB$_C$-mRFP, was constructed as follows. The gene encoding DnaB$_C$ was obtained by PCR from pTWIN1 using the 5' primer (5'-GCG GAT CCA GCA CAG GAA AAA GAG TTT CTA TTA AAG ATT TG-3') and the 3' primer (5'-GCG AGC TCA TCC TGT TCG ATG GAG TTG TGT ACA ATG ATG TCA TTC GCG-3') and subcloned into the BamHI/SacI sites of pTrcHis A (Invitrogen) to yield pTrc-DnaB$_C$. The gene encoding mRFP was obtained by PCR from pDsRed-Monomer-N1 (Clontech) using the 5' primer (5'-GCG GAT CCG AGC TCA TGG ACA ACA CCG AGG ACG TCA TCA AGG-3') and the 3' primer (5'-GCG GTA CCC TAC TTA TCG TCG TCA TCC TTG TAA TCC TCG AGC TGG GAG CCG GAG TGG CGG G-3') and subcloned into the SacI/KpnI sites of the pTrc-DnaB$_C$ to yield pTrc-DnaB$_C$-mRFP. In *E. coli*, the plasmid expresses the protein construct MGGS-[His-tag(H6)]-GMASMTGGQQMGRDLYDDDDKDRWGS-[DnaB$_C$(12-154)]-SIEQDEL-[mRFP]-LEDYKDDDDK.

The expression plasmid for I$_C$-**10**, pDis-DHFR-mRFP, was constructed as follows. The gene encoding *e*DHFR was obtained by PCR from pBAD-DHFR[S1] using the 5' primer (5'-GCA GAT CTA GCG GAA TCA GTC TGA TTG CGG CGT TAG CGG-3') and the 3' primer (5'-GCC CGC GGC TCG AGA AGC TTG AGC TCG GAT CCG ATA TCC CGC CGC TCC AGA ATC TCA AAG CAA TAG C-3') and subcloned into the BglII/SacII sites of pDisplay (Invitrogen) to yield pDis-DHFR. The gene encoding IgS-HA-*e*DHFR-Myc-TMD region was obtained by PCR from the pDis-DHFR using the 5' primer (5'-GCG CTA GCA TGG AGA CAG ACA CAC TCC TGC TAT GGG-3') and the 3' primer (5'-GAC TAG CGG CCG CCT AAC GTG GCT TCT TCT GCC-3') and subcloned into the NheI/NotI sites of pEGFP-N1 (Clontech) to yield pDis-DHFR2. The mRFP gene obtained from pTrc-DnaB$_C$-mRFP was subsequently subcloned into the SacI/XhoI sites of the pDis-DHFR2 to yield pDis-DHFR-mRFP. In mammalian cells, the plasmid expresses the protein construct [Igκ–chain signal sequence]-[HA(YPYDVPDYA)]-GAQPARSSG-[*e*DHFR]-DIGSEL-[mRFP]-LEPRLQVD-[Myc(EQKLISEEDL)]-N-[TMD].

The expression plasmid for I$_C$-**9**, pDis-DHFR-DnaB$_C$-mRFP, was constructed

by inserting the $DnaB_C$ gene obtained from pTrc-$DnaB_C$-mRFP into the BamHI/SacI sites of the pDis-DHFR-mRFP. In mammalian cells, the plasmid expresses the protein construct [Igκ–chain signal sequence]-[HA]-GAQPARSSG-[*e*DHFR]-DIGS-[$DnaB_C$ (12-154)]-SIEQDEL-[mRFP]-LEPRLQVD-[Myc]-N-[TMD].

Pyrobest DNA polymerase (Takara) was used for PCR and all PCR amplified sequences were verified by DNA sequencing.

Protein expression and purification

The plasmids were transformed into *E. coli* TOP10 (Invitrogen). The cells were grown in LB medium to an O.D. (600 nm) of 0.8, at which time expression of the protein was induced by the addition of L-arabinose to a final concentration of 1 % (wt/vol) (except for pTrc-$DnaB_C$-mRFP, no addition). After growth for an additional 24 h at 27 ºC, the cells were harvested by centrifugation. The cell pellets were resuspended in 50 mM phosphate, 300 mM NaCl, pH 7.0 and lysed by sonication. The proteins were purified from the soluble fraction of the lysate by TALON Metal Affinity Resins (Clontech) according to the manufacture's protocol and dialyzed against 50 mM Tris-HCl, 150 mM NaCl, pH 8.0. The concentration of the proteins was determined using a BCA Protein Assay Kit (Pierce).

***In vitro* protein *trans*-splicing assay**

Equimolar amounts of each peptide I_N and protein I_C were mixed in a reaction buffer (50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, 5 μM (or 10 eq.) NADPH, 5 μM (or 10 eq.) TCEP, pH 7.0) to a final concentration of 0.5 μM. Reactions were performed at 25 ºC for 20 h and stopped by mixing with SDS-PAGE loading buffer (50 mM Tris, 2 % SDS, 6 % 2-mercaproethanol). The protein samples were fractionated by 12.5 % SDS-PAGE and electrotransfered onto a PVDF membrane. The splicing product P-**6** was detected with streptavidin-horseradish peroxidase conjugate (SAv-HRP; Invitrogen) using the Chemi-Lumi One (Nacalai Tesque).

For the MALDI-TOF-MS analysis, the *trans*-splicing reaction was carried out in 1 mL solution at 25 ºC. After 20 h incubation, the reaction solution was concentrated using a Microcon Ultracel YM-10 (Millipore) at 4 ºC and subsequently subjected to the

negative mode MALDI-TOF-MS using sinapic acid as a matrix.

For the kinetic experiments (Figure S1), aliquots were taken at defined time points and immediately mixed with SDS-PAGE loading buffer. I_C-**3** was detected with anti-His-tag antibody-horseradish peroxidase conjugate (anti-His-HRP; Bethyl Laboratories) and formation of the splicing product P-**6** was detected as described above. The data was analyzed using the program Scion Image (Scion Corporation), calculated as percent formation, and fitted using the program KaleidaGraph (Synergy Software).

For the competitive inhibition experiments (Figure S2), I_C-**3** was pre-incubated with 10 μM (20 eq.) of trimethoprim and then mixed with I_N-**2** to initiate *trans*-splicing. After 20 h incubation, the reactions were analyzed as described above.

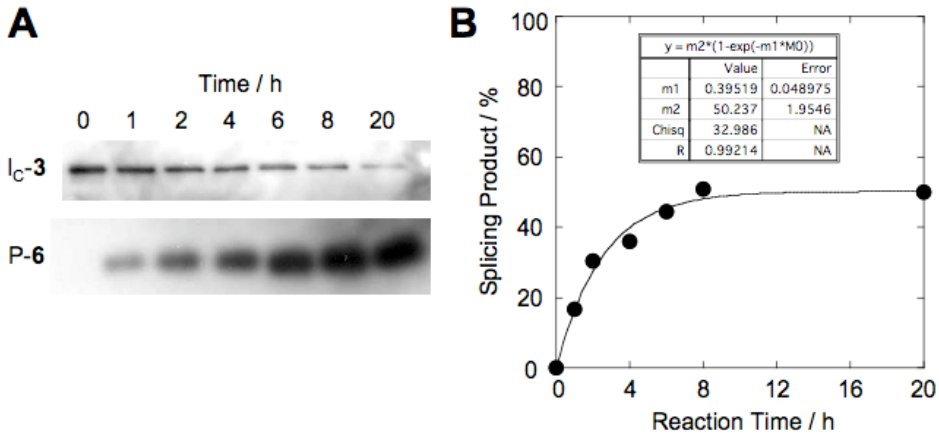


Figure S1. Time-trace of the *trans*-splicing reaction between I_N-**2** and I_C-**3**. (A) Western blotting analysis using anti-His-HRP (top) and SAv-HRP (bottom). (B) A plot of the yield of splicing product P-**6** as a function of time. The k_{obs} for P-**6** formation was obtained by fitting the data to the first-order equation $P = P_0(1\text{-}\exp(-k*t))$,[S2] where P (y in the equation in the inset) is the percentage of splicing product at time t (M0), P_0 (m2) is the maximum percentage of splicing product obtained, and k (m1) is the observed rate constant. The obtained k (m1) value, 0.395 h^{-1} is equal to 1.1×10^{-4} s^{-1}.

Figure S2. Inhibition of the *trans*-splicing reaction with TMP. The *trans*-splicing reaction of I_N-**2** and I_C-**3** was performed in the absence (lane 1) and presence (lane 2) of excess TMP.

Investigation of the effect of linker length on *trans*-splicing efficiency

The I_N peptides shown in Figure 2 were synthesized as described for I_N-**2** (n = 4) and identified by MALDI-TOF-MS (CHCA): n = 0 calcd for $[M+H]^+$ = 2333.71, obsd 2334.0; n = 2 calcd for $[M+H]^+$ = 2828.28, obsd 2828.3; n = 6 calcd for $[M+H]^+$ = 3817.43, obsd 3818.8. The *trans*-splicing reactions were performed between pairs of 0.5 μM (Figure 2) or 5 μM (Figure S3) of I_N peptide and I_C-**3** at 25 ºC for 20 h and analyzed as described above.

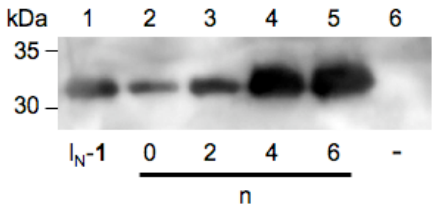


Figure S3. Effects of linker length on the *trans*-splicing efficiency at 5 μM. See also Figure 2.

Cell culture, plasmid transfection, and *trans*-splicing in cell lysates

CHO cells were cultured in Dulbecco's modified Eagle's medium/Nutrient Mixture F-12 Ham (DMEM/F12; Sigma) containing 10 % fetal bovine serum at 37 ºC under 5 % CO_2. Plasmid transfection was carried out using Lipofectamine LTX Reagent (Invitrogen) according to the manufacture's protocol. After 24 h of expression, the cells were washed with PBS, harvested, and lysed in 10 mM HEPES, 150 mM NaCl, 0.5 mM EDTA, 0.5 mM EGTA, 1 % Triton X-100, 1 mM Na_3VO_4, 1mM NaF, 1 mM PMSF, pH 7.4. The cell lysates were mixed with a reaction buffer containing either I_N-**1** or I_N-**2** and incubated at 25 ºC for 20 h. The reactions were analyzed by Western blotting using SAv-HRP as described above.

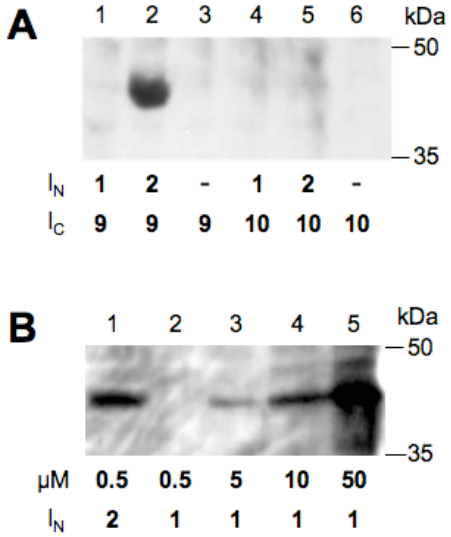


Figure S4. (A) The *trans*-splicing reaction in lysates of CHO cells expressing I_C-**9** or I_C-**10**. The cell lysates were incubated in the presence of 0.5 μM of either I_N-**1** or I_N-**2**. (B) The *trans*-splicing reaction in lysates of I_C-**9**–expressing CHO cells. The concentration of I_N-**1** or I_N-**2** peptide used for the reaction was shown in the figure.

***On-cell trans*-splicing**

CHO cells were transfected as described above and incubated for 24 h. The *on-cell trans*-splicing reactions were performed by incubating the cells with fresh serum-free medium containing 0.5 μM of either I_N-**1** or I_N-**2**, 15 mM HEPES, 5 μM NADPH, and 20 mM reduced glutathione for 8 h at 25 ºC. The cells were washed and

subsequently treated with 5 μg/mL of Alexa Fluor 488-labeled streptavidin (Molecular Probes) at room temperature for 20 min in PBS. After washing again, the cells were observed with a Leica TCS NT confocal laser microscope (Leica Microsystems). The fluorescence images were acquired using the 488 nm line of an argon laser for excitation and a 530 nm band-pass filter for emission, or the 568 nm line of a krypton laser for excitation and a 590 nm long-pass filter for emission. The specimens were viewed using a 100× oil immersion objective.

For Western blotting analysis, the cells after *on-cell trans*-splicing reaction were washed, harvested, and lysed at 4 ºC (on ice) as described above. Note that we have confirmed that no splicing reaction proceeds at 4 ºC (data not shown). The protein samples were analyzed by Western blotting using SAv-HRP as described above.

References

S1) T. Tanaka, N. Kamiya, T. Nagamune, *FEBS Lett.,* 2005, **579**, 2092-2096.

S2) D. D. Martin, M.-Q. Xu, T. C. Evans Jr., *Biochemistry*, 2001, **40**, 1393-1402.