Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 17 (2009) 981–989
Bioorganic & Medicinal Chemistry

Site-specific covalent labeling of proteins inside live cells using small molecule probes

Souvik Chattopadhaya,[a] Rajavel Srinivasan,[b] Dawn S. Y. Yeo,[a] Grace Y. J. Chen[a,b] and Shao Q. Yao[a,b,*]

[a]*Department of Biological Sciences, National University of Singapore, 3 Science Drive 3, Singapore 117543, Republic of Singapore*
[b]*Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543, Republic of Singapore*

Received 14 November 2007; revised 11 January 2008; accepted 24 January 2008
Available online 30 January 2008

Abstract—The study of dynamic movement and interactions of proteins inside living cells in real time is critical for a better understanding of cellular mechanisms and functions in molecular detail. Genetically encoded fusions to fluorescent protein(s) (FP) have been widely used for this purpose [*Annu. Rev. Biochem*. **1998**, *67*, 509–544]. To obviate some of the drawbacks associated with the use of FPs [*Curr. Opin. Biotechnol*. **2005**, *16*, 1–6; *Nat. Methods* **2006**, *3*, 591–596], we report a small molecule-based approach that exploits the unique reactivity between the cysteine residue at the N-terminus of a target protein and cell-permeable, thioester-based small molecule probes resulting in site-specific, covalent tagging of proteins. This approach has been demonstrated by the in vivo labeling of proteins in both bacterial and mammalian systems thereby making it potentially useful for future bioimaging applications.

1. Introduction

Small molecule-based chemical labeling strategies for site-specific tagging of proteins have provided powerful tools for the study of protein functions in intact cells.[1–23] These include the formation of stable complexes between biarsenical compounds and tetracysteine-containing proteins,[4,5] non-covalent interactions between small molecule ligands with proteins containing the cognate tag,[6–9] enzyme- mediated covalent conjugation of small molecules to protein domains/peptide tags fused to target proteins,[10–15] intein-mediated protein semisynthesis,[16–19] incorporation of unnatural amino acids during protein biosynthesis,[20,21] and more recently the metabolic installation of bioorthogonal chemical reporter tags by using the cell's biosynthetic machinery.[22,23] We and others had previously developed small molecule probes that specifically react with N-terminal cysteine-containing proteins in vitro.[24–27] Herein, we report the first use of this approach for in vivo tagging and visualization of proteins in both bacterial and mammalian systems.

Proteins with an N-terminal cysteine are widely used in protein engineering and can be generated by a number of different ways.[24–26] Our approach takes advantage of intein-mediated protein splicing (Fig. 1A) for the in vivo generation of the target protein bearing an N-terminal cysteine residue. Subsequent site-specific labeling of the protein occurs via the well-known native chemical ligation (NCL) reaction between this cysteine and a membrane-permeant thioester-containing small molecule probe (Fig. 1B) resulting in the formation of a covalent protein-probe adduct. NCL is one of the few highly specific chemical transformations that are compatible with the complex and demanding cellular milieu,[28] and requires fewer genetic manipulations at the level of proteins and host cells alike. Typically other in vivo compatible chemical reactions like the ketone-hydrazine reaction[20] and the Staudinger ligation reaction[29] require the introduction of unnatural functionalities into biomolecules and/or tedious genetic manipulations. Following intein cleavage, only one additional amino acid is

Abbreviations: EGFP, enhanced green fluorescent protein; ECFP, enhanced cyan fluorescent protein; PCR, polymerase chain reaction; GST, glutathione-*S*-transferase; FRET, fluorescence resonance energy transfer; NLS, nuclear localization sequence.

Keywords: In vivo protein labeling; Native chemical ligation; Bioimaging; Fluorescence; FRET.

* Corresponding author. Tel.: +65 65162925; fax: +65 67791691; e-mail: chmyaosq@nus.edu.sg

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.01.043

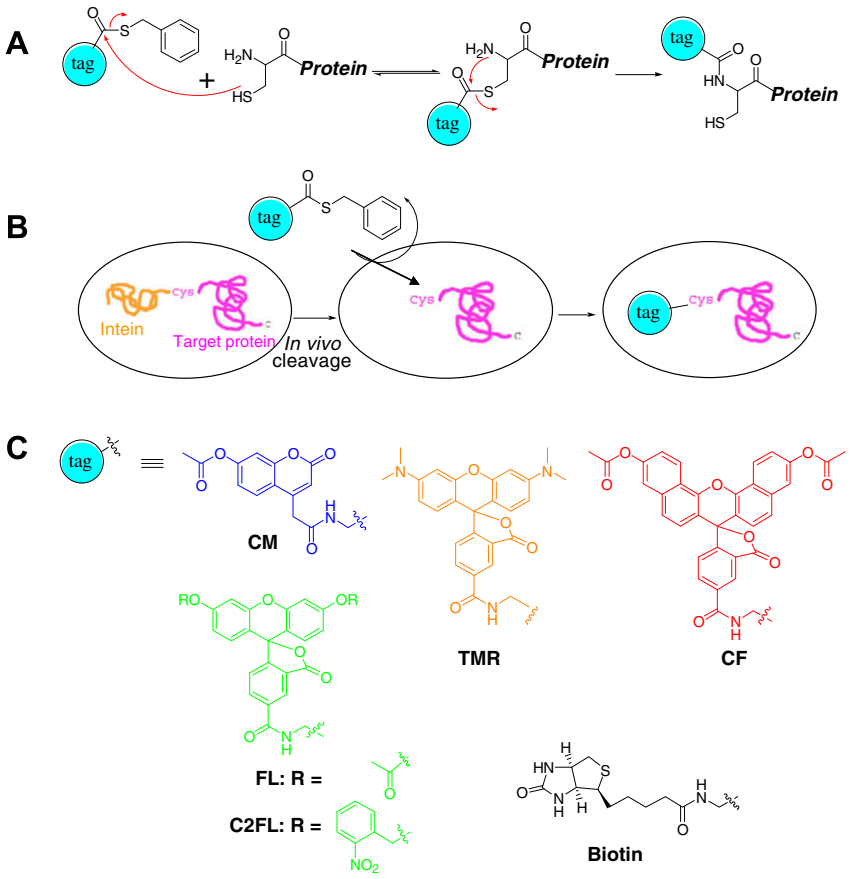

Figure 1. (A) Chemoselective native chemical ligation between an N-terminal cysteine in a protein and a thioester-containing probe, forming a stable amide bond. (B) Our strategy for site-specific covalent labeling of N-terminal cysteine proteins and thioester probes in live cells. (C) Structures of cell-permeable, thioester probes used in this study.

added to the target protein unlike other methods that necessiate the fusion of target proteins to large protein domains/tag that can potentially peturb the folding and/or activity of target proteins. Furthermore, the use of inteins obviates the dependence of endogenous methioninyl aminopetidases (MAP) for the in vivo removal of methionine residues since the cleavage process is significantly influenced by the residues adjacent to the initiator methionine.[30]

A panel of cell-permeable, thioester-containing small molecules were used in our studies (Fig. 1C): (1) mono- or di-acetate forms of coumarin, fluorescein, tetramethylrhodamine, and carboxynaphthofluorescein (**CM**, **FL**, **TMR**, and **CF**, respectively), representing the spectral variants of cell-permeable analogs of the fluorophores (blue, green, orange, and red, respectively). Acetate groups on the probes are spontaneously hydrolyzed inside the cell by endogenous esterases releasing the highly fluorescent free dyes; (2) **C2FL**, representing the 'caged' form of **FL** wherein the two acetates were replaced by photo-labile 2-nitrobenzyl groups. Selective 'unmasking' of **C2FL** by photolysis makes this probe useful for bioimaging techniques where spatio-temporal activation of fluorescence is required; (3) **BIOTIN**, useful for potential in vivo protein–protein interaction studies utilizing biotin–avidin affinity purification/enrichment of protein complexes. The photophysical properties of these probes remained similar to the parent fluorophores.[27]

2. Results and discussion

2.1. In vivo intein cleavage to generate N-terminal Cys proteins

The intein-mediated system was originally designed for protein purification on an affinity column.[25] This meant that one would prefer cell growth conditions which minimize the in vivo cleavage of the fusion protein so as to allow the on-column isolation of the entire fusion (e.g., intein-target protein) before subsequent in vitro cleavage of the target protein takes place. However, we observed with great interest that a substantial amount of the fusion protein always underwent spontaneous in vivo cleavage under normal cell growth conditions (i.e., pH 7 in cell media when grown at 37 °C or even at room temperature) leading to the release of the intein-cleaved target protein inside growing cells. We therefore capitalized on this phenomenon to generate, in vivo, N-terminal cysteine-containing proteins that can be subsequently labeled site-specifically with thioester-containing small molecule probes. Bacterial cells transformed with pTWIN2-GST vector were used as a model to

optimize cell growth conditions such that maximum intein-mediated protein cleavage occurred in vivo to generate GST possessing an N-terminal cysteine residue. We found that inducing protein expression at room temperature for 18 h was sufficient to cleave ~50% GST from its intein fusion (Fig. 2A). We next assessed the in vivo cleave of EGFP and ECFP, transiently expressed in HEK293 cells from pT-Rex-DEST30-intein-EGFP and pT-Rex-DEST30-intein-ECFP-NLS, respectively. Both EGFP and ECFP were engineered such that each possesses an N-terminal cysteine residue upon cleavage from the intein fusion. The pT-Rex-DEST30-intein-ECFP-NLS construct, having a nuclear localization sequence (NLS) attached to the C-terminus of intein-ECFP fusion, allows us to assess the in vivo cleavage of intein fusion inside the nucleus. As shown in Figure 2B and C, both EGFP and ECFP were efficiently cleaved (>95% cleavage for intein-EGFP and >60% for intein-ECFP-NLS) after 48 h of transient transfection and cell growth indicating that the intein-mediated, in vivo protein cleavage occurred efficiently in different cellular compartments inside mammalian cells. Western blot was necessary to follow the in vivo cleavage in mammalian cells as only small amounts of proteins were expressed due to low transfection efficiencies.

2.2. In vivo labeling of N-terminal cysteine-containing proteins in bacteria

Having established that intein-mediated protein cleavage occurred efficiently inside both bacterial and mammalian living cells, we next labeled the N-terminal cysteine-containing proteins, following intein cleavage, with our cell-permeable thioester-containing probes. To label the N-terminal cysteine-containing GST overexpressed in bacteria, the probe (5–100 μM) was directly added to the growing cell media after in vivo protein cleavage was induced, followed by incubation for 0–24 h for the labeling reaction to occur. At the end of the labeling, excessive probe and side products (i.e., probes conjugated to free cysteine) were removed by extensive washings of the cells. For quantification of time-dependent in vivo labeling, a small sample of the labeled cells was periodically analyzed by SDS–PAGE (Fig. 3A and B): consistent with previous reports for in vitro labeling,[27] in vivo labeling was shown to occur in a time-dependent fashion, with >50% of the labeled protein observed within the first 3 h of labeling.

Due to the intrinsic reactivity of thioester toward any thiol-containing molecules, it is possible that our probes may react with internal cysteines of a protein, generating a 'mislabeled' protein in which internal cysteines in the protein is thiol-esterified by the probe via a thiol exchange reaction. We believe, however, this 'thiol-acylation' reaction is readily reversed inside living cells by aqueous hydrolysis to spontaneously release the unlabeled protein. Consequently, of all proteins present in the cell, only those possessing an N-terminal cysteine may be stably labeled (due to the formation of a peptide bond) by our probes. In order to unambiguously confirm this, we expressed and labeled, in vivo, N-terminal cysteine-containing GST followed by direct analysis of the whole-cell protein content using a DTT-free SDS loading buffer (lane 1 in Fig. 3C). Lysis of proteins using a DTT-free loading buffer prior to SDS–PAGE separation allows the detection of any protein which may have been 'mislabeled' at its internal cysteine thiols.[17,25] As shown in Figure 3C, the major fluorescent band observed in DTT-free SDS loading buffer corresponds, both in intensity and molecular weight, to that of the target protein when DTT was present (i.e., lane 1 vs 2). We also observed some negligible yet noticeable background bands observed in lane 1 which were not present in lane 2 (* in Fig. 3C), which may have arisen from non-specific labeling of endogenous proteins. Despite this, we believe these background labelings are tolerable for most bioimaging experiments, and they may be reduced further by more extensive washings of the labeled cells. Put together, our results further confirmed the site-specific covalent nature of our strategy, which allows in vivo labeling N-terminal cysteine-containing proteins with high efficiency and low background.

2.3. Fluorescence microscopy of bacterial cells labeled with different probes

The spectra overlap of EGFP and the probe **TMR** in our labeling strategy provide an ideal donor–acceptor pair for fluorescence resonance energy transfer (FRET), which only occurs when both donor and acceptor are in close proximity.[4,5] This serves to unambiguously confirm the covalent labeling of the N-terminal cysteine proteins with our probes inside live cells. As shown in Figure 4A, a clear FRET signal was observed in all **TMR**-labeled bacterial cells expressing N-terminal cysteine-containing EGFP, indicating the covalent nature

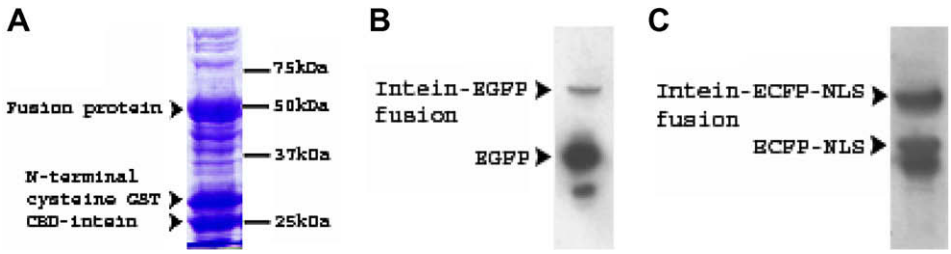

Figure 2. In vivo cleavage efficiency of intein-fused proteins in different organisms. (A) In bacteria, in vivo cleavage occurs at about 50% as seen by coomassie stain of a 12% SDS–PAGE. (B) and (C) Western blots (with anti-EGFP antibody) of in vivo cleavage of intein-fusion proteins in HEK293 mammalian cells: >95% in vivo cleavage for intein-EGFP as shown in (B), and >60% in vivo cleavage for intein-ECFP-NLS as shown in (C). Note that anti-EGFP antibody detects both EGFP and ECFP.

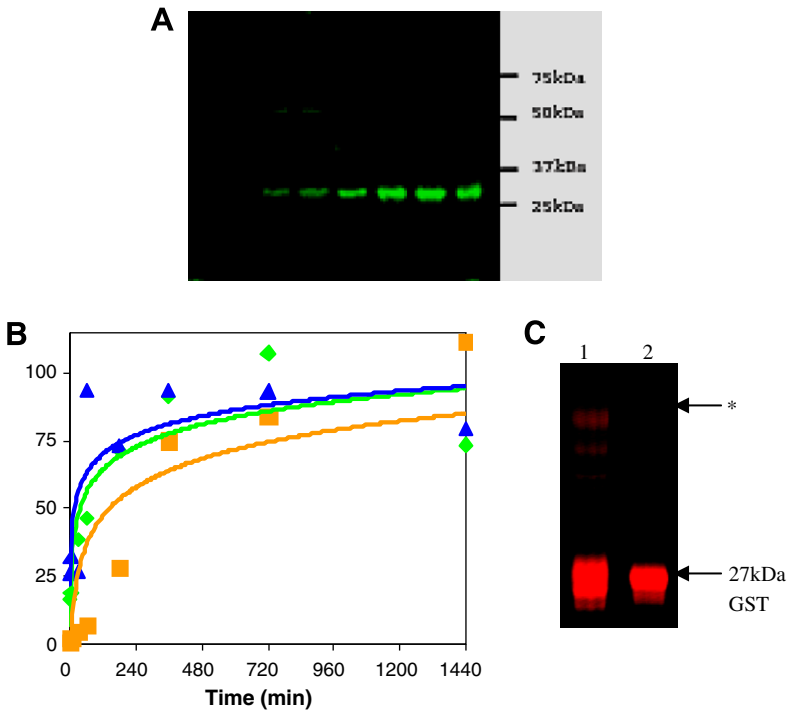

Figure 3. (A) Fluorescence in-gel scanning image showing the time-course of labeling of cells expressing N-terminal cysteine-containing GST with **FL**. Labeled cells after specified time intervals were analyzed on a 12% SDS–PAGE (left to right): 0 min, 10 min, 30 min, 1 h, 3 h, 6 h, 12 h and 24 h. (B) % completion of GST labeling over 24-h intervals with **FL** (♦), **TMR** (■), and **BIOTIN** (▲). At the indicated time point, the extent of labeling was quantitated by the integration of the fluorescence intensity (for **FL** and **TMR** probes) of the labeled protein band using the ImageQuant software (Amersham Biosciences). (C) Site-specific in vivo labeling of N-terminal cysteine-containing GST with 10 μM **TMR**, showing negligible background labeling from acylated internal cysteines. Cells in lanes 1 were lysed in DTT-free SDS-loading buffer, while those in lanes 2 were lysed in regular SDS-loading buffer containing DTT. The 27 kDa GST band in lane 1 appeared smeared due to the lack of reducing agent (i.e., DTT) in the SDS-loading buffer. Some noticeable background bands (indicated by ∗) in lane 1 which disappeared in lane 2 may be from 'mislabeled' endogenous proteins.

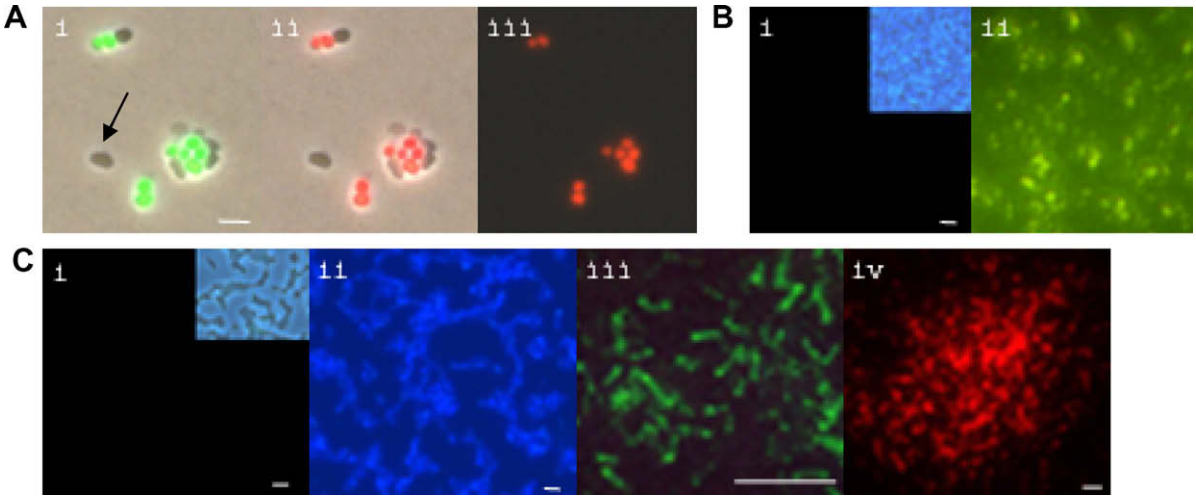

Figure 4. Fluorescence micrographs of live bacterial cells after labeling with 20 μM of the indicated probes: (A) EGFP-expressing cells labeled with **TMR**: (i) overlay of phase contrast image with fluorescence microscopy image (GFP channel). Arrowed cells are those not expressing N-terminal cysteine-containing EGFP; (ii) overlay of phase contrast image with fluorescence microscopy image (TMR channel); (iii) FRET channel of the labeled cells (excitation: 470 ± 20 nm; emission: 605 ± 30 nm). Note that cells not expressing EGFP showed neither TMR nor FRET signal. Scale bar: 1 μm. (B) GST-expressing cells labeled with **C2FL**: (i) fluorescence microscopy image (GFP channel) after 0 min UV photolysis (inset: phase contrast image); (ii) fluorescence microscopy image (GFP channel) after 5 min UV photolysis. Scale bar: 2 μm. (C) GST-expressing cells labeled with other probes: (i) fluorescence microscopy image (TMR channel) of negative control with cells not expressing N-terminal GST but labeled with **TMR** (inset: phase contrast image); (ii) fluorescence microscopy image (coumarin channel) of GST-expressing cells labeled with **CM**; (iii) fluorescence microscopy image (GFP channel) of GST-expressing cells labeled with **FL**; (iv) fluorescence microscopy image (red channel) of GST-expressing cells labeled with **CF**. All cells were labeled for 24 h. Scale bar: 2 μm.

of our labeling. Also evident in Figure 4A is that neither TMR nor FRET signals could be detected in those cells not expressing EGFP, further confirming the labeling strategy only works when there is an N-terminal cysteine-containing protein present inside the cell. A negative control of unlabeled cells expressing EGFP revealed no fluorescence in the FRET channel even with extended exposure time (Supporting Information, Fig. S2).

We next demonstrated the versatility of the strategy by labeling live cells with probes having different fluorescence and other chemical properties (Fig. 4B and C). Bacterial cells expressing N-terminal cysteine GST were labeled with **CM**, **FL**, **TMR**, and **CF** giving rise to cells that have different 'colors' (Fig. 4C). When labeled with **C2FL**, the 'caged' analog of **FL**, labeled proteins inside the cell could be selectively 'lighted' up by UV photolysis (Fig. 4B) indicating the potential of this approach for a variety of other advanced bioimaging techniques.[31]

2.4. In vivo labeling of N-terminal cysteine-containing proteins in mammalian cells

Having successfully demonstrated the strategy for highly specific, covalent protein labeling inside bacterial cells, we next extended the strategy to mammalian cells, which are more useful for bioimaging applications, yet much more challenging because of their cellular complexity. HEK293 cells were used in our studies. Intein-fused EGFP and ECFP mammalian expression vectors were constructed, in which an extra N-terminal cysteine was introduced into both EGFP and ECFP, as described under 4. To facilitate evaluation of the labeling, as well as to assess whether the strategy targets proteins expressed in subcellular compartments within mammalian cells (e.g., nucleus), we fused a nuclear localization sequence (NLS) to ECFP, generating ECFP-NLS. As described in previous paragraphs, in vivo cleavage of the intein-fused protein occurred efficiently in HEK293 cells (Fig. 2B and C), consistently generating >50% cleaved proteins for subsequent labeling experiments.

To determine the site-specific, covalent nature of our labeling strategy in a mammalian system, HEK293 cells transfected with EGFP were incubated with the thioester-containing biotin probe, **BIOTIN**, in cysteine-free media, and the in vivo labeling of EGFP was assessed by SDS–PAGE and Western blots of the resulting cell lysates (Fig. 5A). Western blot was necessary to validate the in vivo labeling in mammalian cells, as only a small amount of the target protein was expressed due to intrinsically low efficiency of the transient transfection. As shown in Figure 5A, endogenously biotinylated proteins, namely methyl-crotonyl-CoA-carboxylase and propionyl-CoA-carboxylase (both 75 kDa), were detected in untransfected, unlabeled HEK 293 cells (lane 1). This is consistent with what was reported previously.[17] Untransfected cells on the other hand, when labeled similarly with the probe, showed a noticeable background labeling (lane 2), which, fortunately was almost completely eliminated in EGFP-transfected, labeled cells (lane 3): the only other major band observed, besides the expected EGFP, was that of endogenous biotinylated proteins. This validated the feasibility of our strategy in mammalian cells with tolerable background labeling. We suspect that the background labeling observed in untransfected cells (e.g., lane 2 in Fig. 5A) was a result of artifacts from the Western blot, although we could not eliminate the possibility of some non-specific labeling in our strategy.

The relative efficiency of our in vivo labeling strategy was assessed by streptavidin absorption experiments (Fig. 5B and C) with EGFP-expressing HEK293 cells labeled with **BIOTIN**. After the labeling reaction, cells were extensively washed to remove any excessive free probe, lysed, and the resulting cell lysate was incubated with an excessive amount of streptavidin beads to isolate

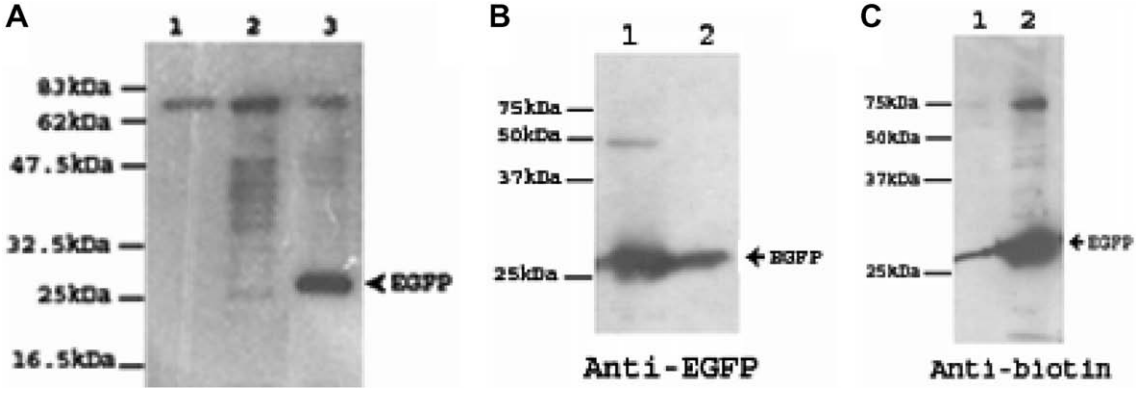

Figure 5. Site-specific labeling of N-terminal cysteine-containing proteins in HEK293 mammalian cells. (A) EGFP-expressing cells labeled with **BIOTIN** for 24 h. Shown is the Western blot (with anti-biotin) of HEK293 cells: lane 1, non-transfected cells; lane 2, non-transfected cells labeled with 100 μM of the probe; lane 3. EGFP-transfected cells labeled with 100 μM of the probe. Note that a band was observed in all lanes, corresponding to endogenously biotinylated methyl-crotonyl-CoA-carboxylase and propionyl-CoA-carboxylase (both 75 kDa). (B) and (C) Determination of percentage in vivo labeling in HEK293 cells expressing EGFP using streptavidin absorption experiments. The same sample was divided equally and 2 separate Western blots using (B) anti-EGFP and (C) anti-biotin were carried out. In lanes 2, biotinylated proteins were pulled-down using streptavidin beads.[17] These include the 27 kDa biotinylated N-terminal cysteine EGFP and 75 kDa endogenous mammalian proteins. Lane 1 contains one-fifth of the flow-through from the absorption experiments.

biotinylated proteins (the bead-bound fraction) from non-biotinlyated proteins (the flow-through fraction). As shown on the Western blot detected with anti-EGFP antibody (Fig. 5B) between the fractions of EGFP in the flow-through (lane 1; 1/5 of total volume) and bound to the beads, it was estimated that <10% all cleaved EGFP was labeled in vivo. A separate but duplicate Western blot was run simultaneously, using anti-biotin antibody, to ensure successful separation of biotinylated/non-biotinylated proteins in the pull-down experiment (Fig. 5C): very few biotinylated EGFP were detected in the flow-through fraction (i.e., lane 1). Work is currently underway to improve the labeling efficiency of our strategy.

2.5. Fluorescence microscopy of mammalian cells

Next, we used fluorescence microscopy to visualize the labeling ECFP in HEK293 cells. The ECFP was localized to the nucleus. After transfection with pT-Rex-DEST30-intein-ECFP-NLS vector, cells were shown to express ECFP almost exclusively in their nuclei (Fig. 6(ii)), indicating successful nuclear localization of the protein. Upon further induction of intein-mediated protein cleavage to release the active N-terminal cysteine-containing ECFP, cells were labeled with **TMR** for 24 h, washed extensively then visualized (Fig. 6(iii)): the majority of the fluorescence detected in the TMR channel was observed to be accumulated inside the nucleus of the transfected cell indicating that successful and site-specific labeling of the nuclear localized ECFP by **TMR** had occurred. Also apparent in the labeled cell, however, were some much weaker but significant fluorescence signals detected at the periphery of the nucleus (Fig. 6(iii)). We believe this was not caused by non-specific covalent labelings, but rather the non-covalent trapping of the probe inside the mitochondrial, as rhodamine dyes (i.e., TMR) are known previously to localize in the mitochondria of mammalian cells.[32] In the future, other dyes may be used to minimize this problem. Finally, since ECFP and **TMR** form a FRET pair, the covalent labeling of our strategy was further confirmed by the observation of a clear FRET signal present exclusively in the nucleus of the labeled transfected cell (Fig. 6B(iv)). A negative control with unlabeled cells expressing ECFP-NLS showed no FRET signal under the same exposure time (Supporting Information, Fig. S3).

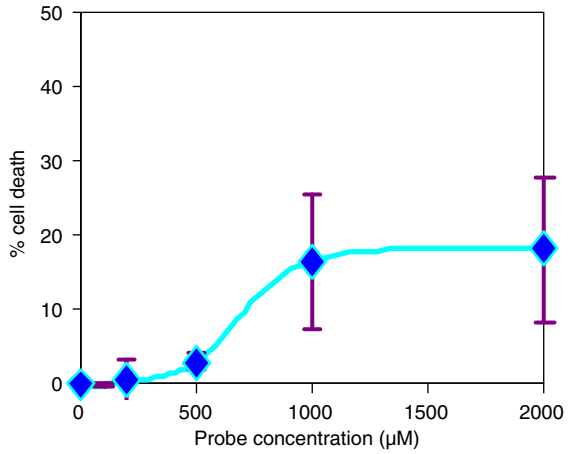

Figure 7. Assessment of probe cytotoxicity-HEK293 cells expressing N-terminal cysteine-containing ECFP-NLS proteins was labeled at different concentration of the **BIOTIN** probe (100 μM to 2 mM) for 24 h. Cell viability was assessed by Trypan blue exclusion assay. Percentage cell death values shown are after negation of cytotoxicity due to vehicle (DMSO) alone. All experiments were performed in triplicates.

In our strategy, when the fluorescent tag is covalently attached to the N-terminus of a protein, the protein effectively becomes a fluorophore itself. In spite of the fact that we observed some background labeling in certain probes (i.e., **TMR** labeling in mitochondria), we believe this problem may be overcome in future by carefully choosing probes which possess unique chemical, physical, and biological properties, which, when combined with advanced bioimaging techniques such as photoactivation, FRET, fluorescence recovery after photobleaching (FRAP) or fluorescence loss after photobleaching (FLIP) should make the strategy applicable to an even wider range of biological experiments.

2.6. Probe toxicity

Lastly, we assessed the cytotoxicity of our probes in mammalian cells. The biotin-containing probe, **BIOTIN**, was taken as an example. Different amounts of

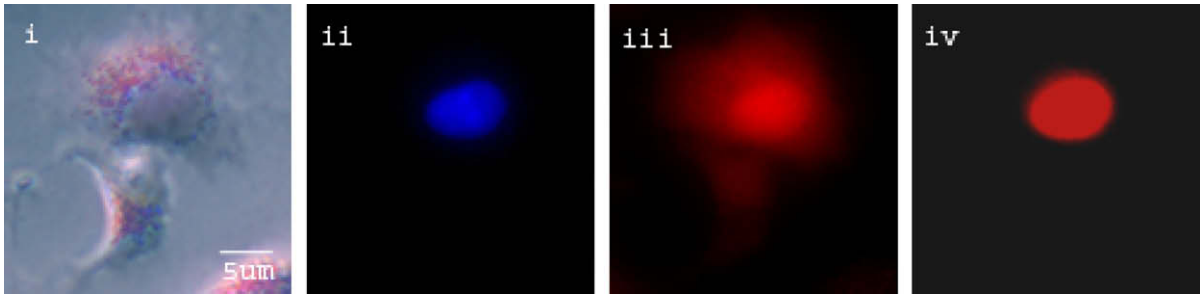

Figure 6. Fluorescence microscopy of ECFP-NLS-expressing HEK293 cells labeled with **TMR** for 24 h. (i) Phase contrast image; (ii) fluorescence image (CFP channel) indicating nuclear localization of ECFP; (iii) fluorescence image (TMR channel), indicating the majority of **TMR** labeling occurred inside the nucleus with some mitochondria staining; (iv) FRET channel (excitation: 436 ± 10 nm; emission: 620 ± 30 nm). FRET signals were detected only in the nucleus of the labeled cell. No signals were detected in either ECFP-transfected but not **TMR**-labeled cells (see Supporting Information, Fig. S3).

BIOTIN were added to HEK293 cells transfected with ECFP-NLS, and the percentage of cell death as a result of **BIOTIN** was monitored. The percentage of cell death was determined for **BIOTIN** concentrations ranging from 100 μM to 2 mM (Fig. 7): with <500 μM of **BIOTIN** added to the cells, minimal cell death was observed (less than 5%). With the addition of up to 2 mM of the probe, only ~20% of cell death was observed. This indicates that our labeling conditions (e.g., typically 10–100 μM of the probe) should have negligible effect on the cell viability, further validating the compatibility of our labeling strategy with live cell experiments.

3. Conclusion

Our labeling approach has several advantages over existing strategies.[2–23] Genetic fusions require the introduction of a macromolecular protein, namely GFP (27 kDa, or its derivatives),[2] avidin (55 kDa)[32] or hAGT (23 kDa),[10] to the target protein, which may subsequently affect its biological and cellular activities. In our strategy, only one extra amino acid, that is, cysteine, is introduced, thus minimizing potential perturbation to the original conformation and activity of the protein. This also compares well with the biarsenical approach which requires the presence of a tetracysteine motif-CCXXCC.[4] Since N-terminal cysteine-containing proteins could be readily generated in vivo,[24–26] our approach does not require complex genetic manipulation of the protein/host, as in the approaches developed by Schultz[20] and Bertozzi.[22] Lastly, with the easy access to potentially a large array of thioester-containing small molecule probes, our approach may provide a general method for in vivo protein labeling in a variety of biological experiments.

In conclusion, we have developed a simple yet highly versatile method for site-specific, covalent labeling of proteins inside live cells. We have shown that this strategy can be applied to different organisms with tolerable background labeling and with a variety of easily accessible small molecule probes, thus making it potentially useful for future bioimaging and proteomics applications.[33] This strategy, however, is not amenable to proteins that have a pre-requisite for an amino-terminal signal sequence for organellar localization.

4. Experimental

4.1. Probe synthesis

All probes used in our studies were synthesized as described elsewhere.[27] The final stock concentration of all the probes reported in this study were 5–10 mM in DMSO.

4.2. Construction of expression plasmids

Two model proteins, EGFP and GST, were PCR amplified from pEGFP (Clontech, USA) and pGEX-4T1 (Pharmacia Biotech, USA) vectors, respectively, and cloned into the pTWIN1/2 expression vector (NEB, USA) following the C-terminus of the *Ssp* DnaB intein tag. The genes were inserted between the first SapI site and the PstI site on the vector, with a cysteine introduced by PCR as the first amino acid of the target protein, generating pTWIN1-EGFP and pTWIN2-GST, respectively. Similarly, the ECFP gene containing a nuclear localization sequence (ECFP-NLS) was amplified from the pECFP-Nuc vector (Clontech, USA) and cloned into the pTWIN1 vector to generate pTWIN1-ECFP-NLS. The intein-fused EGFP and ECFP-NLS genes were then amplified from their respective bacterial constructs (i.e., pTWIN1-EGFP and pTWIN1-ECFP-NLS), cloned into the donor vector pDONR201 using the GATEWAY™ cloning system (Invitrogen, USA), followed by recombination of the cloned genes into a mammalian expression vector pT-Rex-DEST30 (Invitrogen) to generate pT-Rex-DEST30-intein/EGFP and pT-Rex-DEST30-intein/ECFP-NLS, respectively. All constructs were verified by DNA sequencing. The primers used to generate different constructs as shown above are listed below:

pTWIN1-EGFP
5′-GGT GGT TGC TCT TCC AAC TGC AGA GCC ATG GTG AGC AAG GGC-3′
5′-GGT GGT CTG CAG TTA CTT GTA CAG CTC GTC-3′
pTWIN2-GST
5′-GGT GGT TGC TCT TCC AAC TGC AGA GCC ATG TCC CCT ATA CTA-3′
5′-GGT GGT CTG CAG TCA GTC ACG ATG CGG-3′
pTWIN1-ECFP-NLS
5′-GGT GGT TGC TCT TCC AAC TGC AGA GCC ATG GTG AGC AAG GGC-3′
5′-GGT GGT CTG CAG TTA TCT AGA TCC GGT GGA-3′
pT-Rex-DEST30-intein-EGFP
5′-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GGA GAT AGA ACC ATG GCT ATC TCT GGC GAT AGT-3′
5′-GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC CTG CAG TTA CTT GTA CAG-3′
pT-Rex-DEST30-intein-ECFP-NLS
5′-GGT GGT CTG CAG TTA TCT AGA TCC GGT GGA-3′
5′-GGGG AC CAC TTT GTA CAA GAA AGC TGG GTC CTG CAG TTA TCT AGA TCC-3′

4.3. Protein expression and in vivo labeling in bacteria

Bacterial constructs were transformed into the *Escherichia coli* expression strain ER2566 (NEB) and grown in 100 μg/ml ampicillin containing LB media at 37 °C. At OD_{600} = ~0.6, protein expression was induced by the addition of 0.3 mM of IPTG (isopropyl-β-D-thiogalactoside) and the cells were further grown for 12 h at room temperature to allow the expression and folding of the fusion protein, as well as for the fusion protein to undergo self-cleavage and generate the desired N-terminal cysteine protein in vivo. To label the protein in live cells, 20 μM of the probe (5–100 μM also worked) was added directly to the LB media containing grown

cells, and incubated at room temperature for 24 h. For the time-course labeling experiments, a small sample of cells was removed at each time interval, quenched with 10 mM of cysteine and lysed by boiling in SDS sample buffer. The resulting sample was analyzed directly on a 12% SDS–PAGE to ensure that all labeled products were detected. Labeling was visualized by in-gel fluorescence scanning with a Typhoon™ 9200 fluorescence scanner (Amersham Biosciences, USA). Biotinylated proteins were detected by Western blotting. Briefly, following SDS–PAGE, the resulting gel was electroblotted onto a polyvinylidene difluoride (PVDF) membrane (BioRad, USA) and blocked for 1 h with 5% non-fat dry milk in PBST (phosphate buffered saline, pH 7.4, with 0.1% Tween 20). The membrane was incubated with horseradish peroxidase (HRP)-conjugated anti-biotin antibody (Cell Signaling Technologies, USA) at a 1:1000 dilution in PBST for 1 h with gentle agitation. Blots were then washed with PBST (3×15 min) and protein bands were detected by chemiluminescence using the ECL™ kit (Amersham).

4.4. Protein expression and in vivo labeling in mammalian cells

HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, penicillin (100 U/ml) and streptomycin (100 μg/ml) at 37 °C with 5% CO_2. Cells were seeded at 2.4×10^6 cells per 100 mm tissue culture plate. After overnight incubation, cells were transiently transfected with either pT-Rex-DEST30-intein-EGFP or pT-Rex-DEST30-intein-ECFP-NLS using PolyFect Transfection Reagent (Qiagen, USA). After 36 h of protein expression, cells were washed in 1× PBS and cysteine-free DMEM (Sigma) was added. The biotin-thioester probe, **BIOTIN**, was added to a final concentration of 100 μM (10–100 μM also worked) and cells were incubated for another 24 h. Cells were harvested by centrifugation at 1000 rpm for 10 min and resuspended in 1× PBS. Washings were repeated at least three times, and cells were lysed in PBS using glass beads. Streptavidin MagneSphere® Paramagnetic Particles (Promega, USA) were used to pull-down all biotinylated proteins in the cell lysate. Protein samples were incubated with excess streptavidin magnetic beads for 1 h at 4 °C to ensure all biotinylated proteins were absorbed onto the beads. Beads were then washed thrice in 1× PBS, boiled in 1× SDS loading buffer and loaded onto a 12% SDS–PAGE. Western blotting was used to assess biotinylated proteins, as described above. For the quantification of in vivo labeling efficiency, a streptavidin absorption experiment was carried out to separate biotinylated proteins from non-biotinylated ones, as previously described.[17] Sample fractions bound to the streptavidin beads and from the flow-through were analyzed by both anti-biotin and anti-EGFP Western blotting.

4.5. Probe toxicity assay

The biotin-containing thioester probe was used as a representative in this study. A stock solution of 5 mM **BIOTIN** dissolved in DMSO was used for all experiments. HEK293 cells transfected with pT-Rex-DEST30-intein-ECFP-NLS were labeled with **BIOTIN** at concentrations of 100, 200, 500, 1000, and 2000 μM for 24 h. 'Blank' experiments with DMSO only (adjusted to the respective final concentrations) were performed simultaneously. Cells were inspected periodically with a microscope. At the end of 24 h, cells viability was assessed by the Trypan blue (0.04%) exclusion assay. The percentage of cell death from the **BIOTIN** experiments was subtracted from the 'blank' experiments to obtain the final percentage of cell death for a particular **BIOTIN** concentration. All experiments were performed in triplicates and the average values plotted.

4.6. Fluorescence microscopy

For bacteria, labeled cells were harvested by centrifugation at 4000 rpm for 10 min. Upon resuspension in 1× PBS buffer (pH 7.4) containing 10% glycerol, cells were left standing for 30 min. This procedure was repeated three times to ensure the complete removal of any free probe. Cells were mounted on clean glass slides coated with 1.5% agarose. Fluorescence images were recorded with the AxioSkop™ 40 fluorescence microscope (Zeiss, Germany) equipped with a cooled CCD camera (AxioCam, Zeiss) using a 63× or 100× oil objective. For HEK293 mammalian cells, following their growth as described above to induce protein expression, cells were washed in 1× PBS and cysteine-free DMEM (Sigma) was added, followed by 100 μM of **TMR** (10–100 μM also worked) and 0.01 mM $CaCl_2$. Upon incubation for 24 h, labeled cells were washed thrice with 1× PBS and imaged with the AxioVert™ 2000 fluorescence microscope (Zeiss) equipped with a cooled CCD camera (AxioCam, Zeiss) using a 63× objective. Different fluorescence images were obtained with different excitation/emission filter sets: coumarin channel (excitation = 365 nm and emission = 420 nm LP); CFP channel (excitation = 436 ± 20 nm and emission = 480 ± 40 nm); GFP channel (excitation = 470 ± 20 nm and emission = 530 ± 25 nm); TMR and Red channels (excitation = 546 ± 12 nm and emission = 590 nm LP). The FRET channel for CFP/TMR pair was recorded using filters with an excitation = 436 ± 20 nm and emission = 620 ± 20 nm (Chroma). The FRET channel for GFP/TMR pair was recorded using filters with excitation = 470 ± 20 nm and emission = 605 ± 30 nm (Zeiss).

Acknowledgments

Funding support was provided by the National University of Singapore (NUS), Ministry of Education (MOE) and the Agency for Science, Technology and Research (A*STAR) of Singapore.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.01.043.

References and notes

1. Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.
2. Chen, I.; Ting, A. Y. *Curr. Opin. Biotechnol.* **2005**, *16*, 1–6.
3. Marks, K. M.; Nolan, G. P. *Nat. Methods* **2006**, *3*, 591–596.
4. Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269–272.
5. Roberti, M. J.; Bertoncini, C. W.; Klement, R.; Erijman, E. A. J.; Jovin, T. M. *Nat. Methods* **2007**, *4*, 345–351, and references cited therein.
6. Miller, L. W.; Sable, J.; Goelet, P.; Sheetz, M. P.; Cornish, V. W. *Angew. Chem., Int. Ed.* **2004**, *43*, 1672–1675.
7. Farinas, J.; Verkman, A. S. *J. Biol. Chem.* **1999**, *274*, 7603–7606.
8. Marks, K. M.; Braun, P. D.; Nolan, G. P. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9982–9987.
9. Guignet, E. G.; Hovius, R.; Vogel, H. *Nat. Biotechnol.* **2004**, *22*, 440–444.
10. Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. *Nat. Biotechnol.* **2003**, *21*, 86–89.
11. Pol, L. V.; George, N.; Krumm, H.; Johnsson, K.; Johnsson, N. *J. Am. Chem. Soc.* **2005**, *127*, 12770–12771.
12. Yin, J.; Lin, A. J.; Buckett, P. D.; Resnick, M. W.; Golan, D. E.; Walsh, C. T. *Chem. Biol.* **2005**, *12*, 999–1006.
13. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. *Nat. Methods* **2005**, *2*, 99–104.
14. Lin, C. W.; Ting, A. Y. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.
15. Zhang, Y.; So, M.-K.; Loening, A. M.; Yao, H.; Gambhir, S. S.; Rao, J. *Angew. Chem., Int. Ed.* **2006**, *45*, 4936–4940.
16. Giriat, I.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 7180–7181.
17. Lue, R. Y. P.; Chen, G. Y. J.; Hu, Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2004**, *126*, 1055–1062.
18. Pellois, J. P.; Muir, T. W. *Angew. Chem., Int. Ed.* **2005**, *44*, 2–5.
19. Reents, R.; Wagner, M.; Schlummer, S.; Kuhlmann, J.; Waldmann, H. *ChemBioChem* **2005**, *6*, 86–94.
20. Zhang, Z.; Smith, B.; Wang, L.; Brock, A.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735–6746.
21. Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A. *Angew. Chem., Intl. Ed.* **2006**, *45*, 7364–7367.
22. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321–322.
23. Sawa, M.; Hsu, T.-L.; Itoh, T.; Sugiyama, M.; Hanson, S. R.; Vogt, P. K.; Wong, C.-H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12371–12376.
24. Tolbert, T. J.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2171–2174.
25. Xu, M. Q.; Evans, T. C. *Methods* **2001**, *24*, 257–277.
26. Baker, R. T. *Curr. Opin. Biotechnol.* **1996**, *7*, 541–546.
27. Yeo, S. Y. D.; Srinivasan, R.; Uttamchandani, M.; Chen, G. Y. J.; Zhu, Q.; Yao, S. Q. *Chem. Commun.* **2003**, 2870–2871.
28. Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
29. Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007–2010.
30. Ben-Bassat, A.; Bauer, K.; Chang, S.-Y.; Myambo, K.; Boosman, A.; Chang, S. *J. Bacteriol.* **1987**, *169*, 751–757.
31. Schwartz, J. P.; Patterson, G. H. *Science* **2003**, *300*, 87–91.
32. Russell, C. S. J.; Lee, W. G. *Biophys. J.* **1999**, *76*, 469–477.
33. Chen, G. Y. J.; Uttamchandani, M.; Lue, R. Y. P.; Lesaicherre, M. L.; Yao, S. Q. *Curr. Top. Med. Chem.* **2003**, *3*, 705–724.