Covalent split protein fragment–DNA hybrids generated through N-terminus-specific modification of proteins by oligonucleotides†

Shuji Takeda,[a] Shinya Tsukiji,[a] Hiroshi Ueda[a] and Teruyuki Nagamune*[a,b,c]

Received 2nd January 2008, Accepted 20th March 2008
First published as an Advance Article on the web 23rd April 2008
DOI: 10.1039/b720013g

Semisynthetic protein–DNA hybrid molecules have recently attracted much attention as valuable tools for bioanalytical chemistry and nanobiotechnology. Here we describe a synthetic method for conjugating oligonucleotides to the N-terminus of recombinant proteins. Our strategy involves the conversion of amine-terminated oligonucleotides to thioester-functionalized oligonucleotides by using a bifunctional reagent bearing an *N*-hydroxysuccinimide ester and benzyl thioester group, followed by native chemical ligation with proteins containing an N-terminal cysteine. We applied this technique to construct split luciferase fragment–DNA hybrid systems in which the catalytic activity of split luciferase is restored by the re-assembly of each fragment through a specific DNA–protein or DNA–DNA interaction. Split protein fragment–DNA hybrids will offer new opportunities to explore the potential of protein–DNA conjugates for various applications.

Introduction

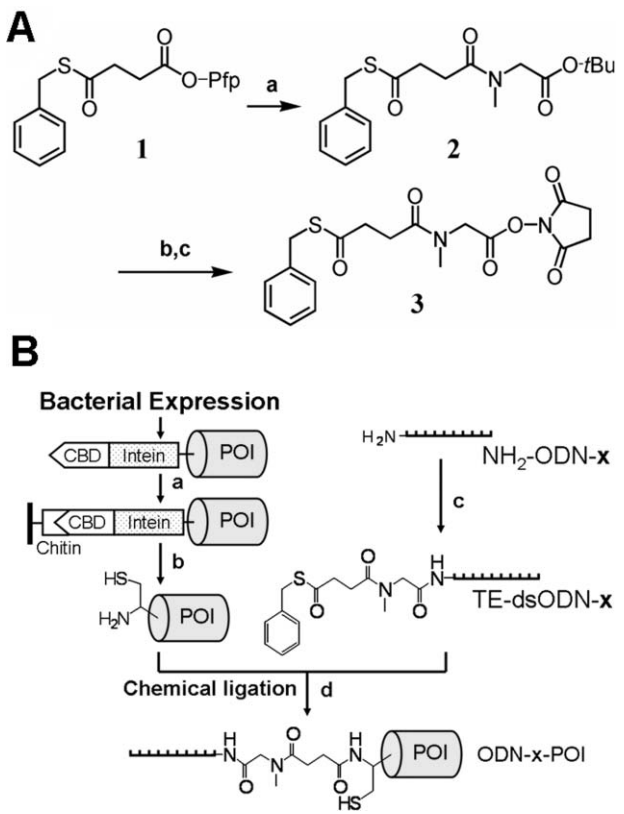
Creating (semi-)artificial proteins with new functions that are useful for a variety of applications is one of the major challenges in protein engineering and nanobioscience. With significant advances in DNA-based technologies, such as the fabrication of DNA microarrays and 2D and 3D DNA nano-architectures,[1] the construction of DNA machines,[2] and the *in vitro* selection of DNA enzymes and aptamers,[3] there has been considerable interest in integrating the unique properties of DNA molecules into proteins and enzymes. In this respect, a growing number of protein–DNA hybrid molecules have been described.[4] For example, semisynthetic nucleases,[5] biosensors for nucleic acids,[6] and signal-amplifiable "tadpoles" for sensitive detection of analytes[7] were constructed by covalently introducing oligonucleotides to proteins. It was also demonstrated that the insertion of single-strand DNA springs allows the preparation of enzymes of which the activity can be controlled by mechanical tension induced by DNA hybridization.[8] Additionally, protein–DNA conjugates have proved to be valuable as molecular scaffolds for the construction of protein microarrays[9] and nanometre-sized protein-based supramolecular objects.[10]

Despite the broad utility of protein–DNA hybrid molecules, the synthetic method for conjugating oligonucleotides to proteins is still very limited. Most of the hybrids reported so far were prepared by classical bioconjugation reactions using bifunctional crosslinkers. Although this strategy provides a general means for covalently attaching oligonucleotides to lysine or cysteine residues within proteins, many proteins display multiple copies of the targeted residues on their surfaces, resulting in the generation of product mixtures. As a result, site-specific DNA modification could be achieved only when the protein of interest contains a single cysteine residue. We[11] and other groups[7,12,13] have recently demonstrated that expressed protein ligation[14] is applicable as a general tool for the preparation of homogeneous protein–DNA conjugates. The method is based on chemical ligation of a recombinant protein containing a C-terminal α-thioester, generated by an intein-fusion technology, with a cysteine-modified oligonucleotide to form an amide bond. A key advantage of this technique is that oligonucleotides can be joined specifically and efficiently to the C-terminus of proteins under mild conditions. However, there may be many situations in which the C-terminus of the protein is functionally important. Therefore, the development of an alternative strategy for conjugating oligonucleotides to the N-terminus of proteins is also highly desirable.

We describe herein a new chemical method for ligating oligonucleotides to recombinant proteins specifically at their N-terminus. Our strategy involves the facile introduction of a thioester moiety into amine-terminated oligonucleotides using a bifunctional reagent bearing an *N*-hydroxysuccinimide ester and a benzyl thioester group, and the subsequent native chemical ligation[15] of the thioester-derivatized oligonucleotides with proteins containing an N-terminal cysteine residue (Scheme 1). Furthermore, we applied this technique to construct split protein fragment–DNA hybrid systems in which the re-assembly of each split protein fragment is mediated by specific protein–DNA or DNA–DNA interactions. Because the functional reconstitution of catalytically (more) active split proteins occurs only upon complementary association,[16] the split protein fragment–DNA conjugates should be capable of transducing molecular complexation events into an enzymatic signal. Such properties will offer new opportunities to

[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: RP-HPLC and SDS-PAGE analyses. See DOI: 10.1039/b720013g



Scheme 1 (A) Synthesis of thioester modification reagent **3**. *Reagents and conditions*: (a) sarcosine *tert*-butyl ester hydrochloride (1 equiv.), DIEA (2 equiv.), CH_2Cl_2, rt, 4 h (**2**, 61%); (b) 50% TFA in CH_2Cl_2; (c) NHS (1.2 equiv.), HBTU (1.2 equiv.), DIEA (1.2 equiv.) DMF, rt, 16 h (**3**, 26%). (B) Strategy for the N-terminus-specific DNA modification of protein of interest (POI): (a) affinity capture and purification of chitin binding domain (CBD)-intein-POI fusion by chitin beads; (b) intein-catalyzed self-cleavage reaction; (c) preparation of thioester-functionalized oligonucleotides using **3** (40–60 equiv. for ODN) in 10 mM borate buffer (pH 8.5) containing 50% DMF; (d) native chemical ligation.

extend the use of protein–DNA hybrid molecules as a platform for bioanalytical chemistry and nanobiotechnology.

Results and discussion

Development of the N-terminus-specific DNA modification method

Our strategy for developing an N-terminus-specific DNA modification method makes use of native chemical ligation (NCL).[15] To this end, it is necessary to have access to both recombinant proteins containing an N-terminal cysteine and oligonucleotides that have a thioester group. The former can be obtained by several techniques such as TEV protease-mediated digestion[17] and intein-fusion expression.[18] On the other hand, to date, no method is available for preparing thioester-appended oligonucleotides. Because solid-phase oligonucleotide synthesis requires treatment with base for full deprotection, the base-sensitive thioester group needs to be incorporated into DNA in a post-synthetic manner. Therefore, we designed a new bifunctional crosslinker **3** containing an *N*-hydroxysuccinimide ester and a benzyl thioester group for the facile introduction of the thioester functionality to commercially

Table 1 Oligonucleotides used in this study

ODN	Sequence (5′–3′)	Amino-modified site
a	TCG ACA TCA AGC	5′
b	AGC ACT TCC ACG	3′
c	TTT ACG CCC ACG CTT T	5′
c′-fl	AAA GCG TGG GCG TAA A	5′
5d	GTA ATC ATG GTC ATA GCT GTT	5′
5d′	AAC AGC TAT GAC CAT GAT TAC	5′
3d′	AAC AGC TAT GAC CAT GAT TAC	3′
5e	GGA TCC TCT AGA TCG ACC TG	5′
3e	GGA TCC TCT AGA TCG ACC TG	3′

available amino-terminated oligonucleotides. The reagent **3** was synthesized as shown in Scheme 1(A). In brief, the starting material **1** was prepared according to the previous report[19] and reacted with sarcosine *tert*-butyl ester in dichloromethane to yield the compound **2**. Deprotection of **2** by TFA treatment and subsequent condensation with *N*-hydroxysuccinimide using HBTU and diisopropylethylamine (DIEA) in DMF afforded the crosslinker **3**.

We tested the reactivity of **3** with amino-oligonucleotides. The oligonucleotides, ODN-**a** and ODN-**b** (Table 1), were incubated with **3** in a 50 mM borate buffer at pH 8.5 containing 50% DMF. The reaction was monitored by reverse-phase HPLC (RP-HPLC) and MALDI-TOF-MS analysis. It was confirmed that the reaction was sufficiently rapid and complete within 1 h, producing the desired benzylthioester-terminated ODNs in quantitative yield with no significant side reactions (see Fig. S1, ESI†). It should be noted that the thioester functionality has stability sufficient for the purification by RP-HPLC and subsequent lyophilization.

Next, the obtained benzyl thioester-functionalized ODNs (TE-ODNs) were subjected to NCL with an enhanced green fluorescent protein that possesses a cysteine at its N-terminus (Cys-EGFP). The Cys-EGFP was prepared as described previously.[20] The ligation reaction was monitored by SDS-PAGE under reducing conditions. A new band corresponding to the ligation product was clearly observed after 48 h incubation at 4 °C, indicating the covalent attachment of the oligonucleotide to the protein (Fig. 1). Ligation yields were approximately 20%. All these results demonstrated that the N-terminus-specific modification of

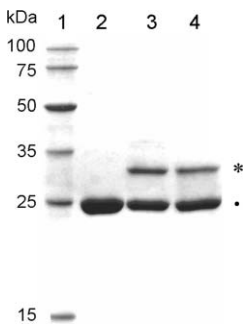


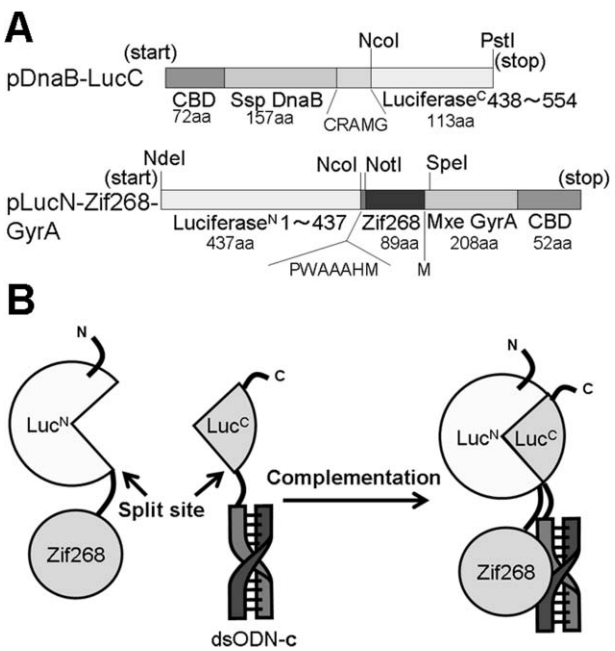
Fig. 1 N-terminus-specific DNA modification. *Reagents and conditions*: 3.7 μM Cys-EGFP, 200 μM TE-ODN-**a** or -**b**, in ligation buffer for 48 h at 4 °C. Lane 1, molecular weight marker; lane 2, Cys-EGFP; lane 3, Cys-EGFP mixed with TE-ODN-**a**; lane 4, Cys-EGFP mixed with TE-ODN-**b**. Cys-EGFP and the ligation product ODN-**a**–EGFP or ODN-**b**–EGFP are denoted by • and *, respectively.

proteins with DNA was accomplished by using the crosslinking reagent **3**.[21]

Design of a split luciferase re-assembly system mediated by a protein–DNA interaction

Split protein fragment complementation is now recognized as an important tool for analyzing protein–protein interactions.[16] A key aspect of this technique is that the activity of the split protein is regained only when a specific protein–protein interaction occurs and induces the re-assembly of each fragment. Thus, we applied the principle of protein fragment complementation to construct a "semi-synthetic" split reporter system in which a protein–DNA interaction induces an enzymatic signal.

The overall scheme of the split protein fragment–DNA hybrid system is illustrated in Scheme 2. As a proof-of-principle study, we chose the zinc finger protein, Zif268, a DNA-binding protein,[22] and its target DNA sequence as an interaction pair. In addition, *firefly* luciferase was used as a reporter protein and split at the position between amino acids 437 and 438, a flexible hinge region that connects the N- and C-terminal domains of the enzyme.[23] The Zif268 was genetically fused to the C-terminus of the N-terminal half of luciferase to give Luc^N-Zif268. The Luc^N-Zif268 was expressed as a fusion with *Mycobacterium xenopi* GyrA intein-chitin binding domain (CBD) (IMPACT system, New England Biolabs). Using this method, after bacterial expression, the Luc^N-Zif268 can be readily affinity-purified on chitin beads and cleaved off from the beads by incubating with dithiothreitol (DTT). In initial experiments, we found that the yield of Luc^N-Zif268 was low (approximately 300 μg per 1 L culture) due to the occurrence of an unwanted *in vivo* self-cleavage reaction. Thus, according to the protocol of IMPACT system, to suppress the *in vivo* self-processing of GyrA intein, a methionine residue was added at the C-terminus of Luc^N-Zif268.[24] The methionine addition resulted in an approximately 50-fold improvement in the yield of Luc^N-Zif268. It is important to note that Luc^N-Zif268 could be expressed in a soluble fraction. The purity and identity of Luc^N-Zif268 were confirmed by SDS-PAGE (Fig. 2(B)). Because the Zif268 domain was obtained in an apo-form, the Luc^N-Zif268 was reconstituted with zinc ion before assay.



Scheme 2 (A) Schematic illustration of the plasmid constructs. (B) Illustration of split enzyme fragment re-assembly driven by interaction of Zif268 and dsDNA.

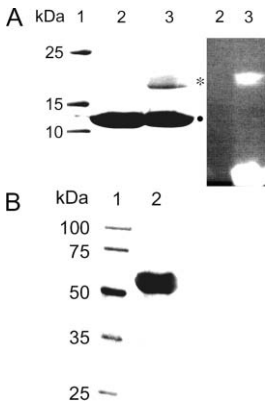


Fig. 2 (A) SDS-PAGE analysis of preparation of dsODN-**c**–Luc^C. Left, Coomassie staining image; Right, fluorescence image. Lane 1, molecular weight marker; lane 2, Cys-Luc^C (12 kDa); lane 3, reaction mixture of TE-dsODN-**c** and Cys-Luc^C. Cys-Luc^C and ligation product dsODN-**c**–Luc^C are denoted by • and *, respectively. (B) SDS-PAGE analysis of Luc^N-Zif268. Lane 1, molecular weight marker; lane 2, Luc^N-Zif268 (60 kDa).

The C-terminal half of luciferase was also expressed by genetically fusing *Synechocystis* sp. DnaB intein-CBD (IMPACT system). The fusion protein was designed to expose a cysteine residue at its N-terminus (Cys-Luc^C) after an intein-mediated self-cleavage reaction.[18] The Cys-Luc^C was bacterially expressed and purified by chitin beads from a soluble fraction. A 5′-amino-modified oligonucleotide containing Zif268 binding sequences, 5′GCG TGG GCG3′ (ODN-**c**), was annealed to a complementary 5′-fluorescein-labeled DNA (ODN-**c′-fl**). The obtained double-strand DNA (dsODN-**c**) was reacted with **3** as described above and purified by gel filtration, affording thioester-functionalized TE-dsODN-**c**. Subsequently, TE-dsODN-**c** was applied to the NCL reaction with Cys-Luc^C. SDS-PAGE analysis using Coomassie staining and fluorescence imaging clearly indicated the formation of the ligation product, dsODN-**c**–Luc^C (Fig. 2(A)). The dsODN-**c**–Luc^C was purified to homogeneity by agarose gel extraction.

The DNA binding activity of the zinc finger motif within Luc^N-Zif268 was investigated by a gel shift assay (Fig. 3). It was clearly observed that the electrophoretic mobility of fluorescent dsODN-**c** was retarded by mixing with holo-Luc^N-Zif268. The fluorescence intensity of the shifted band was increased in proportion to the concentration of holo-Luc^N-Zif268. On the other hand, an apo-form of Luc^N-Zif268 showed no mobility shift. These results indicated that the zinc finger motif of Luc^N-Zif268 is functional and is capable of forming a complex with dsDNA in a holo-form.

We subsequently examined the enzymatic activity of split luciferase hybrids using D-luciferin as a luminogenic substrate (Fig. 4). Consistent with previous reports,[25] Luc^N-Zif268 showed

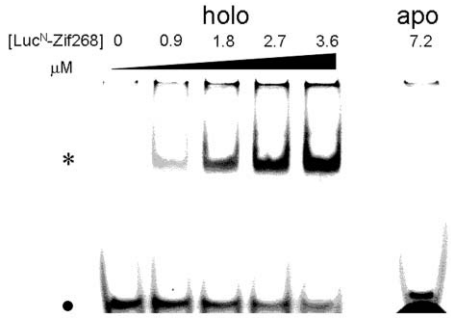


Fig. 3 Polyacrylamide gel shift assay of Luc^N-Zif268 and dsODN-**c**. Fluorescein-labeled dsODN-**c** and Luc^N-Zif268 complexed with dsODN-**c** are denoted by • and *, respectively. Left, various concentrations of holo-Luc^N-Zif298 were mixed with 10 μM dsODN-**c**; Right, apo-Luc^N-Zif268 was mixed with 10 μM dsODN-**c**.

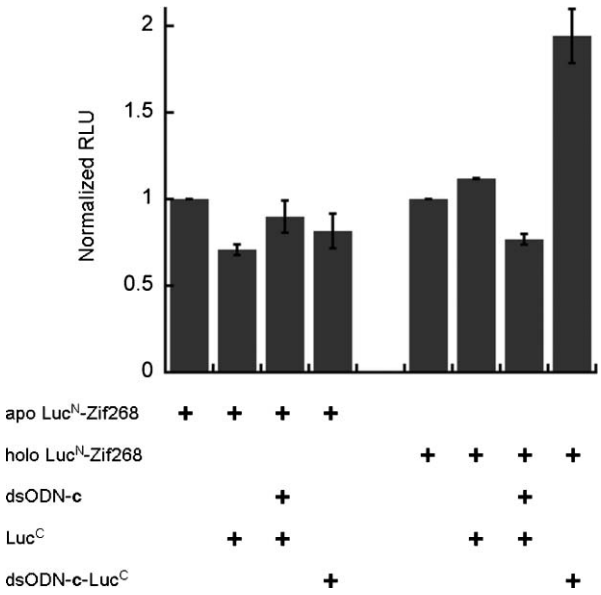
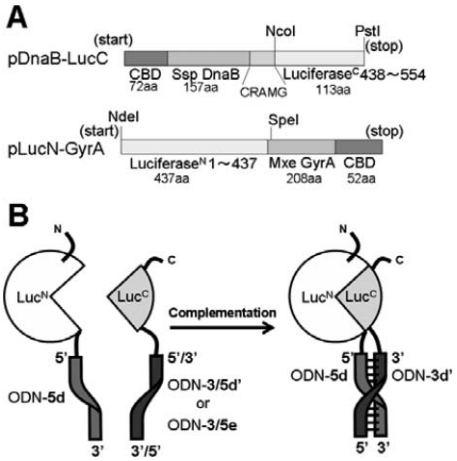


Fig. 4 Split enzyme fragments complementation assay. Conditions: 4 μM apo- or holo-Luc^N-Zif268, 2 μM of each dsODN-**c**–Luc^C, Cys-Luc^C, or free dsODN-**c**, in PBS(−). Each luminescence intensity was normalized against that of the basal Luc^N activity. RLU, relative light unit.

a weak remaining activity, whereas dsODN-**c**–Luc^C displayed no activity. There was no significant difference in the activity of Luc^N-Zif268 between the apo- and the holo-form. Incubation of two hybrids, holo-Luc^N-Zif268 and dsODN-**c**–Luc^C, led to a two-fold enhancement in the luminescent signal.[26] The addition of either dsODN-**c** or Cys-Luc^C to holo-Luc^N-Zif268 did not induce the luciferase activation. Also, apo-Luc^N-Zif268 showed no significant change in the activity in the presence of dsODN-**c**–Luc^C conjugate, dsODN-**c**, or Cys-Luc^C. Taking all results together, we concluded that the functional reconstitution of split luciferase occurs upon the interaction between Zif268 and dsODN-**c** portions in the present semisynthetic split protein system.

DNA hybridization-regulated split luciferase reconstitution

By combining the terminus-specific protein–DNA conjugation methods developed in the present study and previously,[11] it was



Scheme 3 (A) Schematic illustration of the plasmid constructs for complementary DNA hybridization-based protein fragment complementation. (B) Schematic illustration of split enzyme fragment complementation induced by complementary DNA hybridization.

also feasible to control the reconstitution of a split protein by sequence-specific DNA–DNA hybridization. This strategy is shown in Scheme 3. The N-terminal half of split *firefly* luciferase was expressed as a fusion protein with GyrA intein-CBD (IMPACT system). The expressed protein was purified on chitin beads and incubated with 2-mercaptoethanesulfonic acid (MESNA) to obtain it in a C-terminus thioester form (Luc^N-TE). The Luc^N-TE was then modified with a cysteine-tagged single-strand ODN-**5d** by expressed protein ligation as previously described, affording Luc^N–ODN-**5d**. Several hybrids of the C-terminal half of luciferase that have various single-strand DNA (Table 1) at their N-terminus were also prepared as described above (ODN–Luc^C). As shown in Fig. 5, both the N- and C-terminal halves of luciferase could be successfully modified with DNA. After removing excess ODNs by ion-exchange chromatography, size-exclusion filtration, and gel chromatography, the ODN-modified protein fragments were used for the subsequent reconstitution experiments as a mixture with unmodified fragments. It should be noted that a pair of N- and C-terminal halves of unmodified split luciferase fragment shows no recovery of the enzymatic activity (data not shown).

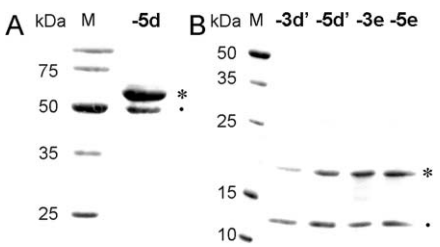


Fig. 5 SDS-PAGE analysis of preparation of ssDNA-modified luciferase fragments. (A) Coomassie staining gel image after ligation of ODN-**5d** with Luc^N-TE. M, molecular weight marker; -**5d**, ligation product. Luc^N-TE and ligation product Luc^N–ODN-**5d** are denoted by · and *, respectively. (B) Coomassie staining gel image of ssDNA-modified Luc^C fragments. M, molecular weight marker. Luc^C and the ligation product ODN-**x**–Luc^C are denoted by • and *, respectively.

DNA hybridization-triggered split luciferase reconstitution was investigated by mixing Luc^N–ODN-**5d** with each of ODN–Luc^C. It was observed that the catalytic activity of Luc^N–ODN-**5d** was enhanced three-times in the presence of ODN-**3d**′–Luc^C of which the ODN sequence is fully complementary to ODN-**5d**′ (Fig. 6).[26] ODN-**3e**–Luc^C and ODN-**5e**–Luc^C containing a non-complementary sequence showed no effect on luciferase activity. Importantly, ODN-**5d**′–Luc^C, which has a complementary sequence attached to the N-terminus of Luc^C but at its 5′-end, did not induce the functional re-assembly due to the reverse orientation of each fragment. These results demonstrate that the DNA sequence- and orientation-directed control of protein function can be achieved in the split protein–DNA hybrid systems.

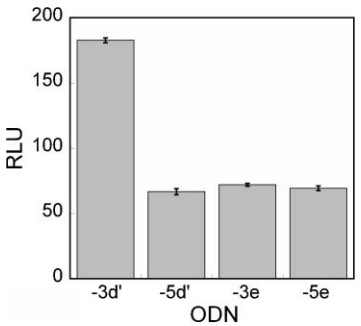


Fig. 6 Recovery of enzymatic activity of split luciferase fragments by re-assembly of complementary DNA hybridization. *Reagents and conditions*: 1 μM Luc^N–ODN-**5d**, 1 μM of each ssDNA-modified Luc^C fragments.

Conclusion

We have described a simple method for conjugating oligonucleotides specifically to the N-terminus of recombinant proteins. The technique involves the thioesterification of amine-terminated oligonucleotides using the bifunctional reagent **3** followed by native chemical ligation with proteins containing a cysteine at their N-terminus. Given the commercial availability of oligonucleotides that have an amine group at the 5′- and 3′-ends as well as internal positions, the present method should be generally applicable to attach proteins to any positions within DNA strands. Furthermore, as a proof-of-principle study, we have constructed two split protein fragment–DNA re-assembly systems that allow the functional reconstitution in response to complementary protein–DNA or DNA–DNA interactions. We have demonstrated that the enzymatic activity of split protein can be precisely regulated in a sequence- and orientation-directed manner. Therefore, in combination with recent advances in DNA-based nanotechnology such as DNA origami, DNA architectures, and DNA-fueled molecular machines,[1,2] the semisynthetic split protein fragment–DNA conjugate may become a platform for the construction of new nanometre-sized functional biomaterials in which the DNA hybridization and protein function is cooperatively coupled. Additionally, the ability to control the re-assembly and thus function of split protein by DNA will provide a new tool for protein engineering and bioanalytical chemistry.

Experimental

Synthesis of bifunctional reagent 3

The starting precursor pentafluorophenyl *S*-benzyl thiosuccinate **1** was synthesized according to the previously reported procedure.[19] To a solution of **1** (0.39 g, 1 mmol) in CH_2Cl_2 (5 ml), DIEA (340 μl, 2 mmol) and sarcosine *tert*-butyl ester hydrochloride (0.18 g, 1 mmol) were added. The reaction solution was stirred for 4 h at room temperature. After evaporation, the residue was re-dissolved in CH_2Cl_2 and washed three times with saturated aqueous citric acid. The organic layer was dried over anhydrous Na_2SO_4, filtered, and evaporated to obtain the compound **2** as yellow oil in 61% yield. ^{1}H NMR: ($CDCl_3$) δ 7.20–7.29 (m, 5H), 4.12 (s, 2H), 4.03 (s, 2H), 3.08 (s, 3H), 2.96 (t, 2H), 2.77 (t, 2H), 1.45 (s, 9H). MALDI-TOF-MS (matrix: dithranol): calcd for $[M + Na]^+ = 374.14$, obsd 374.07.

The compound **2** (636 mg, 1.7 mmol) was dissolved in 20 mL of 50% TFA in CH_2Cl_2. After stirring for 3 h at room temperature, the solvent was removed under reduced pressure. To a solution of the resulting intermediate in DMF (10 mL), *N*-hydroxysuccinimide (230 mg, 2 mmol), DIEA (340 μl, 2 mmol), and HBTU (758 mg, 2 mmol) were added. The mixture was stirred at room temperature for 16 h. After evaporation, the residue was dissolved in $CHCl_3$ and washed with saturated aqueous citric acid. The organic layer was dried over Na_2SO_4, filtered, and evaporated. The crude residue was purified by column chromatography on silica gel with ethyl acetate as the eluent to give **3** as a clear oil (169 mg, 26%). ^{1}H-NMR ($CDCl_3$): δ 7.24–7.29 (m, 5H), 4.51 (s, 2H), 4.14 (s, 2H), 3.13 (s, 2H), 2.97 (t, 2H), 2.87 (s, 4H), 2.76 (t, 2H). MALDI-TOF-MS (matrix: dithranol): calcd for $[M + Na]^+ = 415.09$, obsd 414.98.

N-terminus-specific modification of proteins by oligonucleotides

All oligonucleotides were purchased from Sigma and lyophilized before use. The lyophilized ODN-**a** (93.3 nmol) or -**b** (87.8 nmol) was dissolved in 200 μl of borate buffer (50 mM sodium borate, pH 8.5). To the solution, a solution of **3** in DMF (200 μl, 20 mM) was added. After incubation at room temperature for 3 h, the reaction mixture was applied to a NAP-5 gel filtration column (GE Healthcare Bioscience) to obtain TE-ODNs. TE-ODNs were analyzed by HPLC and MALDI-TOF-MS (matrix: 3-hydroxypicolinic acid). HPLC analyses were carried out on a μBondasphere C18 column (5 micron, 150 × 3.8 mm, Waters) eluted with 0.1 M ammonium acetate buffer (pH 7.0) containing 0–90% acetonitrile with a linear gradient over 60 min at a flow rate of 1 mL min^{-1}, detected at 260 nm. TE-ODN-**a**: $t_R = 14.5$ min. MALDI-TOF-MS (matrix: 3-hydroxypicolinic acid): calcd for $[M + H]^+ = 4047.8$, obsd 4047.8. TE-ODN-**b**; $t_R = 14.1$ min, calcd for $[M + H]^+ = 4156.8$, obsd 4158.3.

Cys-EGFP was obtained as described elsewhere.[20] TE-ODN-**a** or -**b** (4 μl, 500 μM) was mixed with Cys-EGFP (6 μl, 37 μM) in ligation buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA, pH 8.5) and incubated for 48 h at 4 °C. The reaction mixtures were analyzed by SDS-PAGE under reducing conditions.

Construction of a Luc^N-Zif268 expression plasmid

The gene encoding an N-terminal half [1–437] of *firefly* luciferase was amplified by PCR using pGEM-luc (Promega) as a template.

The primer sequences were 5′-CCC GAA TTC ATA TGG AAG ACG CCA AAA ACA TAA AGA AAG GCC C-3′ (Luc^N-R, *Nde*I) and 5′-CCC AAC CAT GGG CGG TCA ACT ATG AAG AAG TGT TCG-3′ (Luc^N-MCS-F, *Nco*I). The amplified DNA fragment and expression vector pTWIN1 (New England Biolabs) were digested with *Nde*I and *Nco*I, purified with a spin column kit (Qiagen), and ligated with Ligation High (Toyobo). The plasmid, pTWIN1-Luc^N-MCS, was cloned and verified by DNA sequencing. The zinc finger domain Zif268 gene was amplified from cDNA of human Egr-1 by PCR using two PCR primers; 5′-CCC CCC ATG GCG CGG CCG CCC ATA TGG AAC GCC CTT ACG CTT GCC CGG TGG AG-3′ (Zif268-R, *Not*I) and 5′-CCC CAA CTA GTG CAT CTC CCG TGA TGC ACA TGT CCT TCT GCC GCA AGT GGA TCT TGG TAT GCC-3′ (Zif268-Met-F, *Spe*I). To suppress the unwanted *in vivo* processing reaction, we added a single methionine residue to the C-terminus of Zif268. The PCR product was digested with *Not*I and *Spe*I and subcloned into pTWIN1-Luc^N-MCS. The plasmid, pLuc^N-Zif268-GyrA, was verified by DNA sequencing.

Bacterial expression, purification, and zinc reconstitution of Luc^N-Zif268

E. coli strain BL21(DE3)pLysS transformed with pLuc^N-Zif268-GyrA was grown overnight at 37 °C in 5 ml LB broth containing 50 μg ml^{-1} ampicilin. Overnight cultures were added to a 250 ml TB medium containing 50 μg ml^{-1} ampicilin and 34 μg ml^{-1} chloramphenicol. At an OD_{660} of 0.6, protein expression was induced by 1 mM IPTG. Cells were cultured at 16 °C overnight and collected by centrifugation. The cell pellet was resuspended in wash buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA, pH 8.5) followed by sonication. After centrifugation, the soluble fraction was loaded onto a chitin column and thoroughly washed with over 10 column volumes of wash buffer. Intein cleavage was induced by incubating the beads with 3 bed volumes of cleavage buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA, 20 mM DTT, pH 8.5). On-column cleavage reaction was allowed to proceed for 24 h at 4 °C, affording Luc^N-Zif268.

The obtained Luc^N-Zif268 was reconstituted with zinc ion in zinc buffer (20 mM Tris-HCl, 100 mM NaCl, 0.5 mM $ZnCl_2$, pH 8.0) to obtain a holo-form of Luc^N-Zif268. Both the apo- and holo-forms of Luc^N-Zif268 were diluted in PBS(−) prior to all assays. The protein concentration was determined by BCA assay.

Plasmid construction and expression of Cys-Luc^C

The gene encoding a C-terminal half [438–554] of *firefly* luciferase was amplified by PCR. The sequence of primers were 5′-CCT TCC ATG GGG TTG AAG TCT TTA ATT AAA TAC AAA GG-3′ (Luc^C-R, *Nco*I) and 5′-GGG AAT TCG GAT CCT TAC AAT TTG GAC TTT CCG CCC TTC TTG GCC-3′(Luc^C-F, *Pst*I). The amplified fragment was digested with *Nco*I and *Pst*I and subcloned into expression vector pTWIN1-MBP1 (New England Biolabs). The plasmid, pDnaB-CRA-Luc^C, was cloned and sequenced.

E. coli strain BL21(DE3)pLysS was transformed with pDnaB-CRA-Luc^C and subjected to protein expression as described above. The soluble fraction of cell lysate was loaded onto a chitin column washed with over 10 column volumes of wash buffer. Intein cleavage was induced by incubating the beads with 3 bed volumes of buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA, pH 7.0). On-column cleavage reaction proceeded at 4 °C for 2 days. The purity and concentration of obtained Cys-Luc^C were analyzed with a BCA assay.

N-terminus-specific modification of Cys-Luc^C with dsODN-c

The 5′-amino-modified oligonucleotide (ODN-**c**: 5′-amino-TTT ACG CCC ACG CTT T-3′) was annealed to the 5′-fluorescein-labeled complementary oligonucleotide (ODN-**c′-fl**: 5′-fluorescein-AAA GCG TGG GCG TAA A-3′). The obtained dsODN containing a Zif268 binding sequence (dsODN-**c**) was then modified with **3** in borate buffer (50 mM sodium borate, pH 8.5). The native chemical ligation reaction procedure between TE-dsODN-**c** and Luc^C was as that previously described.[11] The reaction mixture was then applied to agarose gel electrophoresis for purification. The ligation product, dsODN-**c**–Luc^C, was extracted from the gel and concentrated using a Microcon centrifuge filter kit (Millipore). The concentration of dsODN-**c**–Luc^C was determined by analyzing Coomassie Brilliant Blue-stained SDS-gel using the software Scion Image (Scion Corporation).

Gel shift assay

The fluorescein-labelled dsODN-**c** (10 μM) was mixed with various concentrations of either an apo- or a holo-form of Luc^N-Zif268 in buffer[22] (20 mM Tris-HCl, 100 mM NaCl, 10% Glycerol, 0.1% TritonX-100, 5mM DTT, 0.1mg ml^{-1} BSA; for the holo-form experiments, $MgCl_2$ and $ZnSO_4$ were added at final concentration of 5 mM and 20 μM, respectively), and incubated for several minutes at room temperature. The protein–DNA complexes were separated from free DNA by electrophoresis in a 10% polyacrylamide gel. The fluorescence image of the gel was obtained using a UV transilluminator.

Protein–DNA-interaction-mediated split luciferase reassembly

For split luciferase reconstitution assays, Luc^N-Zif268 (4 μM) was mixed with 2 μM of each dsODN-**c**–Luc^C, Cys-Luc^C, or free dsODN-**c** in 50 μL of PBS(−). After a 30 min incubation at room temperature, a 50 μl of substrate solution (50 mM phosphate, 5 mM $MgSO_4$, 3 mM DTT, 10 mM ATP, 30 μM coenzyme A, 300 μM D-luciferin, pH 7.8) was added. Luminescence intensity was measured for 5 s with a multiwell luminometer AB-2100 (ATTO) and normalized against that of the basal Luc^N activity.

Plasmid construction and expression of Luc^N-TE

The gene encoding an N-terminus half of *firefly* luciferase was amplified by PCR using pGEM-luc as a template. Two primers, (Luc^N-R and Luc^N-F: 5′-GGG CAA CTA GTG CAT CTC CCG TGA TGC ACC GGT CAA CTA TGA AGA AGT GTT CGT CTT CGT CCC-3′) were used. The amplified fragment was digested with *Nde*I and *Spe*I and subcloned into the expression vector pTWIN1-MBP1 (New England Biolabs). The obtained plasmid, pTWIN1-Luc^N-TE, was verified by DNA sequencing.

The expression procedure of Luc^N-TE was the almost same as that of Luc^N-Zif268 described above, except that the intein self-cleavage reaction on the column and elution were induced by thioesterification buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA, pH 8.5).

Preparation of Luc^N–ODN-5d by expressed protein ligation

Cysteine-tagged single-strand ODN-**5d** was prepared and purified by gel chromatography according to our previously described procedure.[11] The obtained ODN was mixed with Luc^N-TE and incubated in ligation buffer at 4 °C for 48 h. The purification of the ligation product was carried out by anion-exchange chromatography on a Poros HQ (Perspective Biosystems) and concentrated by ultra-filtration using Amicon Ultra (MWCO 30kDa) (Millipore). Next, the solution was subjected to gel filtration chromatography on Superdex 75 HR 10/30 (GE Healthcare Bioscience) using PBS(−) as an eluent. The concentration of obtained Luc^N–ODN-**5d** was determined from the Coomassie Blue-stained SDS-gel.

Preparation of a ssODN–Luc^C series

Thioesterification of single-strand oligonucleotides, NCL with Cys-Luc^C, and subsequent purification was carried out as described above to obtain ODN-**3d**′–, ODN-**5d**′–, ODN-**3e**–, and ODN-**5e**–Luc^C. The concentration of these conjugates was determined from a Coomassie Blue-stained SDS-gel.

Re-assembly of split luciferase mediated by DNA–DNA hybridization

In a 96-well plate, Luc^N–ODN-**5d** (1 μM) was mixed with an equimolar of each ssODN–Luc^C in 100 μL of PBS(−) and incubated for 30 min at room temperature. A 100 μl of substrate solution was then added and the luminescence intensity was measured for 5 s as described above.

Acknowledgements

This work was partially supported by NEDO (New Energy and Industrial Technology Development Organization). S. Takeda acknowledges the JSPS Research Fellowships. We thank RIKEN BioResource Centre for providing human Egr-1 cDNA (RDB 1198).

References and notes

1 (*a*) C. M. Niemeyer and D. Blohm, *Angew. Chem., Int. Ed.*, 1999, **38**, 2865–69; (*b*) N. C. Seeman, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 225–248; (*c*) N. C. Seeman, *Trends Biotechnol.*, 1999, **17**, 437–443; (*d*) U. Feldkamp and C. M. Niemeyer, *Angew. Chem., Int. Ed.*, 2006, **45**, 1856–1876; (*e*) P. K. W. Rothemund, *Nature*, 2006, **440**, 297–302.
2 (*a*) B. Yurke, A. J. Turberfield, A. P. Millis, Jr., F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605–608; (*b*) J.-S. Shin and N. A. Pierce, *J. Am. Chem. Soc.*, 2004, **126**, 10834–10835; (*c*) Y. Tian, Y. He, Y. Chen, P. Yin and C. Mao, *Angew. Chem., Int. Ed.*, 2005, **44**, 4355–4358.
3 (*a*) D. S. Wilson and J. W. Szostak, *Annu. Rev. Biochem.*, 1998, **68**, 611–647; (*b*) R. R. Breaker, *Chem. Rev.*, 1997, **97**, 371–390; (*c*) S. Tsukiji, K. Ramaswamy and H. Suga, *Pure Appl. Chem.*, 2004, **76**, 1525–1536.
4 (*a*) C. M. Niemeyer, *Trends Biotechnol.*, 2002, **20**, 395–401; (*b*) C. M. Niemeyer, *Rev. Mol. Biotechnol.*, 2001, **82**, 47–66.
5 D. R. Corey and P. G. Schultz, *Science*, 1987, **238**, 1401–1403.
6 (*a*) S. Howorka, S. Cheley and H. Bayley, *Nat. Biotechnol.*, 2001, **19**, 636–639; (*b*) A. Saghatelian, K. M. Guckian, D. A. Thayer and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2002, **125**, 344–345.
7 (*a*) I. Burbulis, K. Yamaguchi, A. Gordon, R. Carlson and R. Brent, *Nat. Methods*, 2005, **1**, 31–37; (*b*) I. Burbulis, K. Yamaguchi, R. Yu, O. Resnekov and R. Brent, *Nat. Methods*, 2007, **4**, 1011–1013.
8 B. Choi and G. Zocci, *J. Am. Chem. Soc.*, 2006, **128**, 8541–8548.
9 (*a*) C. M. Niemeyer, T. Sano, C. L. Smith and C. R. Cantor, *Nucleic Acids Res.*, 1994, **22**, 5530–5539; (*b*) R. Wacker and C. M. Niemeyer, *ChemBioChem*, 2004, **5**, 453–459; (*c*) J. Tominaga, Y. Kemori, Y. Tanaka, T. Maruyama, N. Kamiya and M. Goto, *Chem. Commun.*, 2007, 401–403.
10 (*a*) C. M. Niemeyer, J. Koehler and C. Wuerdemann, *ChemBioChem*, 2002, **3**, 242–245; (*b*) C. M. Niemeyer, W. Bürger and J. Peplies, *Angew. Chem., Int. Ed.*, 1998, **37**, 2265–2268; (*c*) B. P. Duckworth, Y. Chen, J. W. Wollack, Y. Sham, J. D. Mueller, T. A. Taton and M. D. Distefano, *Angew. Chem., Int. Ed.*, 2007, **46**, 8819–8822.
11 (*a*) S. Takeda, S. Tsukiji and T. Nagamune, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 2407–2410; (*b*) S. Takeda, S. Tsukiji and T. Nagamune, *TetrahedronLett.*, 2005, **46**, 2235–2238.
12 (*a*) M. Lovrinovic, M. Spengler, C. Deutsch and C. M. Niemeyer, *Mol. BioSyst.*, 2005, **1**, 64–69; (*b*) M. Lovrinovic and C. M. Niemeyer, *Biochem. Biophys. Res. Commun.*, 2005, **335**, 943–948; (*c*) M. Lovrinovic, L. Fruk, H. Schröder and C. M. Niemeyer, *Chem. Commun.*, 2007, 353–355; (*d*) M. Lovrinovic and C. M. Niemeyer, *ChemBioChem*, 2007, **8**, 61–67.
13 Niemeyer and coworkers also used the expressed protein ligation method for preparing peptide nucleic acid (PNA)–protein conjugates: M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R. S. Goody and C. M. Niemeyer, *Chem. Commun.*, 2003, 822–823.
14 T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.
15 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, 1994, **266**, 776–779.
16 (*a*) T. J. Magliery, C. G. M. Wilson, W. Pan, D. Mishler, I. Ghosh, A. D. Hamilton and L. Regan, *J. Am. Chem. Soc.*, 2005, **127**, 146–157; (*b*) A. Galarneau, M. Primeau, L.-E. Trudeau and S. W. Michnick, *Nat. Biotechnol.*, 2002, **20**, 619–622; (*c*) J. N. Pelletier, F.-X. Campbell-Valois and S. W. Michnick, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 12141–12146; (*d*) I. Remy and S. W. Michnick, *Mol. Cell. Biol.*, 2004, **24**, 1493–1504; (*e*) S. Cabantous, T. C. Terwilliger and G. S. Waldo, *Nat. Biotechnol.*, 2005, **23**, 102–107; (*f*) N. Johnsson and A. Varshavsky, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10340–10344; (*g*) T. Ozawa, M. Takeuchi, A. Kaihara, M. Sato and Y. Umezawa, *Anal. Chem.*, 2001, **73**, 5866–5874; (*h*) T. Ozawa, Y. Sako, M. Sato, T. Kitamura and Y. Umezawa, *Nat. Biotechnol.*, 2003, **21**, 287–293; (*i*) T. Ozawa, A. Kaihara, M. Sato, K. Tachihara and Y. Umezawa, *Anal. Chem.*, 2001, **73**, 2516–2521; (*j*) A. Kaihara, Y. Kawai, M. Sato, T. Ozawa and Y. Umezawa, *Anal. Chem.*, 2003, **75**, 4176–4181; (*k*) R. Paulmurugan, Y. Umezawa and S. S. Gambhir, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15608–15613; (*l*) R. Paulmurugan and S. S. Gambhir, *Anal. Chem.*, 2003, **75**, 1584–1589; (*m*) R. Paulmurugan and S. S. Gambhir, *Anal. Chem.*, 2005, **77**, 1295–1302; (*n*) C.-D. Hu and T. K. Kerppola, *Nat. Biotechnol.*, 2003, **21**, 539–545; (*o*) M. C. Wehr, R. Laage, U. Bolz, T. M. Fischer, S. Grünewald, S. Scheek, A. Bach, K. Nave and M. J. Rossner, *Nat. Methods*, 2006, **3**, 985–993.
17 T. J. Tolbert and C.-H. Wong, *Angew. Chem., Int. Ed.*, 2002, **41**, 2171–2174.
18 M.-Q. Xu and T. C. Evans, *Methods*, 2001, **24**, 257–277.
19 D. A. Stetsenko and M. J. Gait, *J. Org. Chem.*, 2000, **65**, 4900–4908.
20 B. Lee, S. Takeda, K. Nakajima, J. Noh, J. Choi, M. Hara and T. Nagamune, *Biosens. Bioelectron.*, 2004, **19**, 1169–1174.
21 The present method has also been successfully applied to the conjugation of oligonucleotides to two other proteins, dihydrofolate reductase (DHFR) and *Renilla* luciferase, in yields ranging from 10 to 40% (Fig. S2, ESI†). The ligation yield might be improved by increasing the concentration of recombinant proteins or by using a thiophenol additive: P. E. Dawson, M. J. Churchill, M. R. Ghadiri and S. B. H. Kent, *J. Am. Chem. Soc.*, 1997, **119**, 4325–4329.
22 J.-S. Kim and C. O. Pabo, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 2812–2817.
23 E. Conti, N. P. Franks and P. Brick, *Structure*, 1996, **4**, 287–298.
24 M. W. Southworth, K. Amaya, T. C. Evans, M.-Q. Xu and F. B. Perler, *BioTechniques*, 1999, **27**, 110–120.

25 T. Zako, K. Ayabe, T. Aburatani, N. Kamiya, A. Kitayama, H. Ueda and T. Nagamune, *Biochim. Biophys. Acta*, 2003, **1649**, 183–189.

26 To date, several reports have demonstrated the functional reconstitution of split luciferases (ref. 16*j*–*m*), and some of them achieved over 1000-fold enhancement of catalytic activity upon protein fragment complementation: for a representative example, see ref. 16*m*. However, the increase of luciferase activity in the present system was quite low (approximately 2–4 fold). In this regard, it should be noted that the previous experiments using split luciferase reporters are performed only in live cells or cell lysates, which contain various endogeneous components, such as chaperones, that assist the correct folding and assembly of proteins. In contrast, the present experiments were carried out in a purified *in vitro* system without any chaperone-like components, which might be a reason for the low activity recovery. To our knowledge, no comparable data, *i.e.* split luciferase reconstitution in a purified system, has been reported so far. Also, the catalytic activity enhancement upon reconstitution might be improved by optimizing the spacer length of each fragment and/or split sites.

