J|A|C|S COMMUNICATIONS

Site-Specific Two-Color Protein Labeling for FRET Studies Using Split Inteins

Jin-Yi Yang and Wei Yuan Yang*

Institute of Biological Chemistry, Academia Sinica and Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan

Received April 15, 2009; E-mail: weiyang@gate.sinica.edu.tw

Fluorescence resonance energy transfer (FRET) is an ideal tool for studying dynamic biomolecular motions and associations.[1−3] In FRET measurements, the efficiency (E) of energy transfer between two fluorescent dye molecules reports the intervening distance R ($E = [1 + (R/R_o)^6]^{-1}$; R_o is a dye-pair dependent constant and is usually between 2 and 8 nm). By conjugating dyes at designated locations on individual molecules or biomolecular complexes, the distance between those specific locations can be dynamically monitored through FRET. Coupling nanometer sensing capabilities and single molecule sensitivity, FRET has enabled advances in numerous research areas including DNA/RNA folding and catalysis,[3,4] protein folding,[5] and protein−protein and protein−nucleic acid interactions.[6]

The potential of FRET for monitoring protein conformational changes and dynamics, however, has yet to be fully realized. While fluorescent dyes can be incorporated at desired locations on short peptides through synthesis, placing a fluorophore pair onto specific and structurally important sites of a protein containing more then 100 amino acids remains challenging. Existing successes of labeling a single protein with multiple colors (fluorophores) have been devised on a case-by-case basis.[3,7,8] For single color site-specific fluorophore conjugation on a protein, cysteine−maleimide chemistry is most often used. Proteins are first mutated so that they contain only a single cysteine residue for maleimide-linked fluorophore attachment. Expansion of the cysteine/maleimide system (or any other single color labeling schemes) to accommodate multicolor site-specific protein labeling (required for FRET), however, is not trivial.

Herein we report a general site-specific two-color protein labeling method through the expansion of cysteine-based strategies. The scheme relies on generating protein fragments that contain a single cysteine for fluorophore conjugation. The fragments after labeling are joined together to generate a full-length protein through split intein mediated protein splicing.

Inteins are peptide sequences that mediate self-splicing when inserted within a larger precursor polypeptide.[9,10] During splicing, intein removes itself from the precursor sequence and joins the flanking N- and C-terminal parts into a new protein. In nature, inteins have been found to splice together enzymes including polymerases, proteases, and ATPases. Intein can also exist in split geometry: two halves of an intein (an N-terminal half, IntN, and a C-terminal half, IntC) complement each other spontaneously in solution to form a functional unit.

Two constructs are needed to perform our site-specific two-color protein labeling scheme using split inteins (Figure 1A, **1** and **2**). The protein of interest is first divided at a serine residue into two fragments (abbreviated as ProtN and ProtC, respectively, with the serine becoming the first residue of ProtC). ProtN and ProtC should both carry a cysteine for fluorophore conjugation, and each fragment is joined to one-half of the split intein, giving rise to two fusion proteins, ProtN-IntN-tag and tag-IntC-ProtC. The additional tags within the fusion constructs serve to enhance protein stability, solubility, and the ease of purification. Split inteins' activities can be modulated by the first few residues in ProtC. Oftentimes, a cysteine immediately following IntC (the first residue in ProtC) is required to serve as a nucleophile to drive split intein actions. However, some split inteins maintain activity when this cysteine is replaced by serine,[11,12] thus our design contains a serine residue at the start of ProtC. We selected the DnaE split intein from *Nostoc punctiforme* (*Npu*) to use in our labeling procedure, as it maintains good activity regardless of the second amino acid identity in ProtC[13] and can, therefore, be efficiently utilized on many varieties of proteins of interest.

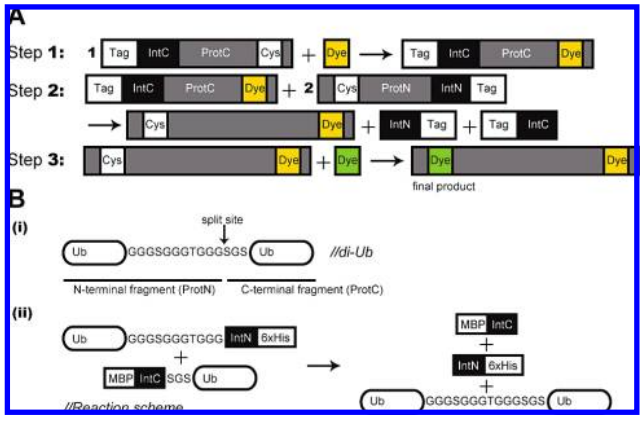

Figure 1. Schematic diagram for dual-fluorophore protein labeling. (A) Protein to be labeled is split into ProtN and ProtC halves at an internal serine residue and expressed as tag-IntC-ProtC and ProN-IntN-tag chimeras. A three-step process follows for generating the dual-labeled protein (Steps 1−3). (B) (i) di-Ub contained two ubiquitin domains joined by a glycine-rich linker. For dual labeling, di-Ub was split at the 2nd serine within the linker region. (ii) When mixed, Ub-IntN-6His and MBP-IntC-Ub, two constructs required for dual labeling di-Ub, generate spliced products MBP-IntC, IntN-6xHis, and di-Ub.

We employ a three-step procedure to label our protein of interest with the desired fluorophore pair (figure 1A). Step 1: the cysteine residue within tag-IntC-ProtC is first labeled with the desired fluorophore. ProtN-IntN-tag remains unlabeled at this step, as the first residue in IntN of split inteins is a catalytically important cysteine residue, and its reaction with maleimide-linked fluorophores will compromise function.[13] Step 2: Following successful incorporation of the first fluorophore, ProtN-IntN-tag and tag-IntC-ProtC are mixed together to trigger protein splicing: this produces the protein of interest (ProtN-ProtC) with an attached dye on ProtC and one unreacted cysteine on ProtN. Step 3: Lastly, our protein of interest undergoes conjugation of a second fluorophore, allowing for intraprotein FRET!

We first confirmed that *Npu* DnaE split intein-mediated protein splicing can occur efficiently when serine is designated as the first

residue in ProtC, as this is required for our labeling scheme. Tethered ubiquitin was used as the protein for validation; it consisted of two ubiquitin domains joined together with a 14 amino acid glycine-rich flexible linker that resulted in an ubiquitin dimer (di-Ub, Figure 1B). di-Ub was divided at the second serine within the flexible linker region to generate the two constructs needed: Ub-IntN-6xHis and MBP-IntC-Ub (Figure 1B; MBP = maltose binding protein). As a comparison, we generated control constructs in which the first serine in ProtC is replaced with CFNGT, the native amino acid sequence that flanks IntC in *Npu* DnaE (detailed in Supporting Information Figure S1). Splicing kinetics was assayed through the formation of di-Ub molecules by Western blotting (Figure 2A and B, against ubiquitin; corresponding SDS-PAGE in Figure S2A and B). While the native cysteine-based (CFNGT) reaction occurred at a remarkable rate (<1 h half-life), the serine based splicing was nevertheless effective (half-life $\approx$ 16 h; Figure 2C). More importantly, the amount of spliced di-Ub (correct molecular weight, Figure S3D) for the serine based construct remained high despite the slower kinetics.

We went on to perform dual fluorophore labeling on di-Ub, which allows for FRET and can be used to report the distance between ubiquitin domains as well as the configuration of the 14-aa flexible linker within di-Ub. We mutated the 47th amino acid (originally a Gly) on the N-terminal ubiquitin domain and the 7th amino acid (originally a Thr) on the C-terminal ubiquitin domain into cysteines (Figure 2D), two sites not buried within ubiquitin's hydrophobic core. When a fluorophore pair with a short known R_o was used, (Alexa Fluor 488/647, $R_o \approx 5.6$ nm), di-Ub displayed minimal FRET (Figure 2E). Consistent with this observation, labeling with a longer R_o fluorophore pair (Alexa Fluor 594/647, $R_o \approx 8.5$ nm) resulted in significant FRET (excitation = 545 nm, to avoid direct excitation of Alexa Fluor 647; homogeneous labeling, Figure S3). The long apparent distance between the two fluorophores suggests that excluded volume effects exerted by the two terminal ubiquitins in di-Ub stretches the 14-aa flexible linker (detailed in Supporting Information Figure S4).

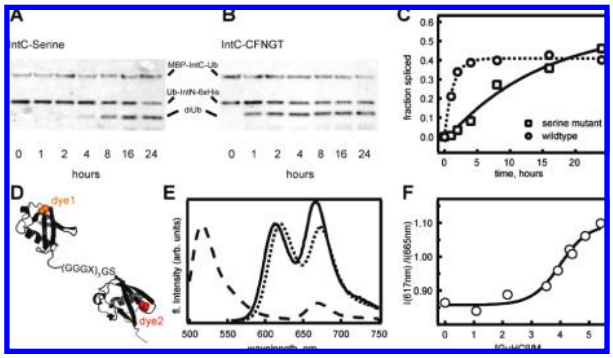

Figure 2. Dual-fluorophore labeled di-Ub. (A and B) Splicing between Ub-IntN-6xHis and MBP-IntC-Ub (molar ratio = 1:3), monitored by Western blotting (anti-Ub antibody). (C) Splicing kinetics for constructs containing either a serine (□) or a CFNGT sequence (○) following IntC. (D) di-Ub topology with fluorophores at the 47th residue (orange) and the 7th residue (red) of the N- and C-terminal ubiquitin domains, respectively. X = S/T. (E) di-Ub fluorescence. Alexa Fluor 488/647 labeled di-Ub in 0 M GuHCl (dashed, excitation = 490 nm), Alexa Fluor 594/647 labeled di-Ub in 0 M GuHCl (solid, excitation = 545 nm), or Alexa Fluor 594/647 labeled di-Ub in 5.4 M GuHCl (dotted, excitation = 545 nm). (F) The ratio between 617 nm (Alexa Fluor 594) and 665 nm (Alexa Fluor 647) emission (excitation = 545 nm) as a function of GuHCl concentration.

FRET also allowed us to monitor the unfolding of di-Ub(s) in guanidine hydrochloride (GuHCl, a protein denaturant). The intramolecular FRET decreased in increased GuHCl concentrations (Figure 2F). The unfolding transition for di-Ub resembled that of monoubiquitins (ubiquitins' unfolding midpoint lies at ~4 M GuHCl), indicating that labeling retained the native properties of the ubiquitin domains. Individual ubiquitin domain unfolding resulted in overall expansion of di-Ub topology and distance lengthening between the two dyes.

Our site-specific dual-color protein labeling method using split intein has several major advantages. It can be applied directly to a wide range of proteins (naturally cysteine-free or made cysteine-free) and allows for FRET based protein conformational monitoring at both the ensemble and single molecule level. The procedures are fast, low in cost, and easy to carry out. The requirement of segmenting a protein at an internal serine site can also be easily satisfied, as serines are abundant in protein sequences and are frequently located in loops that serve as convenient sites for protein splitting. A potential limitation for the method is the instability of the split protein fragments, which can sometimes cause low expression. However, this scheme allows for the incorporation of protein tags that increase protein stability (i.e., MBP, GST, etc.) and solubility, which also results in easier protein handling during expression, purification, labeling, and splicing without affecting the final labeled protein. In the future, newly discovered, engineered, or evolved split inteins may permit other nucleophilic residues such as threonines to be used with this labeling scheme, further relaxing the sequence constraint. Our strategy also has the potential of incorporating the recently discovered s11 intein from *Ssp GyrB* to allow for protein labeling with more than three colors on a single protein for more comprehensive FRET experiments.[11,14,15] This demonstrated the use of split intein will also be applicable toward additional labeling methods other then cysteine−maleimide based chemistry.

Acknowledgment. Funding was provided by the NSC-97-2113-M-001-032-MY2 grant from the National Science Council in Taiwan. We thank Dr. Shu Yuan Yang for proofreading the manuscript.

Supporting Information Available: Supporting Figures S1−S4 and details of the experimental methods. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Weiss, S. *Nat. Struct. Biol.* **2000**, *7*, 724–729.
(2) Selvin, P. R. *Nat. Struct. Biol.* **2000**, *7*, 730–734.
(3) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507–516.
(4) Zhuang, X. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 399–414.
(5) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.
(6) Kapanidis, A. N.; Laurence, T. A.; Lee, N. K.; Margeat, E.; Kong, X.; Weiss, S. *Acc. Chem. Res.* **2005**, *38*, 523–533.
(7) Jager, M.; Michalet, X.; Weiss, S. *Protein Sci.* **2005**, *14*, 2059–2068.
(8) Kapanidis, A. N.; Weiss, S. *J. Chem. Phys.* **2002**, *117*, 10953–10964.
(9) Xu, M. Q.; Evans, T. C., Jr. *Curr. Opin. Biotechnol.* **2005**, *16*, 440–446.
(10) Muralidharan, V.; Muir, T. W. *Nat. Methods* **2006**, *3*, 429–438.
(11) Appleby, J. H.; Zhou, K.; Volkmann, G.; Liu, X. Q. *J. Biol. Chem.* **2009**, *284*, 6194–6199.
(12) Nichols, N. M.; Evans, T. C., Jr. *Biochemistry* **2004**, *43*, 10265–10276.
(13) Iwai, H.; Zuger, S.; Jin, J.; Tam, P. H. *FEBS Lett.* **2006**, *580*, 1853–1858.
(14) Shi, J.; Muir, T. W. *J. Am. Chem. Soc.* **2005**, *127*, 6198–6206.
(15) Hohng, S.; Joo, C.; Ha, T. *Biophys. J.* **2004**, *87*, 1328–1337.

JA9030215