Analytical Biochemistry 521 (2017) 55–58

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Technical note

Efficient sortase-mediated N-terminal labeling of TEV protease cleaved recombinant proteins

Kwabena Sarpong, Ron Bose*

Division of Oncology, Department of Medicine, Washington University School of Medicine, Campus Box 8076, 660 South Euclid Avenue, St. Louis, MO 63110, USA

ARTICLE INFO

Article history:
Received 2 December 2016
Received in revised form
6 January 2017
Accepted 9 January 2017
Available online 11 January 2017

Keywords:
Sortase
Post-labeling purification
TEV
LPXTG
Affinity tag

ABSTRACT

A major challenge in attaching fluorophores or other handles to proteins is the availability of a site-specific labeling strategy that provides stoichiometric modification without compromising protein integrity. We developed a simple approach that combines TEV protease cleavage, sortase modification and affinity purification to N-terminally label proteins. To achieve stoichiometrically-labeled protein, we included a short affinity tag in the fluorophore-containing peptide for post-labeling purification of the modified protein. This strategy can be easily applied to any recombinant protein with a TEV site and we demonstrate this on Epidermal Growth Factor Receptor (EGFR) and Membrane Scaffold Protein (MSP) constructs.

Site-specific protein labeling is the method of choice for most biochemical and biophysical applications, as this offers a high level of precision for the attachment of a fluorophore or other chemical moiety [1,2]. Due to the relatively low abundance of cysteines in proteins [3], chemical labeling of proteins using maleimide chemistry is a common strategy for most applications. However, many proteins contain multiple cysteine residues and mutagenesis of these cysteines is time-consuming and may compromise protein function. An alternate approach is to label primary amines with N-hydroxysuccinimide ester-based fluorophores. However, the relatively high abundance of lysines and pKa requirements renders the utility of amino groups for protein modification a less commonly used strategy. These challenges are compounded by long reaction times to ensure complete modification of the protein. With fluorophores having a MW of <1 kDa, separation of labeled products from the unlabeled protein can also present challenges. Sub-stoichiometric labeling often results in a diminished signal-to-noise ratio and impacts the utility of fluorophore-labeled proteins for biophysical studies.

Enzymatic approaches for site-specific incorporation of fluorophores are an alternative to these chemical labeling strategies. Sortases are membrane-associated transpeptidases that anchor Gram-positive bacterial surface proteins to their cell walls. Since the discovery of sortases, *Staphylococcus aureus* sortase A (SrtA) has been the prototype for understanding the mechanism of action of these enzymes [4]. Proteins anchored to the cell wall by SrtA possess a C-terminal sorting signal that contains a hydrophobic domain sandwiched between the conserved LPXTG recognition motif and a positively charged tail [4]. SrtA catalyses the hydrolysis of the peptide bond between the threonine and glycine residues to generate an acyl-enzyme intermediate that is subsequently attacked by an oligoglycine peptide in a nucleophilic attack [5]. This results in the formation of a new peptide bond between the incoming nucleophilic glycine-containing peptide and the protein. Seminal work by Schneewind and coworkers laid the ground for its utility in biochemical and biotechnological studies showed a recombinant peptide containing the LPXTG motif alone is sufficient for recognition and catalysis [6]. These studies also indicated that a peptide containing 1-3 N-terminal glycines could replace the peptidoglycan involved in the sortase-mediated reaction [6]. Current biochemical evidence has suggested that only one additional

Abbreviations: TEV, Tobacco Etch Virus; EGFR, epidermal growth factor receptor; MSP, membrane scaffold protein; HER, Human Epidermal growth factor Receptor; KD, kinase domain; SrtA, Sortase A; ACP, acyl carrier protein; LAP, Lipoid acid ligase Acceptor Protein; TAMRA, 5-carboxytetramethylrhodamine; EDTA, Ethylenediaminetetraacetic acid; EGTA, Ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid; NTA, nitrilotriacetic acid.

* Corresponding author.
E-mail address: rbose@wustl.edu (R. Bose).

http://dx.doi.org/10.1016/j.ab.2017.01.008
0003-2697/

residue (preferably a glycine) is required at the C-terminus of the LPXTG recognition sequence for efficient sortase binding and catalysis [7].

Most recombinant proteins used for biochemical and biotechnological applications contain affinity tags that ensure their easy and efficient purification [8,9]. Sandwiched between the affinity tags and the proteins are protease recognition sites that offer the cleavage of the affinity tags following purification. Commonly used recognition sites include TEV, Factor Xa, and Thrombin protease cleavage sites. An important requirement for the sortase reaction is the generation of the N-terminal glycine residue which can be done by removing the initial methionine of an expressed protein using methionylaminopeptidase or engineering a thrombin or TEV protease recognition site that exposes an N-terminal glycine following cleavage [10].

Recent years have seen the development and utility of sortase to modify proteins at their carboxyl and amino termini in addition to internal loops [7,11]. Unlike traditional chemical strategies that are easy to use, protein modification employing short genetically encoded tags such as the LPXTG-tag, ACP-tag and LAP-tag offer a high degree of precision. However, back reaction from the final product (containing the LPXTG motif and therefore an efficient substrate for the sortase enzyme) and the reversible nature of the sortase reaction can lead to sub-stoichiometric protein modification and decreased labeling efficiencies. To address this challenge, the equilibrium of the reaction is driven towards product formation by increasing the fold excess of the fluorophore-containing peptide [7,11]. Recent methodologies to address this issue of irreversibility have included the use of a sortase-tagged expressed protein ligation (STEPL) system that circumvents the removal of unconjugated species [12], dialysis to remove reaction by-products [13] and the introduction of tryptophan-derived zippers around the SrtA recognition motif that induces the formation of a stable β-hairpin [14]. Other research groups have solved this problem by utilizing a depsipeptide, which replaces the amide bond between the threonine and glycine residues with an ester linkage [15]. These challenges make protein modification using sortase cumbersome and potentially expensive when the fluorophore-containing peptide is needed in many fold excess. The presence of reaction by-products as a result of back reaction and the reversible nature of the reaction affect the purity and degree of labeling of the final product, and subsequently present challenges in the utilization of fluorophore-labeled proteins.

We have developed a simple approach that combines TEV protease cleavage, sortase modification and affinity purification to N-terminally label proteins. To achieve stoichiometrically-labeled product, a short affinity tag is included in the fluorophore-containing peptide so that post-labeling affinity purification of only the labeled protein can be performed.

We used a *Staphylococcus aureus* Sortase pentamutant (SrtA 5M) that had previously undergone directed evolution to be catalytically more efficient [16]. SrtA 5M contains five mutations and has a 140-fold increase in transpeptidase activity over wildtype SrtA. We engineered SrtA 5M to have a TEV cleavage site before the C-terminal histidine tag, in order to facilitate our downstream purification strategy. This TEV-cleaved SrtA 5M (lacking a histidine tag) has catalytic activity essentially identical to the original SrtA 5M construct (Fig. 1A).

We expressed and purified a recombinant Epidermal Growth Factor Receptor kinase domain (EGFR KD) that contains an N-terminal polyhistidine tag and a TEV cleavage site, ENLYFQG. It is important to note that the ENLYFQS sequence for the TEV protease should not be used in this approach because the resulting N-terminal serine residue is not an effective substrate for SrtA 5M. TEV protease cleavage of this EGFR construct yields a glycine residue at the N-terminus and we will call this protein, Gly-EGFR. We have optimized the TEV protease cleavage procedure to give complete cleavage of this EGFR construct in 6–8 h. We tested whether additional N-terminal glycines (Gly_2-EGFR and Gly_4-EGFR) are better substrates in the sortase-mediated reaction. Our results using a TAMRA-labeled LPETGG peptide showed that the labeling of Gly-EGFR, Gly_2-EGFR, and Gly_4-EGFR were very similar (Fig. 1B). The addition of several N-terminal glycine residues had no effect on the efficiency of the reaction and all future experiments utilized one N-terminal glycine for the sortase-mediated labeling reaction.

N-terminal labeling of Gly-EGFR using sortase and a short peptide (that contains the fluorophore and the sortase recognition motif) initially resulted in <60% modification of the protein with the labeled peptide (data not shown). This was unsatisfactory as unlabeled protein decreases the signal-to-noise ratio and complicates data analysis and experimental interpretation. We hypothesized that the addition of an affinity tag to the short peptide would enable the efficient purification of the labeled EGFR. We designed a peptide that had a 6× histidine tag (for purification of labeled EGFR from unlabeled EGFR), a cysteine residue for maleimide labeling (we also utilized the NH_2-amino group of the peptide for labeling with NHS-succinimidyl esters) and an LPETGG motif at the C-terminus. This 25-mer peptide (MSYYHHHHHHDYDIPTCENLPETGG or H_6-LPETGG peptide, hereafter) was fluorophore-labeled at either the N-terminal methionine residue or at the internal cysteine residue. Labeling of Gly-EGFR with the H_6-LPETGG peptide results in the regeneration of the sequence N-terminal to the EGFR that was cleaved off during TEV protease digestion (albeit subtle mutations at the C-terminus of the peptide, ENLPETG instead of the original ENLYFQG).

Our proposed labeling strategy (Fig. 2A) proceeds by mixing the 0.5 mM fluorophore-labeled H_6-LPETGG peptide with 25 μM Gly-EGFR (300 μg) and 1 μM SrtA 5M (lacking a 6× Histidine tag) in a sortase buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$. Following incubation at 4 °C for 30 min, the reaction is quenched with 10 mM EDTA. We tested the chelating effect of EDTA and EGTA and showed that both chelators at 10 mM concentration were sufficient in quenching the activity of sortase. The reaction mixture which contains labeled EGFR, unlabeled EGFR, the H_6-LPETGG peptide and the sortase enzyme is loaded onto a Superdex 75 10/30 GL column (GE Healthcare Life Sciences, Pittsburg, PA) to remove the excess H_6-LPETGG peptide and sortase enzyme. Fractions containing labeled and unlabeled EGFR are pooled together and incubated with Nickel NTA beads. After 20 min, the flow through (containing unlabeled EGFR) is discarded while the bound labeled EGFR is eluted with imidazole. Fig. 2B shows the use of this strategy to site specifically label TEV protease-cleaved EGFR KD with the H_6-LPETGG peptide labeled with a quencher (CruzQuencher™1 Maleimide, from Santa Cruz Biotechnology, Santa Cruz, CA). Our protein recovery from labeling 300 μg of EGFR was 140 μg, resulting in ~50% protein yield. The absorbance of the fluorophore (A_{max}) and the labeled protein (A_{280}) together with the molar extinction coefficients of the fluorophore (ε_{max}) and protein ($\varepsilon_{protein}$) allow us to calculate the moles of dye per mole protein or the degree of labeling (DOL) using the equation below;

$$DOL = \frac{A_{max} \cdot \varepsilon_{protein}}{(A_{280} - A_{max} \cdot CF_{280}) \cdot \varepsilon_{max}}$$

The correction factor (CF) is included in this equation to account for the absorption of fluorophore at 280 nm and equals the A_{280} of the dye divided by the A_{max} of the dye. We have consistently achieved >0.95 mol of dye per mole of protein in our purified labeled EGFR. Following labeling and purification, *in vitro* kinase assays of unlabeled and fluorophore-labeled EGFR kinase domains show