JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

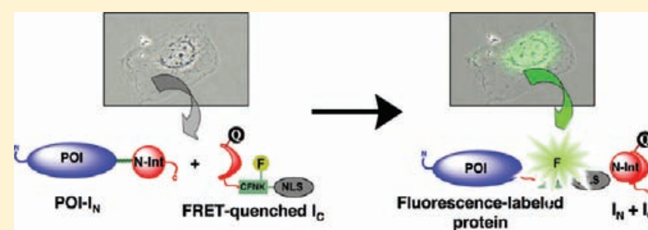
In-Cell Fluorescence Activation and Labeling of Proteins Mediated by FRET-Quenched Split Inteins

Radhika Borra,[†,§] Dezheng Dong,[†,§] Ahmed Y. Elnagar,[†,§] Getachew A. Woldemariam,[†,§] and Julio A. Camarero*[,†,‡]

[†]Department of Pharmacology and Pharmaceutical Sciences, and [‡]Department of Chemistry, University of Southern California, Los Angeles, California 90033, United States

Supporting Information

ABSTRACT: Methods to visualize, track, and modify proteins in living cells are central for understanding the spatial and temporal underpinnings of life inside cells. Although fluorescent proteins have proven to be extremely useful for in vivo studies of protein function, their utility is inherently limited because their spectral and structural characteristics are interdependent. These limitations have spurred the creation of alternative approaches for the chemical labeling of proteins. We report in this work the use of fluorescence resonance emission transfer (FRET)-quenched DnaE split inteins for the site-specific labeling and concomitant fluorescence activation of proteins in living cells. We have successfully employed this approach for the site-specific in-cell labeling of the DNA binding domain (DBD) of the transcription factor YY1 using several human cell lines. Moreover, we have shown that this approach can be also used for modifying proteins to control their cellular localization and potentially alter their biological activity.



INTRODUCTION

Elucidating the distribution, dynamics, and chemical environment of proteins inside living cells is critical for understanding the biomolecular mechanisms of cellular function.[1−3] Labeling of proteins with fluorescent probes or affinity reagents has facilitated in vitro studies of protein structure, dynamics, and protein−protein interactions.[4] However, traditional methods of protein labeling are often inadequate for in vivo studies because they require purification of the protein, chemical labeling, repurification, and reintroduction into cells by invasive methods such as microinjection. These limitations have spawned efforts to noninvasively and site-specifically label proteins in living cells or tissue.

The most prominent method of protein labeling is to genetically encode green fluorescent protein (GFP) or one of its variants as a fusion to the protein of interest.[5,6] Although GFP variants have proven to be extremely useful for in vivo studies of protein function, their utility is somehow inherently limited because of their relatively large sizes, potential for oligomerization, and sometimes slow or incomplete maturation times.

This need for chemically diverse protein labels has led researchers to develop novel ways to label fusion proteins with small molecular probes and/or quantum dots (QDs).[3,7,8] Most of the approaches developed so far for the chemical labeling of proteins in vivo exploit specific, high-affinity noncovalent or covalent interactions between a synthetic ligand and its corresponding receptor. These include hapten−antibody,[9] biotin−avidin,[10] various enzyme−inhibitor combinations,[11−14] nitrilotriacetate (NTA)−oligohistidine sequence,[15] different chemoselective reactions,[16] and Cys-rich peptides that bind biarsenical fluorophores[17] and QDs.[18] Most of these approaches, however, show limited applicability for the simultaneous incorporation of multiple fluorescent tags as well as limited temporal/spatial control for the in vivo labeling process.

One of the most promising approaches for in vivo protein labeling involves the use of intein-mediated protein trans-splicing.[19] Intein-mediated labeling of proteins is highly modular, allowing the covalent site-specific incorporation of a myriad of biophysical probes into proteins.[20−22] The kinetics of protein splicing is also relatively fast, with a number of split inteins having reaction times in the order of several minutes.[21,23,24] Moreover, the recent development of conditional protein splicing, through both chemical and photochemical means, makes possible the chemical modification of proteins in living cells with temporal and spatial control.[25−28] The use of protein trans-splicing for the site-specific labeling of proteins with fluorogenic dyes for in vivo tracking purposes, however, requires that the labeling process must be linked to the simultaneous activation of fluorescence. This is key to the suppression cellular background fluorescence due to the presence of the unreacted intein-fragment (Figure 1). In this work, we report the use of fluorescence resonance emission transfer (FRET)-quenched DnaE split inteins for the site-specific labeling and concomitant fluorescence activation of proteins in living cells. We have successfully used this approach

Received: January 9, 2012
Published: March 10, 2012

 © 2012 American Chemical Society

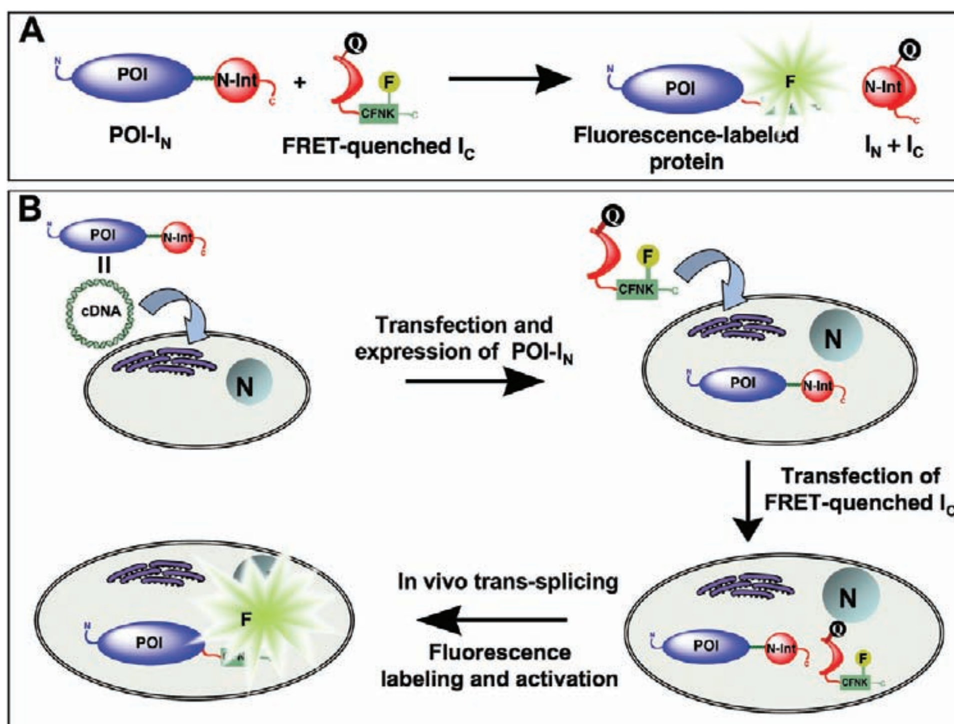


Figure 1. (A) Site-specific labeling and fluorescence activation of a protein of interest (POI) by FRET-quenched protein trans-splicing. Key to this approach is the introduction of fluorescence quencher into the I_C polypeptide, which blocks the fluorescence signal of the fluorophore located at the C-terminus of the I_C polypeptide before protein trans-splicing happens. When protein trans-splicing occurs, the fluorophore is covalently attached to the C-terminus of the POI triggering its fluorescence. (B) The use of this approach for in-cell modification and fluorescence tagging of proteins minimizes the fluorescence background from the unreacted I_C polypeptide, thus facilitating the optical tracking of the labeled protein inside the cell.

for the site-specific in-cell labeling of the DNA binding domain (DBD) of the transcription factor YY1 using several human cell lines. We have also shown that this approach can be easily employed to modify proteins to control their cellular localization and potential biological function.

■ RESULTS AND DISCUSSION

Design and Synthesis of FRET-Quenched DnaE Split Inteins. Intein-mediated protein trans-splicing has been used in a multitude of applications including protein backbone cyclization,[29−32] protein immobilization,[21,33−35] protein semi-synthesis,[36] and segmental isotopic labeling,[37−40] among others. Of particular interest is the use of protein trans-splicing for in vivo site-specific functionalization of proteins.[25−27,41,42] For example, the introduction of biophysical probes such as organic or inorganic fluorescent probes could allow the derivatization of proteins for optical tracking purposes in living cells. To be successful, however, it is key that the fluorescence should only be activated once the trans-splicing reaction has occurred. This will eliminate any detrimental fluorescence background coming from the corresponding unreacted split intein precursor, therefore facilitating the optical tracking of the labeled protein (Figure 1).

In protein trans-splicing, the intein self-processing domain is split in two fragments, called N-intein (I_N) and C-intein (I_C), respectively. These two intein fragments do not have any protein splicing activity individually. However, they can bind each other with high specificity to produce a fully functional protein-splicing domain able to ligate the N- and C-extein segments through a native peptide bond (Figure 1). One of the best-characterized naturally occurring split inteins is the *Syncechocystis* sp. strain PCC6803 (*Ssp*) α-subunit DNA polymerase III (DnaE) intein,[43] which has many known orthologues with high sequence homology in other cyanobacteria species[23,44] (Figure 2C). In the *Ssp* DnaE intein, the I_N and I_C fragments have 123 and 36 residues, respectively. The relatively small size of the I_C fragment facilitates its chemical synthesis, thus allowing the use of synthetic I_C fragments bearing different biophysical probes in the C-extein segment to be used for the chemical modification of proteins through protein trans-splicing.[21,26,27]

We hypothesized that the chemical introduction of a fluorescence quencher on specific locations of the I_C polypeptide could quench, in a reversible fashion, the fluorescence of a suitable fluorogenic probe located in the C-extein fragment, thus rendering any unreacted I_C virtually nonfluorescent. Hence, only after the trans-splicing reaction has occurred and the C-extein fragment has been transferred to the acceptor protein, the fluorescence of the probe would be activated resulting in the concomitant fluorescence activation and labeling of the protein of interest (Figure 1).

To test this hypothesis, we initially used the well-characterized *Ssp* DnaE split intein[43] in combination with fluorescein and dabcyl as fluorescence donor and FRET-quencher, respectively (Figure 2). The fluorescein group was introduced at the C-terminus of the first four residues (Cys-Phe-Asn-Lys) of the C-extein, which are required for efficient trans-splicing.[21,27,45] The dabcyl group was first introduced at the N-terminus of the I_C polypeptide (Q_N, Figure 2). Modification of the N-terminus of the *Ssp* DnaE I_C intein has been shown not to affect the splicing activity of this split intein.[21,24,27] According to the crystal structure of the *Ssp* DnaE split intein,[46] the distance between the fluorophore and the quencher in this case (peptide **3**, Table 1) was estimated to

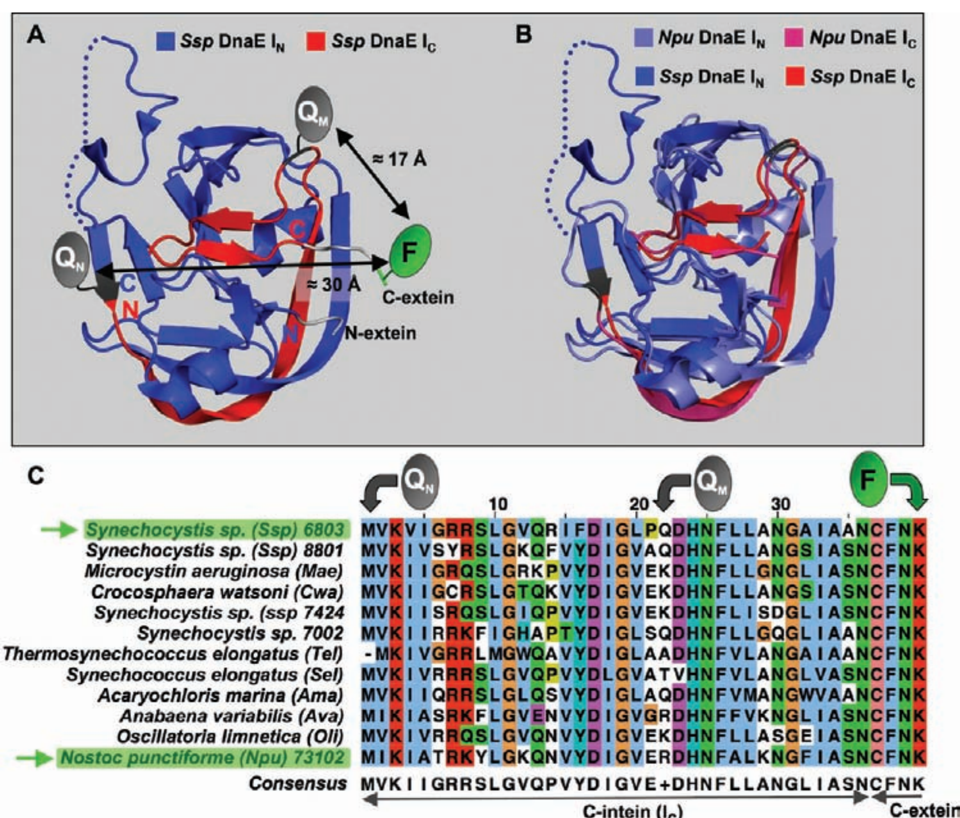


Figure 2. Design of FRET-quenched DnaE split inteins. (A) Crystal structure of the *Ssp* DnaE intein in the prespliced state (PDB code: 1ZDE).[46] DnaE I_C and I_N are shown in red and blue, respectively. The structural secondary elements are also shown. The positions used to place the quencher and fluorophore groups at the I_C and C-extein, respectively (Q_N and Q_M for the N-terminal and side-chain incorporation of fluorescence quencher, respectively), as well as the distances are indicated. (B) Structural alignment of the *Ssp* DnaE (PDB code: 1ZDE) and *Npu* DnaE (PDB code: 2KEQ)[48] split inteins. (C) Multiple sequence alignment of the DnaE I_C for different species indicating the positions used for the introduction of the quencher group in the I_C polypeptide. Multiple sequence alignment was performed using T-Coffee and visualized using Jalview.[59] Molecular representations of the DnaE inteins were generated using the PyMol software package.

Table 1. Sequence of the Different DnaE C-Intein (I_C) Polypeptides Used in This Work[a]

peptide name	compound no.	sequence	molecular weight found (expected[b])/Da	fluorescence intensity/au[c]
SsP Ac-I_C	1	Ac-BVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFNKXKX-NH_2	4753.0 ± 0.1 (4752.6)	ND
Ssp Ac-I_C-Fl	2	Ac-BVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFNKXK (FITC) X-NH_2	5165.1 ± 0.1 (5165.1)	100
Ssp Q_N-I_C-Fl	3	dabcyl-BVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAANCFNKXK (FITC) X-NH_2	5374.0 ± 0.1 (5375.3)	3.7
Ssp Q_M-I_C-F1	4	Ac-BVKVIGRRSLGVQRIFDIGLPK (dabcyl) DHNFLLANGAIAANCFNKXC (Cam-Fl) X-NH_2	5389.0 ± 0.10 (5339.3)	1.3
Npu Ac-I_C	5	Ac-BIKIATRKYLGKQNVYDIGVERDHKFALKNGFIASNCFNKXC (Cam) X-NH_2	5028.0 ± 0.2 (5029.6)	ND
Npu Ac-I_C-Fl	6	Ac-BIKIATRKYLGKQNVYDIGVERDHKFALKNGFIASNCFNKXC (Cam-Fl) X-NH_2	5360.0 ± 0.7 (5360.5]	100
Npu Q_M-I_C-F1	7	Ac-BIKIATRKYLGKQNVYDIGVEK (dabcyl) DHNFALKNGFIASNCFNKXC (Cam-Fl) X-NH_2	5597.0 ± 0.1 (5583.0)[d]	1.4
Npu AC-I_C-NLS-F1	8	Ac-BIKIATRKYLGKQNVYDIGVEKDHNFALKNGFIASNCFNKXC (Cam-Fl) XPKKKRKV-NH_2	6225.8 ± 0.1 (6226.0)	ND
Npu Q_M-I_C-NLS-F1	9	Ac-BIKIATRKYLGKQNVYDIGVEK (dabcyl) DHKFALKNGFIASNCFNKXC (Cam-Fl) XPKKKRKV-NH_2	6449.0 ± 0.14 (6447.6)	ND

[a]Standard single code letters are used for the peptides sequences. Single letter codes B and X stand for norleucine and 6-amino hexanoic acid, respectively. Fluorescein isothiocyanate (FITC) and 5-(iodoacetamide)-fluorescein (IAF) were used to introduce fluorescein into specific Lys or Cys residues, respectively. Dabcyl, Cam, and Cam-Fl stand for 4-dimethylaminoazobenzene-4′-carboxyl, carboxamidomethyl, and fluorescein-carboxamidomethyl, respectively. [b]Average molecular weight. [c]Relative fluorescence intensity to peptide **1**. [d]The difference of +16 Da between the expected and found molecular weight for peptide **7** was attributed to oxidation of the thioether function of carboxamidomethyllated Cys residue used to introduce the fluoresceine group.

be ~30 Å (Figure 2A), well in-range for efficient FRET-quenching. In addition, we also explored introducing the dabcyl group (Q_M, Figure 2) on residue 22 of the *Ssp* DnaE I_C polypeptide (peptide **4**, Table 1). This position is in closer proximity to the

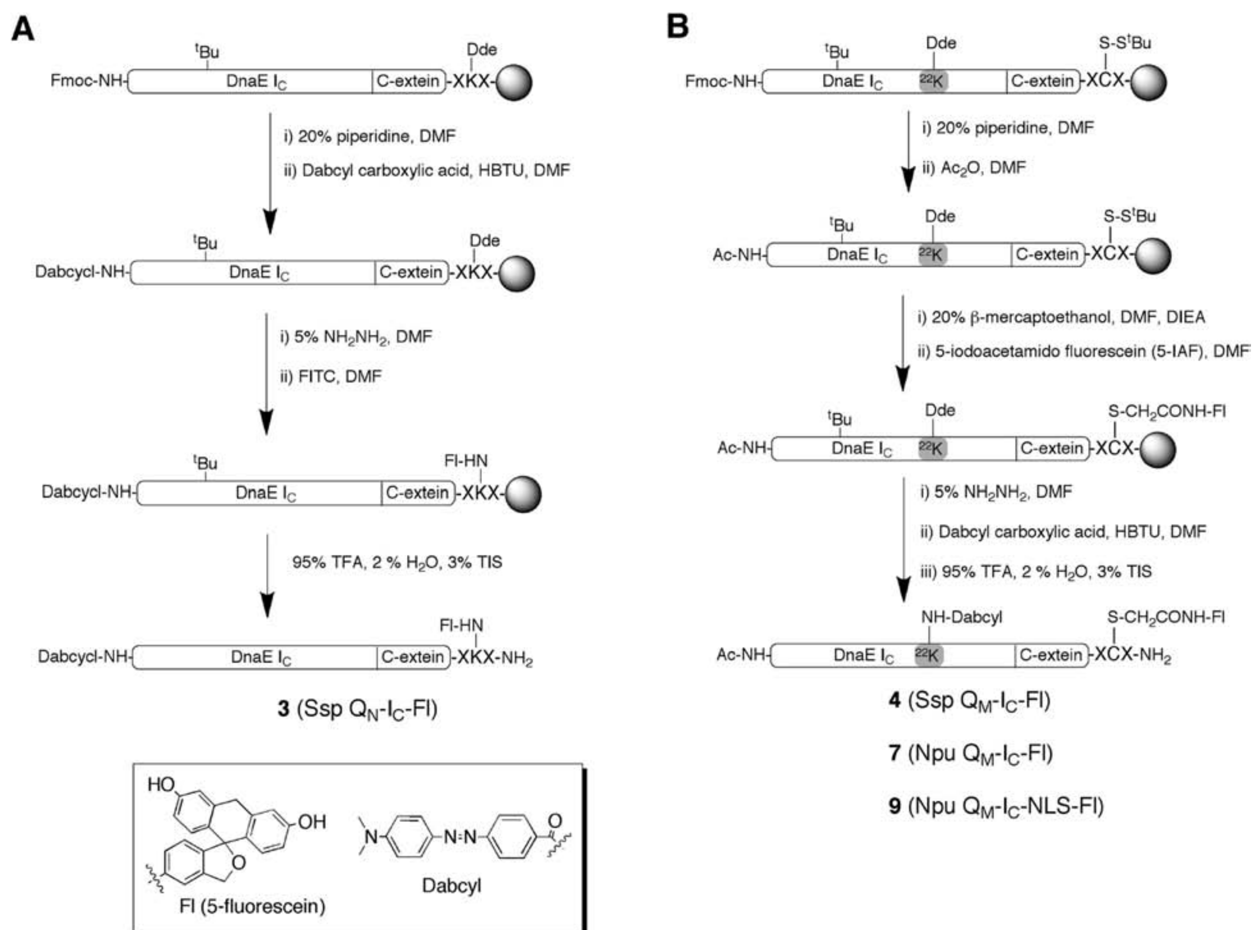


Figure 3. Synthetic approach used for the synthesis of FRET-quenched DnaE I_C polypeptides. All peptides were synthesized by solid-phase peptide synthesis using Fmoc-based chemistry.

C-extein both in the sequence and in the structure of the split intein complex (~17 Å, Figure 2A) and in principle should provide better FRET-quenching than the N-terminal position (peptide **3**). Also, position 22 is not well conserved among other cyanobacterial DnaE I_C split inteins (Figure 2C), suggesting that is not essential for trans-splicing and that the original Gln residue in this position could be replaced by Lys(N^{ε}-dabcyl) with minimal impact of the splicing ability of the modified I_C polypeptide (peptide **4**). We also used this position to prepare a FRET-quenched version of the *Nostoc puntiforme* PCC73102 (*Npu*) DnaE I_C intein (peptide **6**), a highly homologous cyanobacterial DnaE intein (Figure 2B and C).[47,48] This DnaE intein has also be shown to have the highest rate reported for protein trans-splicing ($\tau_{1/2} \approx 60$ s)[24] and a high splicing yield,[24,47] and therefore is a good candidate for the use of protein trans-splicing for in-cell molecular imaging and tracking purposes.

Accordingly, we synthesized peptides **3**, **4**, and **7** (Table 1) using Fmoc-based solid-phase peptide synthesis (SPPS) as depicted in Figure 3. Once the synthesis of peptide **3** was completed, the N^{α}-Fmoc group was deprotected with piperidine and the resulting amino group acylated with dabcyl. Fluorescein was then introduced at the C-terminus of the first four residues of the C-extein through the ε-amino group of a Lys(Dde) residue (Figure 3A and Table 1). This was accomplished by first using hydrazine to remove the Dde group and then acylating the resulting ε-amino group with fluorescein isothiocyanate (FITC). In the case of peptides **4** and **7**, the fluorescein group was introduced first through the thiol group of a Cys(S-^{t}Bu) located at the C-terminus of the first four residues of the C-extein (Figure 3B and Table 1). After removal of the S-^{t}Bu group by treatment with β-mercaptoethanol, the thiol group was alkylated with 5-iodoacetamido fluorescein (5-IAF). The dabcyl group was then introduced on position 22 through the ε-amino group of a Lys(Dde) residue. The Dde protecting group was deprotected as above with hydrazine and acylated with dabcyl.

The *Ssp* and *Npu* wild-type DnaE I_C inteins (peptides **1**, **2**, **5**, and **6**) with and without fluorescein at the C-terminus of the C-extein were also synthesized by Fmoc-SPPS as controls. In all cases, the final cleavage and deprotection of the polypeptides was achieved by treatment with TFA. The peptides were purified and characterized by HPLC and ES−MS (Table 1 and Figure S1).

In Vitro FRET-Quenching of Fluorescein,Dabcyl-Labeled DnaE I_C Polypeptides. First, we estimated the level of fluorescein quenching provided by the dabcyl group at the N-terminus and residue 22 of the Ssp DnaE I_C intein. The dabcyl group was able to efficiently quench the fluorescence of the fluorescein moiety located at the C-extein of the intein in both constructs (Table 1). As anticipated, the fluorescence quenching was slightly more effective when the dabcyl group was located at residue 22 of the DnaE I_C polypeptide (~99%) than when it was placed at the N-terminus (~93%). These results are in good agreement with the distances estimated between the fluorescein and dabcyl groups using the crystal

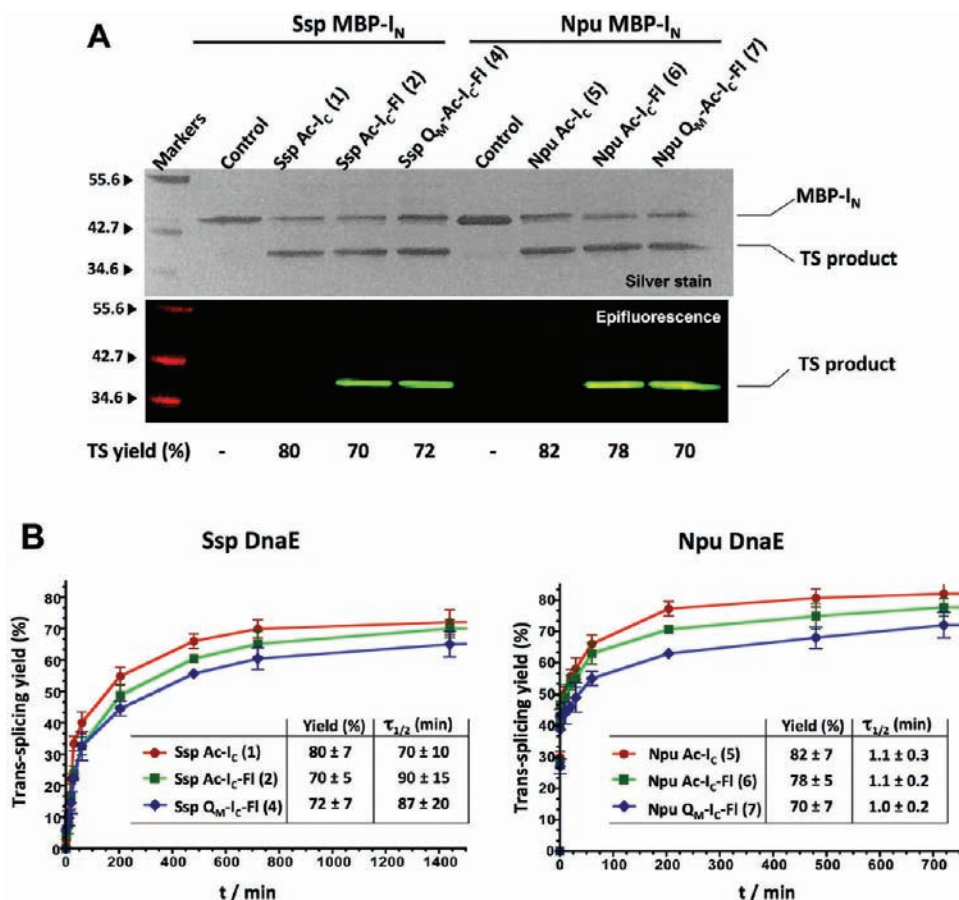


Figure 4. In vitro protein trans-splicing using FRET-quenched DnaE split inteins. (A) SDS-PAGE analysis of the trans-splicing reaction using FRET-quenched I_C polypeptides. Protein detection was performed either by silver staining (top) or by epifluorescence (bottom). Unmodified I_C polypeptides (**1** and **5**) were used as controls. The lane defined as control refers to a reaction containing the corresponding MBP-I_N without addition of any I_C peptide. Trans-splicing reactions between DnaE I_C and MBP-I_N constructs at 1 and 0.1 μM, respectively, were performed in splicing buffer (0.5 mM EDTA, 1 mM TCEP, 50 mM sodium phosphate, 250 mM NaCl buffer, pH 7.2) at room temperature for 24 h. (B) Plot of the fraction of spliced product at different times. Error bars are the standard deviation of three independent experiments.

structure of the *Ssp* DnaE inteins (Figure 2A).[46,48] A similar quenching efficiency (~99%) was also observed for the FRET-quenched *Npu* DnaE I_C intein (Table 1, peptide 7), which is in agreement with the high level of sequence and structural homology between these cyanobacterial split inteins (Figure 2B).

In Vitro Trans-Splicing Activities of Modified DnaE I_C Polypeptides. Next, we evaluated the effect of the position of the dabcyl group on the trans-splicing activity of the different modified DnaE I_C polypeptides (Table 1). We used the maltose binding protein (MBP) fused to the N-terminus of the *Ssp* (MBP-Ssp-I_N) or *Npu* DnaE I_N (MBP-Npu-I_N) inteins as a model protein. The first four residues of the N-extein were also included to facilitate protein trans-splicing.[21,45] The trans-splicing reactions were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) using the corresponding unmodified and fluorescein-labeled I_C polypeptides as controls. As expected, the introduction of the dabcycl group on the N-terminal position of the *Ssp* DnaE intein had no effect on its trans-splicing activity (Figures S4). The modification of the N-terminus of different I_C peptides has been shown not to affect the activity or assembly of the DnaE split intein.[21,24,27] Intriguingly, modification of residue 22 to introduce the fluorescence quencher group had also a minimal effect on the splicing activity of both DnaE split inteins (Figure 4A). In both cases, the trans-splicing yields were similar to those of the corresponding unmodified I_C polypeptides (~80%). Analysis by HPLC and mass spectrometry of the different trans-splicing reactions also confirmed the identity of the trans-spliced product as the C-terminal labeled MPB (Figure S5). It is interesting to note that under the conditions used in this work, no N-cleavage byproduct (i.e., MBP) was detected in the trans-splicing reaction (Figure S5).

The kinetic analysis of the DnaE I_C polypeptides derivatized with the dabcycl group on residue 22 also confirmed that the resulting FRET-quenched I_C inteins had trans-splicing rates similar to those of the corresponding unmodified I_C inteins (Figure 4B). Hence, while peptide **4** (Ssp Q_M-I_C-Fl) showed a slightly slower trans-splicing rate than the wild-type I_C polypeptides ($\tau_{1/2}$ = 87 ± 20 min versus $\tau_{1/2}$ = 70 ± 10 min for the unmodified I_C peptides, Figure 4B), the *Npu* version of this peptide (**7**) showed kinetics identical to that of the wild-type peptides **5** and **6**, with half-life times around 1 min (Figure 4B). The *Npu* DnaE intein is one of the fastest inteins reported so far, and the half-life values found in this work are in agreement with those previously reported.[24,47]

In summary, our data clearly show that modification of residue 22 of the DnaE I_C polypeptide to introduce a fluorescence quencher provides both good quenching yields (~99%) and at the same time has a minimal (*Ssp* DnaE intein) or negligible (*Npu* DnaE) impact on the trans-splicing activity

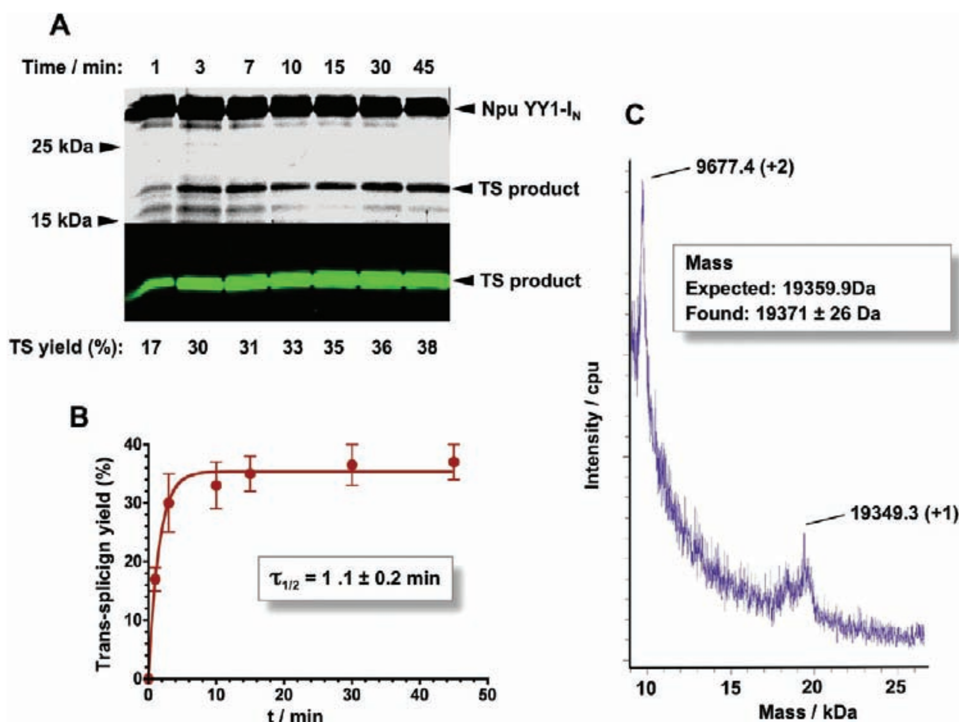


Figure 5. In vitro protein trans-splicing reaction between *Npu* DnaE I_C and YY1-I_N. Trans-splicing reactions between *Npu* DnaE I_C (**6**) and YY1-Npu-I_N constructs at 1 and 0.1 μM, respectively, were performed in splicing buffer (0.5 mM EDTA, 1 mM TCEP, 50 mM sodium phosphate, 250 mM NaCl buffer, pH 7.2) at room temperature for 45 min. (A) SDS-PAGE analysis of the protein trans-splicing reaction. Protein detection was performed by silver staining (top) and epifulorescence (bottom). (B) Plot of the spliced product fraction at different times. (C) MALDI-TOF mass spectrum of trans-spliced YY1 product. Expected mass corresponds to average isotopic mass.

of the corresponding modified I_C inteins. These results, combined with the fast kinetics of the *Npu* DnaE intein, make the modified Npu Q_M-I_C-Fl polypeptide (**7**) the ideal choice to perform in-cell fluorescence activation and labeling of proteins by protein trans-splicing using FRET-quenched split inteins (Figure 1).

In-Cell Trans-Splicing Activity of the FRET-Quenched Npu Q_M-I_C-Fl Polypeptide (7). We tested the splicing activity of peptide **7** (Npu Q_M-I_C-Fl) in cells to evaluate its ability to label proteins with concomitant fluorescence activation. For this purpose, we used the DNA binding domain (DBD) of the transcription factor Yin Yang 1 (YY1). YY1 is a ubiquitously distributed multifunctional transcription factor belonging to the GLI-Kruppel class of zinc finger proteins.[49] The protein is involved in repressing and activating a diverse number of promoters including negative regulation of p53, thus making it of particular interest.[50,51]

First, the DBD of YY1 was fused to the N-terminus of the *Npu* DnaE I_N intein (YY1-Npu-I_N) and recombinantly expressed to test the splicing activity with the Npu I_C intein **6** in vitro. The trans-splicing reaction was monitored by SDS-PAGE at different times, and the proteins were visualized by silver staining and epifluorescence to evaluate the kinetics and yield of the reaction (Figure 5A). The splicing reaction was relatively efficient with a yield ~32% and extremely fast with a half-life that was estimated ~1 min (Figure 5B). The identity of the trans-spliced product was also confirmed by mass spectrometry as the C-terminal labeled DBD of YY1 (Figure 5C). Under these conditions, we were not able to detect any trace of N-cleavage product on the band corresponding to the trans-spliced product.

For in-cell trans-splicing, the protein YY1-Npu-I_N was cloned into a mammalian expression vector and transiently expressed in mammalian cells for 24 h. We demonstrated the generality of this approach by using two different mammalian cell lines, U2OS and HeLa cells. Transient expression of YY1-Npu-I_N was checked in both cell lines by Western blotting reaching an intracellular concentration of ~2 μM after 24 h expression as estimated by Western blot (Figure S6). The cells were then transfected with peptide **7** (Npu Q_M-I_C-Fl) using the commercially available Chariot protein delivery reagent.[52] This system uses the amphipathic peptide Pep-1,[53] which has been shown to be able to deliver cargo polypeptides into several cell lines without the need for covalent coupling.[53,54] Peptides **5** (Npu Ac-I_C) and **6** (Npu Ac-I_C-Fl) were also used as controls. Peptide transfection was carried out for 30 min with final peptide concentrations ranging from 10 to 50 nM in all cases. The best peptide transfection efficiency was obtained for 50 nM I_C using a molecular ratio I_C:Pep-1 of 1:20 (Figure S7). This ratio has been previously described as optimal when using Pep-1.[53] Under these transfection conditions, no cellular cytotoxicity was detected, and the amount of peptide transfected into the cell was relatively constant reaching intracellular concentrations of ~20 μM as estimated by SDS-PAGE (Figure S8). Importantly, the transfection of fluorescent peptide **6** (Npu Ac-I_C-Fl) was able to provide a strong fluorescent signal to the cells, while the FRET-quenched intein peptide **7** was basically nonfluorescent when internalized into cells not expressing the I_N intein fragment (Figure S7). These results confirm the stability of the I_C polypeptides to the intracellular conditions and the efficiency of the FRET-quenched intein to suppress the fluorescence of the fluorescein moiety in the absence of protein trans-splicing.

Figure 6. In-cell trans-splicing reaction using a FRET-quenched *Npu* DnaE split intein. (A) HeLa or U2OS cells were first transiently transfected with a plasmid encoding YY1-Npu-I_N. After 24 h of protein expression, the cells were transfected with different fluorescent-labeled Npu DnaE I_C polypeptides using the Chariot protein delivery system for 60 min. Following the transfection step, fresh DMEM supplemented with serum was added for another 60 min. Cells were then extensively washed and examined by fluorescence microscopy. Bar represents 25 μm. (B) Quantification of labeling yield for in-cell trans-splicing reaction. U2OS cells were first transiently transfected with a plasmid encoding YY1-Npu-I_N, and then with peptide Npu Q_M-I_C-Fl (7) as described above. Cells were collected after 2 h of in-cell trans-splicing reaction, lysed, and the soluble fraction analyzed by SDS-PAGE. Identification of labeled YY1 DBD protein and quantification of in-cell trans-splicing yield was performed by Western blot (right panel) and epifluorescence (left panel), respectively.

The in-cell trans-splicing reaction in HeLa and U2OS cells was monitored by fluorescence microscopy and SDS-PAGE (Figure 6). After 2 h, the cells transfected with the FRET-quenched I_C intein 7 and expressing YY1-Npu-I_N already showed a significant amount of fluorescence (Figure 6A). In-cell reactions longer than 2 h did not significantly increase the level of fluorescence, suggesting that the trans-splicing reaction is essentially completed in 2 h (Figure 6A). This 2 h time

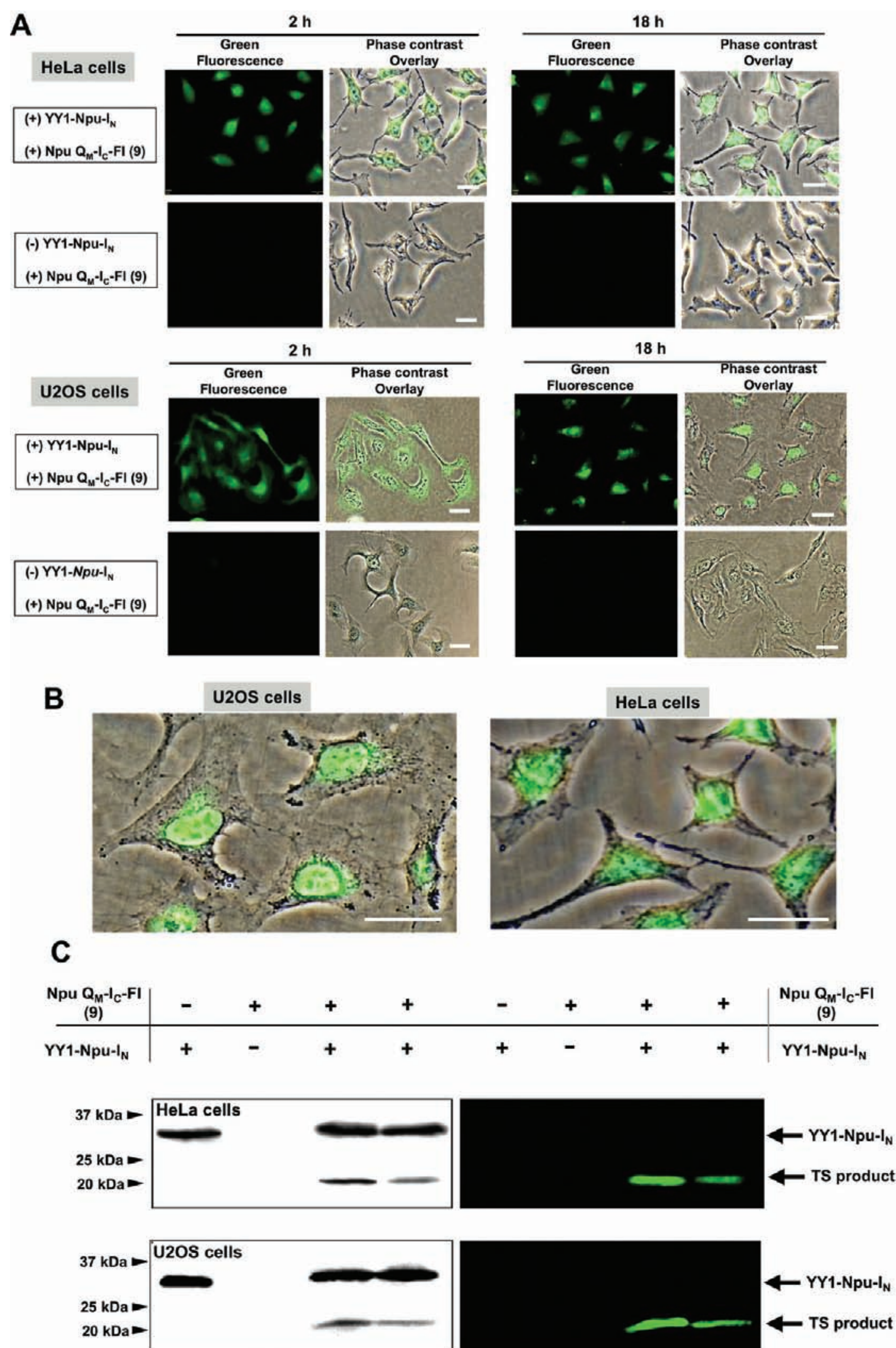


Figure 7. In-cell site-specific labeling of YY1 DBD with a nuclear localization signal and concomitant fluorescence activation using protein trans-splicing. (A) HeLa or U2OS cells were first transiently transfected with a plasmid encoding YY1-Npu-I_N and with different fluorescent-labeled Npu DnaE I_C polypeptides as described in Figure 6. Cells were then extensively washed and examined by fluorescence microscopy. Bar represents 25 μm. (B) Magnification of U2OS and HeLa cells shown in the previous panel after the TS reaction for 18 h with peptide **9** and YY1-Npu-I_N. (C) Quantification of labeling yield for in-cell trans-splicing reaction. Identification of labeled YY1 DBD protein and quantification of in-cell trans-splicing yield was performed by Western blot (right panel) and epifluorescence (left panel), respectively.

dx.doi.org/10.1021/ja300209u | J. Am. Chem. Soc. 2012, 134, 6344−6353

period includes the peptide transfection (30 min) plus cell recovery (90 min). Time optimization of these processes should allow reducing the time required for the in-cell labeling reaction. As expected, the cells transfected only with FRET-quenched I_C intein **7** remained basically nonfluorescent even after 18 h of incubation (Figure 6A). This indicates the requirement of the I_N intein to trigger trans-splicing and activation of fluorescence. Intriguingly, this also highlights the relative high stability of the I_C polypeptide to intracellular degradation. Although we do not have a complete explanation for this unexpected result, it is worth noting that DnaE I_C polypeptides have been described to have a tendency to aggregate in vitro under physiological conditions in the absence of the I_N polypeptide.[21,27] This process could explain the extra intracellular stability of the I_C fragment; however, further studies will be required to fully elucidate the mechanism of such phenomenon. Quantification of the cellular fluorescence signal in cells transfected with YY1-Npu-I_N revealed a ~30-fold increase of fluorescence in both U2OS and HeLa cells when compared to cells transfected only with the FRET-quenched I_C intein **7**.

Analysis by SDS-PAGE using Western blot and epifluorescence of in-cell trans-splicing also verified the labeling of the protein YY1 and confirmed that the trans-splicing reaction was finished (~20% yield) after 2 h incubation for both cell lines (Figures 6B and S9). The identity of the in-cell trans-spliced product was further confirmed by LC/MS/MS (Figure S10). In agreement with the fluorescence results, longer reaction times did not increase the amount of labeled protein (Figure S9). This indicates a good correlation between the level of fluorescence detected in the live cells and the percentage of protein labeled. The in-cell protein trans-splicing yield (~20%) is only slightly smaller than that when the reaction is performed in vitro (~32%). This difference could be attributed to the more oxidative environment of the eukaryotic cellular cytosol.

In-Cell Modification of Protein Localization with Concomitant Fluorescence Activation. Encouraged by these results, we tested this approach to see if it could also be used to modify the intracellular location of a protein with concomitant fluorescence activation to allow optical tracking inside a living cell. We introduced a nuclear localization signal (NLS) and a fluorescent label moiety at the C-terminus of DBD YY1. The DBD of YY1 lacks the original nuclear localization signal found in full length YY1 and therefore when expressed in mammalian cells is mostly distributed in the cytosol (Figure 6). Thus, the simultaneous introduction of a NLS and a fluorescent label into the C-terminus of DBD YY1 should allow the localization and visual tracking of the resulting protein into the nucleus.

We synthesized a FRET-quenched *Npu* I_C intein incorporating a SV40 nuclear localization signal[55] and a fluorescein moiety at the C-terminus of the C-extein (peptide **9**) (Npu Q_M-I_C-NLS-Fl, see Table 1). We also used a nonquenched fluorescein-labeled version of the same peptide (**8**) (Npu Ac-I_C-NLS-Fl, see Table 1) as control. Both peptides were synthesized by Fmoc-based SPPS as described earlier (Figure 3). Peptides **8** and **9** were internalized very efficiently in cells using the Chariot protein delivery reagent as described earlier (Figures S7 and S8). As expected, the I_C polypeptides containing the NLS sequence were localized mainly in the nucleus and perinuclear region of the cell (Figures S7 and S11). In contrast, I_C peptide **6**, which lacks the NLS sequence, showed mostly a cytosolic distribution (Figure S11). In-cell trans-splicing reactions with NLS-containing I_C polypeptides were also performed in U2OS and HeLa cells and monitored by fluorescence microscopy as described before. As shown in Figure 7, the cells transfected with the FRET-quenched I_C polypeptide **9** and expressing YY1-Npu-I_N increased their level of fluorescence after 2 h. In agreement with the previous results, when the same peptide was transfected into cells not expressing the I_N intein construct, the cells remained nonfluorescent (Figure 7A). The in-cell trans-splicing reaction was also complete after 2 h, and increasing the reaction time to 18 h did not raise the level of fluorescence signal (Figure 7) or the labeling efficiency (~20%, Figure S9). More importantly, the fluorescence signal in cells transfected with the FRET-quenched I_C polypeptide **9** and expressing YY1-Npu-I_N accumulated in the nucleus and perinuclear regions (Figure 7). This effect was more pronounced after 18 h, indicating a steady-state between the nuclear import and the diffusion-driven export of the labeled protein. Quantification of the fluorescence signal in these cells also showed a 30-fold increase over the fluorescence found in cells transfected with only with FRET-quenched I_C intein **9** in the absence of protein YY1-Npu-I_N (Figure 7).

CONCLUSIONS

In summary, we have described a new approach for the fast and efficient modification and simultaneous fluorescence labeling of proteins inside living cells using FRET-quenched protein trans-splicing. This new strategy allows the use of protein trans-splicing for the site-specific labeling of proteins with fluorophores and at the same time suppresses the background fluorescence from unreacted I_C polypeptide. This is key for in vivo tracking purposes. We have shown that this approach is general and can be used for the site-specific modification of proteins to alter its cellular localization with concomitant fluorescent activation. As shown in this work, FRET-quenched I_C polypeptides containing small- or medium-sized fluorescent-labeled C-exteins can be easily accessed chemically through SPPS. I_C polypeptides containing longer C-extein fragments could be also readily obtained through semisynthesis by ligation of the quenched-labeled I_C peptide thioester to a recombinantly obtained fluorescent-modified C-extein through native chemical ligation.[56,57] Site-specific labeling of recombinant fragments can be accomplished using standard intein expression vectors.[57] Hence, the use of this technique in combination with synthetic or semisynthetic FRET-quenched I_C polypeptides should allow the modification of proteins without limit on the size of the polypeptide fragments to be introduced. This could range from small regulatory peptide sequences, such as a NLS sequence shown in this study, to larger protein domains required for the full biological activity of the resulting protein.

This approach could also be readily adapted to other naturally or artificially split inteins.[58] The use of different orthogonal split inteins should in principle allow the simultaneous labeling of proteins with different fluorescent probes inside living cells for in-cell multicolor optical tracking. Moreover, the recent development of photomodulated protein trans-splicing[26−28] should make feasible the use of photocaged FRET-quenched inteins in vivo with spatial and temporal control through highly focused and coherent light sources, which should allow simultaneous in vivo tracking of the modified proteins.

ASSOCIATED CONTENT

Supporting Information

Experimental details, synthesis, purification and characterization of I_C peptides, cloning expression and purification of I_N proteins, in vitro and in vivo protein splicing protocols, fluorescence spectroscopy and microscopy, and cell culture. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

jcamarer@usc.edu

Author Contributions

§These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Research Grant R01-GM090323 (J.A.C.) and by the Department of Defense Congressionally Directed Medical Research Program Grant PC09305 (J.A.C.).

REFERENCES

(1) Xie, X. S.; Yu, J.; Yang, W. Y. *Science* **2006**, *312*, 228–30.
(2) Chen, I.; Ting, A. Y. *Curr. Opin. Biotechnol.* **2005**, *16*, 35–40.
(3) Miller, L. W.; Cornish, V. W. *Curr. Opin. Chem. Biol.* **2005**, *9*, 56–61.
(4) Rao, J.; Dragulescu-Andrasi, A.; Yao, H. *Curr. Opin. Biotechnol.* **2007**, *18*, 17–25.
(5) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269–72.
(6) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 906–18.
(7) Marks, K. M.; Nolan, G. P. *Nat. Methods* **2006**, *3*, 591–6.
(8) Johnsson, N.; Johnsson, K. *ACS Chem. Biol.* **2007**, *2*, 31–8.
(9) Farinas, J.; Verkman, A. S. *J. Biol. Chem.* **1999**, *274*, 7603–6.
(10) Howarth, M.; Takao, K.; Hayashi, Y.; Ting, A. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7583–8.
(11) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. *ACS Chem. Biol.* **2008**, *3*, 373–82.
(12) Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9955–9.
(13) Marks, K. M.; Braun, P. D.; Nolan, G. P. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9982–7.
(14) Calloway, N. T.; Choob, M.; Sanz, A.; Sheetz, M. P.; Miller, L. W.; Cornish, V. W. *ChemBioChem* **2007**, *8*, 767–74.
(15) Guignet, E. G.; Hovius, R.; Vogel, H. *Nat. Biotechnol.* **2004**, *22*, 440–4.
(16) Bertozzi, C. R. *Acc. Chem. Res.* **2011**, *44*, 651–3.
(17) Martin, B. R.; Giepmans, B. N.; Adams, S. R.; Tsien, R. Y. *Nat. Biotechnol.* **2005**, *23*, 1308–14.
(18) Michalet, X.; Pinaud, F. F.; Bentolila, L. A.; Tsay, J. M.; Doose, S.; Li, J. J.; Sundaresan, G.; Wu, A. M.; Gambhir, S. S.; Weiss, S. *Science* **2005**, *307*, 538–44.
(19) Perler, F. B. *IUBMB Life* **2005**, *57*, 469–76.
(20) Shi, J.; Muir, T. W. *J. Am. Chem. Soc.* **2005**, *127*, 6198–206.
(21) Kwon, Y.; Coleman, M. A.; Camarero, J. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 1726–1729.
(22) Kurpiers, T.; Mootz, H. D. *ChemBioChem* **2008**, *9*, 2317–25.
(23) Dassa, B.; Amitai, G.; Caspi, J.; Schueler-Furman, O.; Pietrokovski, S. *Biochemistry* **2007**, *46*, 322–30.
(24) Zettler, J.; Schutz, V.; Mootz, H. D. *FEBS Lett.* **2009**, *583*, 909–14.
(25) Mootz, H. D.; Blum, E. S.; Tyszkiewicz, A. B.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 10561–9.
(26) Vila-Perello, M.; Hori, Y.; Ribo, M.; Muir, T. W. *Angew. Chem., Int. Ed.* **2008**, *47*, 7764–7.
(27) Berrade, L.; Kwon, Y.; Camarero, J. A. *ChemBioChem* **2010**, *11*, 1368–72.
(28) Binschik, J.; Zettler, J.; Mootz, H. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 3249–52.
(29) Evans, T. C. Jr.; Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X. Q.; Xu, M. Q. *J. Biol. Chem.* **2000**, *275*, 9091–4.
(30) Abel-Santos, E.; Scott, C. P.; Benkovic, S. J. *Methods Mol. Biol.* **2003**, *205*, 281–294.
(31) Camarero, J. A.; Kimura, R. H.; Woo, Y. H.; Shekhtman, A.; Cantor, J. *ChemBioChem* **2007**, *8*, 1363–6.
(32) Austin, J.; Wang, W.; Puttamadappa, S.; Shekhtman, A.; Camarero, J. A. *ChemBioChem* **2009**, *10*, 2663–70.
(33) Girish, A.; Sun, H.; Yeo, D. S.; Chen, G. Y.; Chua, T. K.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2447–51.
(34) Camarero, J. A.; Kwon, Y.; Coleman, M. A. *J. Am. Chem. Soc.* **2004**, *126*, 14730–1.
(35) Camarero, J. A. *Biopolymers* **2008**, *90*, 450–8.
(36) Mootz, H. D. *ChemBioChem* **2009**, *10*, 2579–89.
(37) Camarero, J. A.; Shektman, A.; Campbell, E.; Chlenov, M.; Gruber, T. M.; Bryant, D. A.; Darst, S. A.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8536–8541.
(38) Muralidharan, V.; Muir, T. W. *Nat. Methods* **2006**, *3*, 429–38.
(39) Iwai, H.; Zuger, S. *Biotechnol. Genet. Eng. Rev.* **2007**, *24*, 129–45.
(40) Flavell, R. R.; Muir, T. W. *Acc. Chem. Res.* **2009**, *42*, 107–16.
(41) Dhar, T.; Mootz, H. D. *Chem. Commun.* **2011**, *47*, 3063–5.
(42) Shah, N. H.; Vila-Perello, M.; Muir, T. W. *Angew. Chem., Int. Ed.* **2011**, *50*, 6511–5.
(43) Wu, H.; Hu, Z.; Liu, X. Q. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9226–9231.
(44) Wei, X. Y.; Sakr, S.; Li, J. H.; Wang, L.; Chen, W. L.; Zhang, C. C. *Res. Microbiol.* **2006**, *157*, 227–34.
(45) Martin, D. D.; Xu, M. Q.; Evans, T. C. Jr. *Biochemistry* **2001**, *40*, 1393–402.
(46) Sun, P.; Ye, S.; Ferrandon, S.; Evans, T. C.; Xu, M. Q.; Rao, Z. *J. Mol. Biol.* **2005**, *353*, 1093–105.
(47) Iwai, H.; Zuger, S.; Jin, J.; Tam, P. H. *FEBS Lett.* **2006**, *580*, 1853–8.
(48) Oeemig, J. S.; Aranko, A. S.; Djupsjobacka, J.; Heinamaki, K.; Iwai, H. *FEBS Lett.* **2009**, *583*, 1451–6.
(49) Wang, C. C.; Chen, J. J.; Yang, P. C. *Expert Opin. Ther. Targets* **2006**, *10*, 253–66.
(50) Gronroos, E.; Terentiev, A. A.; Punga, T.; Ericsson, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12165–70.
(51) Sui, G.; Affarel, B.; Shi, Y.; Brignone, C.; Wall, N. R.; Yin, P.; Donohoe, M.; Luke, M. P.; Calvo, D.; Grossman, S. R. *Cell* **2004**, *117*, 859–72.
(52) Chariot(tm), A. M. o. N. A. h. w.; activemotif.com.
(53) Morris, M. C.; Depollier, J.; Mery, J.; Heitz, F.; Divita, G. *Nat. Biotechnol.* **2001**, *19*, 1173–6.
(54) Henriques, S. T.; Costa, J.; Castanho, M. A. *Biochemistry* **2005**, *44*, 10189–98.
(55) Miyamoto, Y.; Imamoto, N.; Sekimoto, T.; Tachibana, T.; Seki, T.; Tada, S.; Enomoto, T.; Yoneda, Y. *J. Biol. Chem.* **1997**, *272*, 26375–81.
(56) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779.
(57) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
(58) Volkmann, G.; Iwai, H. *Mol. Biosyst.* **2010**, *6*, 2110–21.
(59) Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J. *Bioinformatics* **2004**, *20*, 426–7.