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

Tomomi Ando,[a] Shinya Tsukiji,[a] Tsutomu Tanaka[a] and Teruyuki Nagamune*[abc]

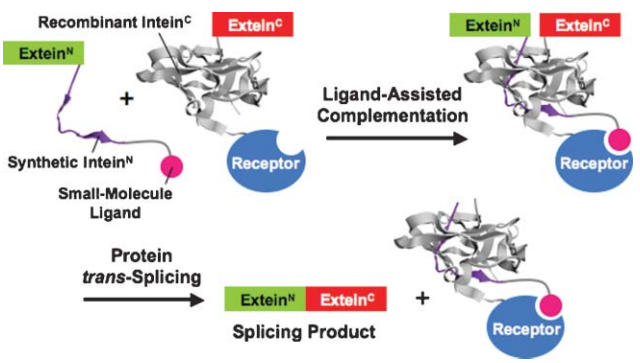
Received (in Cambridge, UK) 22nd August 2007, Accepted 17th September 2007
First published as an Advance Article on the web 10th October 2007
DOI: 10.1039/b712843f

A new split intein-based protein ligation tool that is synthetically accessible and can be used for protein semisynthesis on the cell surface and potentially inside cells has been constructed.

The ability to incorporate site-specifically a range of chemical probes such as fluorescent molecules, photocrosslinkers, unnatural amino acids, and post-translational modifications into cellular proteins will provide invaluable tools for the study of living systems. Consequently, the development of strategies that enable the specific modification of proteins in cellular contexts is an increasingly active area of chemical biology. Although several chemical labeling methods have been described to date, most of them allow only limited types of modification at the side chain of specific residue(s) within remote fusion protein-/peptide-tags.[1] A protein semisynthesis approach[2] is a powerful alternative strategy. The semisynthesis of proteins is based on the chemoselective ligation of a chemically modified peptide to a recombinant protein, thus enabling the precise incorporation of an unlimited variety of chemical probes into the target protein.[2] Despite such significant potential, the utility of this approach is currently limited to *in vitro* conditions due to the lack of protein ligation techniques which are applicable *in vivo*.[3,4] Here we report a new split intein-based protein ligation method that can be used on the cell surface and potentially inside cells.

Inteins are attractive scaffolds for the creation of *in vivo* protein ligation tools, because they can split into two halves to mediate protein splicing *in trans*.[5] The *trans*-splicing process allows N- and C-terminal extein polypeptides, either synthetic or recombinant in origin, to be joined together by a native peptide bond with concomitant removal of the intein complex.[5] In pioneering work, Giriat and Muir demonstrated the feasibility of *in vivo* protein semisynthesis by using the naturally split *Synechocystis* sp. (*Ssp*) DnaE intein.[4] However, the general applicability of the *Ssp* DnaE intein as well as other existing split inteins would be severely limited by their moderate-to-large sizes (*ca.* 36–180 amino acids), which make chemical synthesis of intein fragments extremely difficult or impractical. In addition, for many *trans*-splicing inteins, the intrinsic low affinity between the complementary fragments is a limitation.[5] We focused on the rational design of a new split intein that overcomes these drawbacks. The *Ssp* DnaB intein was chosen as our starting point, because it was recently found that the DnaB intein split at residue 11, generating I_N and I_C, is capable of *trans*-splicing in bacteria[6] and also *in vitro*.[7] The I_N fragment contains only 11 amino acids, so is suitable for preparation and further modification by conventional solid-phase peptide synthesis (SPPS). However, the *in vitro* study reported a reactant concentration, required for *trans*-splicing of the pair of these I_N and I_C fragments, of over 10 μM,[7] which will restrict its broad use for cellular applications. Consistent with this, we observed no *trans*-splicing of the I_N–I_C pair under submicromolar concentrations (*vide infra*). Therefore, inspired by previous two-[8] and three-hybrid strategies,[9] we aimed to confer high affinity to the complementary I_N–I_C pair by integrating an auxiliary receptor–ligand interaction as outlined in Scheme 1. Covalently fusing a small-molecule ligand and its cognate receptor protein to the C-terminus of I_N and the N-terminus of I_C, respectively, should generate a semisynthetic peptide–protein *trans*-splicing system in which the short synthetic I_N fragment can heterodimerize with the recombinant I_C efficiently to reconstitute the splicing-active complex.

To test our design strategy, a series of model I_N peptides and I_C proteins were prepared, in which the chemical probe, the N-terminal extein, is a biotin tag, and the target protein, the C-terminal extein, is the monomeric red fluorescent protein (mRFP) (Fig. 1(A)). We used *E. coli* dihydrofolate reductase (*e*DHFR) and trimethoprim (TMP) as the receptor–ligand pair, because TMP has nanomolar affinity for *e*DHFR and can be easily derivatized without substantial loss of its binding affinity.[10] We synthesized a 4′-carboxy-substituted TMP (Scheme S1, ESI†) and attached it to the C-terminus of I_N *via* a linker consisting a tetramer of PEG_4-based amino acid.[11] The peptides were readily



Scheme 1 Schematic representation of the small-molecule-integrated semisynthetic split intein.

[a]*Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
[b]*Department of Bioengineering, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. E-mail: nagamune@bioeng.t.u-tokyo.ac.jp; Fax: +81-3-5841-8657; Tel: +81-3-5841-7328*
[c]*Center for NanoBio Integration, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
† Electronic supplementary information (ESI) available: Full experimental procedures. See DOI: 10.1039/b712843f

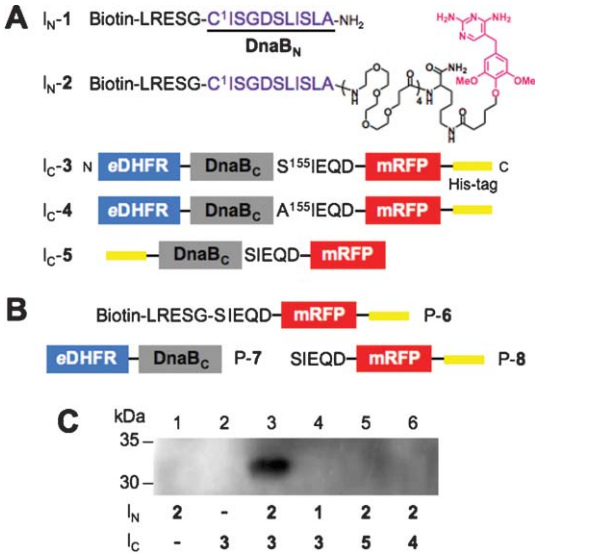


Fig. 1 *In vitro trans*-splicing assays. (A) The I_N peptides and I_C proteins used in this study. The LRESG and SIEQD sequences in I_N and I_C fragments are native exteins of *Ssp* DnaB intein. TMP is shown in pink. (B) Protein products of the *trans*-splicing between I_N-**2** and I_C-**3**. (C) Western blotting analysis of P-**6** formation using SAv-HRP. The reactions were performed between pairs of 0.5 μM of I_N peptide and I_C protein at 25 °C for 20 h.

synthesized by standard SPPS and purified by reversed-phase HPLC. Each of the protein constructs were bacterially expressed and purified from the soluble fraction by affinity chromatography.

Initial *trans*-splicing assays were performed by incubating each of the I_N and the I_C at an equimolar concentration of 0.5 μM and analyzed by Western blotting using streptavidin–horseradish peroxidase conjugate (SAv-HRP). While the incubation of I_N-**1** and I_C-**3** for 20 h resulted in no splicing (Fig. 1(C), lane 4), the formation of splicing product P-**6** was clearly observed in the reaction of I_N-**2** and I_C-**3** (Fig. 1(C), lane 3). The product band showed cross-reactivity to the anti-His-tag antibody as expected (data not shown). The formation of splicing product P-**6**, excised intein fragment P-**7**, and C-terminal cleavage product P-**8** was further confirmed by MALDI-TOF-MS analysis of the reaction mixture.[12] In contrast, no splicing was observed between I_N-**2** and active site-mutated (S155A) I_C-**4** (Fig. 1(C), lane 6), which is consistent with product formation occurring *via* protein *trans*-splicing. The total yield of P-**6** was estimated to be *ca.* 50%.[13] This value lies in the range reported for other split inteins.[7,9] In addition, kinetic experiments revealed that the *trans*-splicing reaction follows first-order kinetics with a k_{obs} of 1.1 × 10^{-4} s^{-1} (Fig. S1, ESI†).[14] It should be noted that a significant amount of splicing product P-**6** could be detected even within 1 h (Fig. S1, ESI†).

To further confirm that the interaction between TMP ligand and *e*DHFR plays a critical role in this peptide–protein *trans*-splicing system, the reactions were carried out using I_N-**2** and I_C-**3** in the presence of a 20-fold molar excess of TMP, and using I_N-**2** and I_C-**5** lacking the *e*DHFR portion. In neither case was the product formed (Fig. S2, ESI† and Fig. 1(C), lane 5). These results provide clear evidence that the specific *e*DHFR–TMP interaction facilitates the *trans*-splicing reaction.

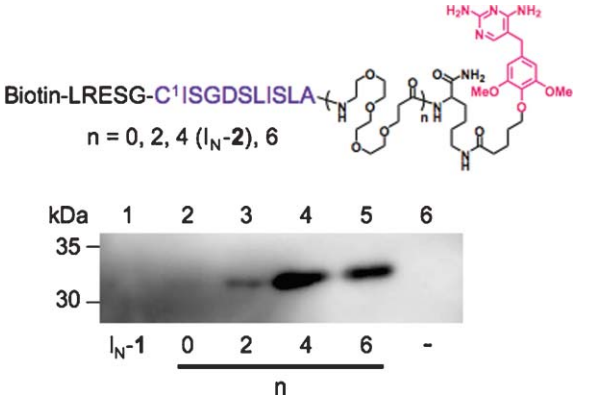


Fig. 2 Effects of linker length of TMP-integrated I_N peptide on *trans*-splicing efficiency. The reactions were performed between pairs of 0.5 μM of I_N peptide and I_C-**3** at 25 °C for 20 h and analyzed by Western blotting using SAv-HRP.

We also investigated the effect of linker length of TMP-integrated I_N peptide on *trans*-splicing efficiency and found the tetramer of PEG_4-amino acid to be optimal among peptides we tested (Fig. 2). It was clearly indicated that a linker of appropriate length is essential for the reconstitution of splicing-active intein complex in this system. Importantly, at a reaction concentration of 5 μM, the formation of P-**6** could be detected using I_N-**1** and I_C-**3**, whereas the pair of I_N-**2** and I_C-**3** still showed a three-fold higher *trans*-splicing efficiency by virtue of the auxiliary interaction (Fig. S3, ESI†).

To demonstrate the *in vivo* applicability of this ligation tool, we next performed the semisynthesis of a protein on the surface of mammalian cells. We chose cell membrane-bound mRFP as the model protein (Fig. 3(A)). Chinese hamster ovary (CHO) cells

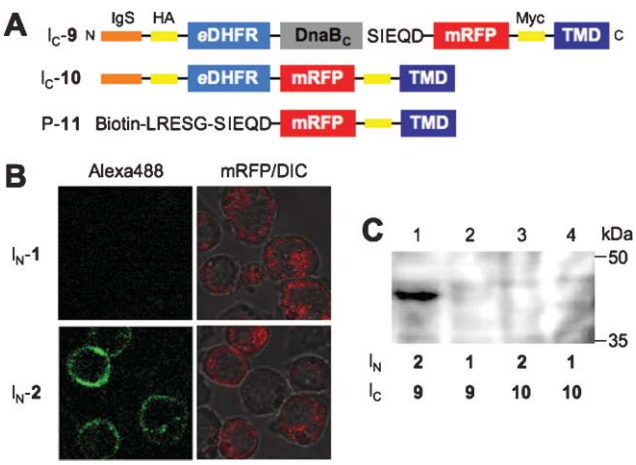


Fig. 3 *On-cell trans*-splicing. (A) The I_C constructs used in this study and the splicing product. IgS is the Igκ signal sequence and TMD is the transmembrane domain of the platelet-derived growth factor receptor. (B) Biotinylation of cell surface-displayed mRFP. CHO cells expressing I_C-**9** were treated with 0.5 μM of either I_N-**1** (top) or I_N-**2** (bottom) at 25 °C for 8 h. The biotinylation was detected with SAv-488. Confocal images show Alexa Fluor 488 fluorescence (left) and overlays of mRFP fluorescence and DIC images (right). (C) Western blotting analysis of the biotinylation of mRFP-TMD (lane 1) using SAv-HRP. Negative controls are shown with I_N-**1** (lane 2) or with cells expressing I_C-**10** (lanes 3 and 4).

were transiently transfected with plasmid encoding I_C-**9**, and cell surface expression of the intein-mRFP fusion was confirmed by immunofluorescence staining using FITC-labeled anti-HA and anti-Myc antibodies (data not shown).[15] Cells were incubated with I_N-**2** for 8 h and washed to remove excess peptide. Cells were then stained with Alexa Fluor 488-labeled streptavidin (SAv-488) and observed by confocal laser scanning microscopy. The images showed that only surfaces of transfected cells were specifically labeled by biotin (Fig. 3(B)). Cells treated with SAv-488 alone showed no fluorescence staining (data not shown). The formation of biotinylated P-**11** was unambiguously confirmed by Western blotting using SAv-HRP (Fig. 3(C), lane 1). No labeling was observed after incubation with I_N-**1** or on cells expressing I_C-**10** that lacks the intein domain (Fig. 3(C), lanes 2–4). These results are in accordance with the *in vitro* experiments described above. To our knowledge, this work represents the first demonstration of a semisynthesis of cell surface protein on living cells.

In conclusion, we have designed and constructed a new split intein tool for *in vivo* protein ligation. The small-molecule-ligand-integrating strategy allowed the creation of semisynthetic split intein in which the I_N fragment is short, hence synthetically accessible, but is capable of *trans*-splicing efficiently through the intein fragment complementation assisted by the auxiliary ligand–receptor interaction. Given the high orthogonality of *e*DHFR–TMP interaction in mammalian cells,[10] this system should be applicable inside living cells by combining it with intracellular peptide delivery methods.[4,16] Because inteins are promiscuous with respect to the extein sequences, the present protein ligation tool would find general use for semisynthesis of proteins containing a variety of chemical probes in cellular contexts.

We are grateful to Toshihiro Ninomiya, Dr Shuji Takeda, and Satoshi Goda for their early contribution to this work. This work was supported by the Center for NanoBio Integration.

Notes and references

1 Selected reviews: I. Chen and A. Y. Ting, *Curr. Opin. Biotechnol.*, 2005, **16**, 35–40; T. Gronemeyer, G. Godin and K. Johnsson, *Curr. Opin. Biotechnol.*, 2005, **16**, 453–458; L. W. Miller and V. W. Cornish, *Curr. Opin. Chem. Biol.*, 2005, **9**, 56–61; K. M. Marks and G. P. Nolan, *Nat. Methods*, 2006, **3**, 591–596.

2 C. J. A. Wallace, *Curr. Opin. Biotechnol.*, 1995, **6**, 403–410; T. W. Muir, *Annu. Rev. Biochem.*, 2003, **72**, 249–289; D. Schwarzer and P. A. Cole, *Curr. Opin. Chem. Biol.*, 2005, **9**, 561–569; B. L. Nilsson, M. B. Soellner and R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 91–118.

3 As a few examples, Yao and coworkers demonstrated that native chemical ligation and expressed protein ligation can be performed in living cells: D. S. Y. Yeo, R. Srinivasan, M. Uttamchandani, G. Y. J. Chen, Q. Zhu and S. Q. Yao, *Chem. Commun.*, 2003, 2870–2871; R. Y. P. Lue, G. Y. J. Chen, Y. Hu, Q. Zhu and S. Q. Yao, *J. Am. Chem. Soc.*, 2004, **126**, 1055–1062.

4 I. Giriat and T. W. Muir, *J. Am. Chem. Soc.*, 2003, **125**, 7180–7181.

5 See the following review and references cited therein: V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429–438.

6 W. Sun, J. Yang and X.-Q. Liu, *J. Biol. Chem.*, 2004, **279**, 35281–35286.

7 C. Ludwig, M. Pfeiff, U. Linne and H. D. Mootz, *Angew. Chem., Int. Ed.*, 2006, **45**, 5218–5221.

8 T. Ozawa, S. Nogami, M. Sato, Y. Ohya and Y. Umezawa, *Anal. Chem.*, 2000, **72**, 5151–5157; T. Ozawa, A. Kaihara, M. Sato, K. Tachihara and Y. Umezawa, *Anal. Chem.*, 2001, **73**, 2516–2521; T. Ozawa, M. Takeuchi, A. Kaihara, M. Sato and Y. Umezawa, *Anal. Chem.*, 2001, **73**, 5866–5874.

9 H. D. Mootz and T. W. Muir, *J. Am. Chem. Soc.*, 2002, **124**, 9044–9045; H. D. Mootz, E. S. Blum, A. B. Tyszkiewicz and T. W. Muir, *J. Am. Chem. Soc.*, 2003, **125**, 10561–10569; H. D. Mootz, E. S. Blum and T. W. Muir, *Angew. Chem., Int. Ed.*, 2004, **43**, 5189–5192.

10 L. W. Miller, Y. Cai, M. P. Sheetz and V. W. Cornish, *Nat. Methods*, 2005, **2**, 255–257; N. T. Calloway, M. Choob, A. Sanz, M. P. Sheetz, L. W. Miller and V. W. Cornish, *ChemBioChem*, 2007, **8**, 767–774.

11 We designed the linker length by predicting the tertiary structure of the fusion complex of I_N and *e*DHFR-I_C using the crystal structures of *Ssp* DnaB intein (PDB ID: 1MI8) and *e*DHFR-methotrexate complex (1DRE).

12 MALDI-TOF-MS analysis was performed in negative mode with sinapic acid as a matrix: P-**6**, calc. 28806, obs. 28789; P-**7**, calc. 35279, obs. 35286; P-**8**, calc. 28037, obs. 28020.

13 Because we could not distinguish the bands of P-**6** and P-**8** by SDS-PAGE/Western blotting due to the similar molecular weight, we estimated the yield of P-**6** by determining the peak ratio of P-**6** and P-**8** by MALDI-TOF-MS analysis (*ca.* 55 : 45). However, the potential differences in ionization energy cannot be fully ruled out.

14 This rate constant is comparable to that for the split *Ssp* DnaE intein (0.7×10^{-4} s^{-1}: D. D. Martin, M.-Q. Xu and T. C. Evans, Jr., *Biochemistry*, 2001, **40**, 1393–1402), the rapamaycin-inducible split VMA intein (1.9×10^{-4} s^{-1}),[9] and the split *Ssp* DnaB intein (I_N–I_C pair) (0.4×10^{-4} s^{-1})[7].

15 We confirmed that the construct I_C-**9** is capable of *trans*-splicing with I_N-**2** in a crude cell lysate (Fig. S4, ESI†). It was demonstrated that over 10 μM of I_N-**1** was required to produce the splicing product P-**11** in an amount comparable to that obtained using 0.5 μM of I_N-**2** in the cell lysate (Fig. S4, ESI†), indicating again the advantage of the present ligation system.

16 J. S. Wadia and S. F. Dowdy, *Curr. Opin. Biotechnol.*, 2002, **13**, 52–56; A. Joliot and A. Prochiantz, *Nat. Cell Biol.*, 2004, **6**, 189–196; M. C. Morris, J. Depollier, J. Mery, F. Heitz and G. Divita, *Nat. Biotechnol.*, 2001, **19**, 1173–1176; T. Takeuchi, M. Kosuge, A. Tadokoro, Y. Sugiura, M. Nishi, M. Kawata, N. Sakai, S. Matile and S. Futaki, *ACS Chem. Biol.*, 2006, **1**, 299–303; M. Okuyama, H. Laman, S. R. Kingsbury, C. Visintin, E. Leo, K. L. Eward, K. Stoeber, C. Boshoff, G. H. Williams and D. L. Selwood, *Nat. Methods*, 2007, **4**, 153–159.