Published in final edited form as:

J Am Chem Soc. 2018 July 11; 140(27): 8420–8423. doi:10.1021/jacs.8b05200.

Protein Labeling via a Specific Lysine-Isopeptide Bond using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

Scott A. McConnell[1], **Brendan R. Amer**[1], **John Muroski**[1], **Janine Fu**[1], **Chungyu Chang**[3], **Rachel R. Ogorzalek Loo**[1], **Joseph A. Loo**[1], **Jerzy Osipiuk**[2], **Hung Ton-That**[3], and **Robert T. Clubb**[1,†]

[1]Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics and the Molecular Biology Institute, University of California, Los Angeles, CA, USA

[2]Structural Biology Center, Argonne National Laboratory, Argonne, IL, USA

[3]Department of Microbiology & Molecular Genetics, University of Texas Health Science Center, Houston, TX, USA

Abstract

Proteins that are site-specifically modified with peptides and chemicals can be used as novel therapeutics, imaging tools, diagnostic reagents and materials. However, there are few enzyme-catalyzed methods currently available to selectively conjugate peptides to internal sites within proteins. Here we show that a pilus-specific sortase enzyme from *Corynebacterium diphtheriae* (^{Cd}SrtA) can be used to attach a peptide to a protein via a specific lysine-isopeptide bond. Using rational mutagenesis we created ^{Cd}SrtA^{3M}, a highly activated cysteine transpeptidase that catalyzes *in vitro* isopeptide bond formation. ^{Cd}SrtA^{3M} mediates bioconjugation to a specific lysine residue within a fused domain derived from the corynebacterial SpaA protein. Peptide modification yields greater than >95% can be achieved. We demonstrate that ^{Cd}SrtA^{3M} can be used in concert with the *S. aureus* SrtA enzyme, enabling dual, orthogonal protein labeling via lysine-isopeptide and back-bone-peptide bonds.

Enzymatic methods that site-specifically functionalize proteins are of significant interest, as they can enable the creation of novel protein-conjugates for medical and research applications[1–5]. The *Staphylococcus aureus* sortase (^{Sa}SrtA) has been developed into a powerful protein engineering tool[6–10]. It catalyzes a transpeptidation reaction that covalently modifies the target protein via a backbone peptide bond, by joining peptide segments that contain a LPXTG ‘sort-tag’ and an N-terminal oligoglycine amine group[11,12]. Several groups have now optimized this reaction to modify proteins with a range of molecules,

†To whom correspondence should be addressed: Robert T. Clubb. Department of Chemistry and Biochemistry, University of California, Los Angeles, 611 Charles Young Drive East, Los Angeles, CA 90095, USA; Tel. (+1) 310 206 2334; Fax (+1) 310 206 4779; rclubb@mbi.ucla.edu.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website. This includes PDF file that shows additional data and procedures used in this paper.

Notes

The authors declare no competing financial interests.

including drugs, lipids, sugars, fluorophores, and peptides[13–21]. While ^{Sa}SrtA is potent tool, it is almost exclusively used to modify target proteins at their N- or C-termini, while it labels internal lysine side chains as a side reaction with low sequence specificity[15,22,23]. Here we show that a mutationally activated sortase enzyme from *Corynebacterium diphtheriae* (^{Cd}SrtA) can site-specifically install a peptide on a protein via a lysine-isopeptide bond. ^{Cd}SrtA and ^{Sa}SrtA have orthogonal activities, enabling dual peptide-fluorophore labeling of a protein via lysine isopeptide- and backbone peptide-bonds, respectively.

Gram-positive bacteria use specialized sortase enzymes to construct pili: long, thin fibers (0.2–3.0 μm × 2-10 nm) that project from the cell surface to mediate bacterial adherence to host tissues, biofilm formation and host immunity modulation[24–26]. These structures are distinct from pili produced by Gram-negative bacteria because their protein subunits (called pilins) are crosslinked by lysine-isopeptide bonds that confer enormous tensile strength[27,28]. Recently, we reconstituted *in vitro* the assembly reaction that builds the archetypal SpaA-pilus in *C. diphtheriae*, the causative agent of pharyngeal diphtheria[29]. ^{Cd}SrtA functions as a pilin polymerase, performing a repetitive transpeptidation reaction that covalently links adjacent SpaA pilin subunits together via lysine-isopeptide bonds. As shown in scheme 1, ^{Cd}SrtA crosslinks adjacent SpaA proteins by connecting their N- (^{N}SpaA, residues 30-194) and C-terminal (^{C}SpaA, residues 350-500) domains, which contain a reactive WxxxVxVYPK pilin motif and LPLTG sorting signal sequences, respectively. In the reaction, ^{Cd}SrtA first cleaves the LPLTG sequence in ^{C}SpaA between the threonine and glycine, forming an acyl-enzyme intermediate in which the catalytic C222 residue in ^{Cd}SrtA is joined to ^{C}SpaA's threonine carbonyl atom.

This transient intermediate is then nucleophilically attacked by the reactive K190 within ^{N}SpaA's pilin motif resulting in a T494-K190 isopeptide bond between ^{C}SpaA and ^{N}SpaA domains within adjacent pilin subunits. Previously, we demonstrated that wild-type ^{Cd}SrtA is catalytically inactive *in vitro* due to the presence of an N-terminal polypeptide segment, called a lid, that masks the enzyme's active site (Fig. 1A)[30–34]. Moreover, we showed that it was possible to activate the enzyme by introducing D81G and W83G lid mutations and we demonstrated that a soluble catalytic domain harboring these mutations (^{Cd}SrtA^{2M} , residues 37-257 of ^{Cd}SrtA with D81G/W83G mutations) site-specifically ligates the isolated ^{N}SpaA and ^{C}SpaA domains *in vitro*[29].

Toward the goal of creating a lysine modifying bioconjugation reagent we improved the ligation activity of ^{Cd}SrtA^{2M} we defined substrate determinants that are required for catalysis. In addition to the aforementioned D81 and W83 mutations in ^{Cd}SrtA^{2M}, inspection of the crystal structure reveals three lid residues that may stabilize its positioning over the active site (I79, N85, K89). The ligation activities of triple mutants of ^{Cd}SrtA containing the D81G and W83G alterations, as well I79R, N85A or K89A substitutions were determined. A D81G/W83G/N85A triple mutant, hereafter called ^{Cd}SrtA^{3M}, has the highest level of ligation activity (Figs. 1B and S1). After a 24 hour incubation with the isolated ^{N}SpaA and ^{C}SpaA domains, ^{Cd}SrtA^{3M} produces 10.6-fold more cross-linked ^{N}SpaAx^{C}SpaA product than ^{Cd}SrtA WT and 35% more product than ^{Cd}SrtA^{2M} (Fig. S1). The mutations in ^{Cd}SrtA^{3M} presumably further displace its lid, thereby facilitating enhanced binding of ^{C}SpaA's LPLTG sorting signal and subsequent acylation by C222. This is substantiated by

our finding that the ^{Cd}SrtA^{3M} triple mutant exhibits the highest level of activity in a HPLC-based sorting signal cleavage assay that reports on formation of the acyl-enzyme intermediate (Fig. S1) and previous studies that have shown that alterations in the lid increase C222 reactivity with 4,4'-dithiodipyridine[29]

^{N}SpaA and ^{C}SpaA are joined by ^{Cd}SrtA^{3M} via their respective pilin motif and LPXTG sorting signal elements. To elucidate determinants required for recognition of the K190 nucleophile, ^{Cd}SrtA^{3M} was incubated with a peptide containing the pilin motif (DGWLQDVHVYPKHQALS) and either ^{C}SpaA or a peptide containing its C-terminal sorting signal (KNAGFELPLTGGSGRI) (Fig. S2). In both instances, no detectable product was observed, indicating that ^{Cd}SrtA^{3M} requires additional tertiary elements within ^{N}SpaA to recognize K190. In contrast, when ^{Cd}SrtA^{3M} is incubated with ^{N}SpaA and the peptide containing the C-terminal sorting signal, >95% of ^{N}SpaA is labeled with the peptide (Fig. 1C). Moreover, LC-MS/MS analysis of the crosslinked species reveals that the components are joined via a site-specific isopeptide bond between the threonine within the sorting signal peptide and the Nε amine of K190 in ^{N}SpaA (Fig. S3A).

We next demonstrated that ^{Cd}SrtA^{3M} can be used to label a target protein via an isopeptide bond with either a peptide fluorophore or another protein. In the labeling reaction a target protein is first expressed as a fusion with the ^{N}SpaA domain containing the pilin motif (hereafter called PM), and then reacted with a LPLTG-containing biomolecule and ^{Cd}SrtA^{3M} (Fig. 2A). To demonstrate peptide fluorophore attachment using ^{Cd}SrtA^{3M}, we incubated the enzyme with ^{N}SpaA and a fluorescent FITCKNAGFELPLTGGSGRI peptide (^{FITC}LPLTG). After incubating for either 24 or 48 hours the reaction components were separated by SDS-PAGE and visualized by either coomassie staining or FITC fluorescence at 530nm. ^{Cd}SrtA^{3M} labels ^{N}SpaA with the fluorescent peptide, yielding a ^{FITC}LPLTx^{N}SpaA cross-linked product (Fig. 2B, right). Fluorophore labeling is specific, as ^{N}SpaA harboring a K190A mutation is unreactive in control experiments (Fig. 2B, left).

To demonstrate that ^{Cd}SrtA^{3M} can also be used to join proteins together via an isopeptide bond, the isolated PM was reacted with green fluorescent protein engineered to contain a C-terminal LPLTGGSGRI sorting signal sequence (GFP-LPLTG). Incubation of these proteins with ^{Cd}SrtA^{3M} resulted in the appearance a higher molecular weight GFP-LPLTx^{N}SpaA cross-linked product (Fig. 2C, S3B). Notably, the ^{Cd}SrtA^{3M} protein-protein ligation reaction is versatile, as labeling can be achieved with the PM fused to either the N- or C-terminus of the target protein.

The ^{Cd}SrtA and ^{Sa}SrtA enzymes recognize distinct nucleophiles, suggesting that they can be used orthogonally to selectively label a single target protein at different sites. To demonstrate orthogonal labeling we created a fusion protein that contained the Small Ubiquitin-like Modifier (SUMO) protein harboring a pentaglycine peptide and PM at its N- and C-termini, respectively (G_5-SUMO$_{PM}$). Our dual modification approach involves sequential reaction of the G_5-SUMO$_{PM}$ substrate with each sortase and peptide fluorophores containing the cognate sorting signal, as outlined in Fig. 3A. To selectively modify Gly_5-SUMO$_{PM}$ (species 1), it was first incubated with ^{Cd}SrtA^{3M} and FITC-LPLTG$_{pep}$ to create at high yield G_5-SUMO$_{PM}$-FITC (species 2) (Fig. 3B). After removal of excess FITC-LPLTG peptide using a

desalting column, the target protein was then labeled at its N-terminus with AlexaFluor$_{546}$-LPATG using ^{Sa}SrtA. This was achieved by incubating species 2 with ^{Sa}SrtA and AlexaFluor$_{546}$-LPATG to produce the doubly labeled protein (species 3). Separation of the reaction products by SDS-PAGE confirms dual labeling, as the expected fluorescence for each probe is detected at ~33 kD during the procedure (Fig. 3B). In particular, FITC-labeled Gly$_5$-SUMO$_{PM}$ is produced after treatment with ^{Cd}SrtA^{3M} (488/530nm excitation/emission), and persists after treatment with ^{Sa}SrtA that catalyzes the second conjugation step with AlexaFluor$_{546}$ (532/605nm excitation/emission). We note that a similar labeling strategy can presumably be used for fusion proteins that contain the SpaB basal pilin instead of ^{N}SpaA, as we have recently shown that ^{Cd}SrtA^{3M} can also use SpaB as a nucleophile *in vitro*[29]. A strength of our approach is the distinct nucleophile and sorting signal substrate specificities of each sortase, which limits cross reactivity. In addition to recognizing distinct nucleophiles, our findings indicate that the sortases have unique sorting signal substrate specificities; ^{Cd}SrtA^{3M} is unable to hydrolyze or use as a transpeptidation substrate sorting signals containing the sequence LPATG that is readily used by ^{Sa}SrtA, but instead it is selective for peptides containing LPLTG (Figs. S4, S5). Moreover, the isopeptide bond created by ^{Cd}SrtA^{3M} is not significantly hydrolyzed by ^{Sa}SrtA or ^{Cd}SrtA after 24 hours (Fig. S6). Thus, ^{Cd}SrtA acts preferentially on its LPLTG sorting signal substrate, preventing potential reversal of LPATG peptides installed by ^{Sa}SrtA. Similarly, the isopeptide linkages installed by ^{Cd}SrtA are not a substrate for reversal by ^{Sa}SrtA or ^{Cd}SrtA.

The bioconjugation chemistry catalyzed by ^{Cd}SrtA^{3M} enables site-specific lysine labeling of a protein, creating an isopeptide linkage that may be less susceptible to proteolysis than conventional peptide bonds. An attractive feature of ^{Cd}SrtA^{3M} is its high degree of specificity for the ε-amine nucleophile within the pilin motif, which enables selective labeling. Transglutaminases can also modify protein lysine residues, but unlike ^{Cd}SrtA^{3M}, these enzymes exhibit minimal substrate specifity[35,36]. Similarly, ^{Sa}SrtA can modify lysines as a side reaction that occurs with minimal specificity and at low efficiency because the lysine ε-amine is not ^{Sa}SrtA's natural substrate[15,22,23]. Chemical methods that modify amino acid side chains have also been developed, but they often require cysteine or non-natural amino acid incorporation into the protein and in some instances harsh reaction conditions[37]. The bioconjugation chemistry catalyzed by ^{Cd}SrtA^{3M} is functionally similar to the non-enzymatic SpyTag/SpyCatcher system[38,39], but its enzymatic activity affords greater control making ^{Cd}SrtA^{3M} an attractive new tool to engineer proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421 and National Institutes of Health Grants AI52217 (R.T.C. and H. T-T.), DE025015 (H. T-T), GM103479 (J.A.L.) and U.S. Department of Energy, Office of Biological and Environmental Research contract DE-AC02-06CH11357 (J.O.). S.A.M. was supported by a Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185). NMR equipment used in this research was purchased using funds from shared equipment grant NIH S100D016336.

REFERENCES

(1). Proft T Biotechnol. Lett 2009 52, 1–10.

(2). Matsumoto T; Tanaka T; Kondo A Biotechnol. J 2012 7 1137–1146. [PubMed: 22829516]

(3). Walper SA; Turner KB; Medintz IL Chit. Opin. Biotechnol 2015 54, 232–241.

(4). Krall N; Da Cruz FP; Boutureira O; Bemardes GJL Nat. Chem 2016 8, 103–113. [PubMed: 26791892]

(5). Rashidian M; Dozier JK; Distefano MD Bioconjug. Chem 2013 24, 1277–1294. [PubMed: 23837885]

(6). Mazmanian SK; Liu G; Ton-That H; Schneewind O Science (80-.). 1999 285, 760–763.

(7). Antos JM; Chew GL; Guimaraes CP; Yoder NC; Grotenbreg GM; Popp MWL; Ploegh HL J. Am. Chem. Soc 2009, 131, 10800–10801. [PubMed: 19610623]

(8). Williamson DJ; Fascione MA; Webb ME; Turnbull WB Angew. Chemie -Int. Ed 2012, 51, 9377–9380.

(9). Levary DA; Parthasarathy R; Boder ET; Ackerman ME PLoS One 2011, 6.

(10). Tsukiji S; Nagamune T ChemBioChem 2009, 10, 787–798. [PubMed: 19199328]

(11). Popp MW: Antos JM; Grotenbreg GM; Spooner E; Ploegh HL Nat. Chem. Biol 2007 3 707–708. [PubMed: 17891153]

(12). Mao H; Hart SA; Schink A; Pollok BA J. Am. Chem. Soc 2004 126, 2670–2671. [PubMed: 14995162]

(13). Samantaray S; Marathe U; Dasgupta S; Nandicoori VK; Roy RP J. Am. Chem. Soc 2008 130 2132–2133. [PubMed: 18229923]

(14). Antos JM.; Miller GM; Grotenbreg GM; Ploegh HL J. Am. Chem. Soc 2008 150, 16338–16343.

(15). Möhlmann S; Mahlert C; Greven S; Scholz P; Harrenga A ChemBioChem 2011 12, 1774–1780. [PubMed: 21656631]

(16). Wagner K; Kwakkenbos MJ; Claassen YB; Maijoor K; Böhne M; van der Sluijs KF; Witte MD; van Zoelen DJ; Cornelissen LA; Beaumont T; Bakker AQ; Ploegh HL; Spits H Proc. Natl. Acad. Sci 2014, 111, 16820–16825. [PubMed: 25385586]

(17). Beerli RR; Hell T; Merkel AS; Grawunder U PLoS One 2015, 10.

(18). Amer BR; MacDonald R; Jacobitz AW; Liauw B; Clubb RT J. Biomol. NMR 2016, 64, 197–205. [PubMed: 26852413]

(19). Dorr BM; Ham HO; An C; Chaikof EL; Liu DR Proc. Natl. Acad. Sci 2014, 111, 13343–13348. [PubMed: 25187567]

(20). Chen I; Dorr BM; Liu DR Proc. Natl. Acad. Sci 2011, 108, 11399–11404. [PubMed: 21697512]

(21). Antos JM; Truttmann MC; Ploegh HL Recent Advances in Sortase-Catalyzed Ligation Methodology. Current Opinion in Structural Biology. 2016, pp 111–118. [PubMed: 27318815]

(22). Dasgupta S; Samantaray S; Sahal D; Roy RP J. Biol. Chem 2011, 286, 23996–24006. [PubMed: 21566128]

(23). Bellucci JJ; Bhattacharyya J; Chilkoti A Angew. Chemie - Int. Ed 2015, 54, 441–445.

(24). Danne C; Dramsi S Res. Microbiol 2012, 163, 645–658. [PubMed: 23116627]

(25). Spirig T; Weiner EM; Clubb RT Mol. Microbiol 2011, 82, 1044–1059. [PubMed: 22026821]

(26). Ton-That H; Schneewind O Trends Microbiol. 2004, 12, 228–234. [PubMed: 15120142]

(27). Echelman DJ; Alegre-Cebollada J; Badilla CL; Chang C; Ton-That H; Fernández JM Proc. Natl. Acad. Sci 2016, 113, 2490–2495. [PubMed: 26884173]

(28). Yeates TO; Clubb RT Biochemistry: How Some Pili Pull. Science. 2007, pp 1558–1559.

(29). Chang C; Amer BR; Osipiuk J; McConnell SA; Huang I-H; Hsieh V; Fu J; Nguyen HH; Muroski J; Flores E; Ogorzalek Loo RR; Loo JA; Putkey JA; Joachimiak A; Das A; Clubb RT; Ton-That H Proc. Natl. Acad. Sci. U. S. A 2018, 115, E5477–E5486. [PubMed: 29844180]

(30). Jacobitz AW; Naziga EB; Yi SW; McConnell SA; Peterson R; Jung ME; Clubb RT; Wereszczynski JJ Phys. Chem. B 2016, 120, 8302–8312.

(31). Manzano C; Izoré T; Job V; Di Guilmi AM; Dessen A Biochemistry 2009, 48, 10549–10557. [PubMed: 19810750]

(32). Cozzi R; Zerbini F; Assfalg M; D'Onofrio M; Biagini M; Martinelli M; Nuccitelli A; Norais N; Telford JL; Maione D; Rinaudo CD FASEB J 2013, 27, 3144–3154. [PubMed: 23631841]

(33). Persson K Acta Crystallogr. Sect. D Biol. Crystallogr 2011, 67, 212–217.

(34). Manzano C; Contreras-Martel C; El Mortaji L; Izoré T; Fenel D; Vernet T; Schoehn G; Di Guilmi AM; Dessen A Structure 2008, 16, 1838–1848. [PubMed: 19081060]

(35). Yokoyama K; Nio N; Kikuchi Y Appl. Microbiol. Biotechnol 2004, 64, 447–454. [PubMed: 14740191]

(36). Fontana A; Spolaore B; Mero A; Veronese FM Adv. Drug Deliv. Rev 2008, 60, 13–28. [PubMed: 17916398]

(37). Spicer CD; Davis BG Nat. Commun 2014, 5.

(38). Reddington SC; Howarth M Curr. Opin. Chem. Biol 2015, 29, 94–99. [PubMed: 26517567]

(39). Zakeri B; Fierer JO; Celik E; Chittock EC; Schwarz-Linek U; Moy VT; Howarth M Proc. Natl. Acad. Sci 2012, 109, E690–E697. [PubMed: 22366317]

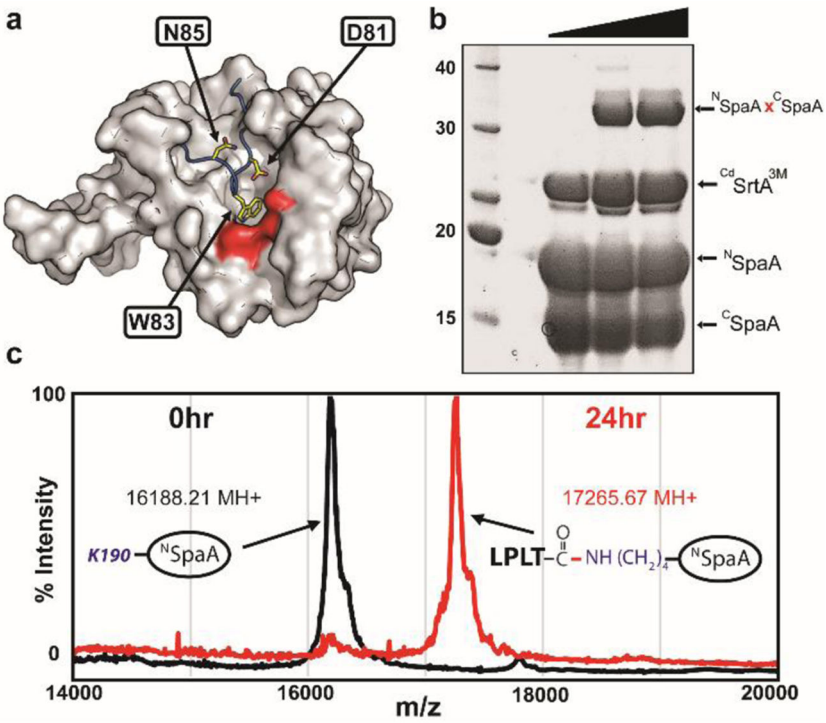


Figure 1.

Mutationally activated ^{Cd}SrtA catalyzes lysine-isopeptide bond formation. (A) The structure of ^{Cd}SrtA^{WT} harbors an inhibitory "lid" structure (blue). Side chains that were mutated to activate the enzyme are shown as yellow sticks. The surface of the catalytic site is colored red. (B) Protein-protein ligation using the activated ^{Cd}SrtA^{3M} enzyme. SDS-PAGE analysis of the reaction demonstrating formation of the lysine-isopeptide linked ^{N}SpaAx^{C}SpaA product. The reaction (100μM enzyme, 300μM ^{C}SpaA and ^{N}SpaA) was sampled at 0, 24 and 48 hours. (C) High yield protein-peptide labeling with ^{Cd}SrtA^{3M}. MALDI-MS data showing that >95% ^{N}SpaA is labeled with peptide containing the sort-tag, LPLTG$_{peptide}$. MALDI-MS spectra recorded at 0 (black) and 24h (red) are overlaid.

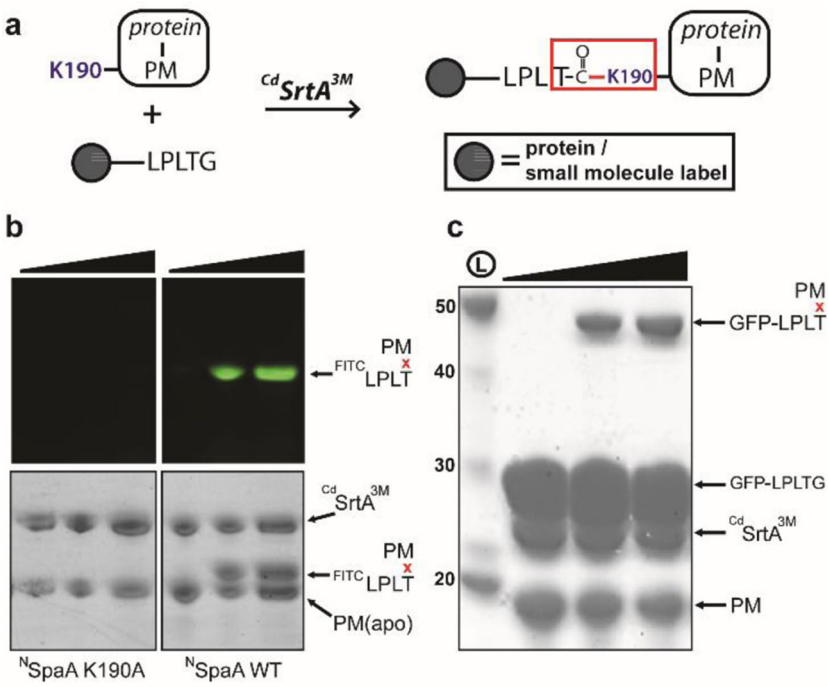


Figure 2.

Labeling proteins via a lysine-isopeptide bond with $^{Cd}SrtA^{3M}$. (A) Schematic showing $^{Cd}SrtA^{3M}$ catalyzed labeling of pilin motif (PM) fusion protein with a protein containing the LPLTG sorting signal or a LPLTG peptide with a functional label. (B) SDS-PAGE analysis of a fluoropeptide modification reaction containing $^{Cd}SrtA^{3M}$ (100μM) and $^{FITC}LPLTG$ (1mM) and either $^{N}SpaA$ (K190A) (lanes 1-3) or $^{N}SpaA$ WT (lanes 4-6) (both 100μM). Top and bottom panels are the same gels visualized by fluorescence or by Coomassie staining, respectively. Reaction progress was measured at 0 (lanes 1,4), 24 (lanes 2,5) and 48hrs (lanes 3,6). (C) Protein-protein ligation with $^{Cd}SrtA^{3M}$. As in panel (B), except reactions contained GFP-LPLTG (300μM) instead of the fluoropeptide. Reactants were visualized with Coomassie staining at 0,24 and 48 hrs (lanes 1-3, respectively).

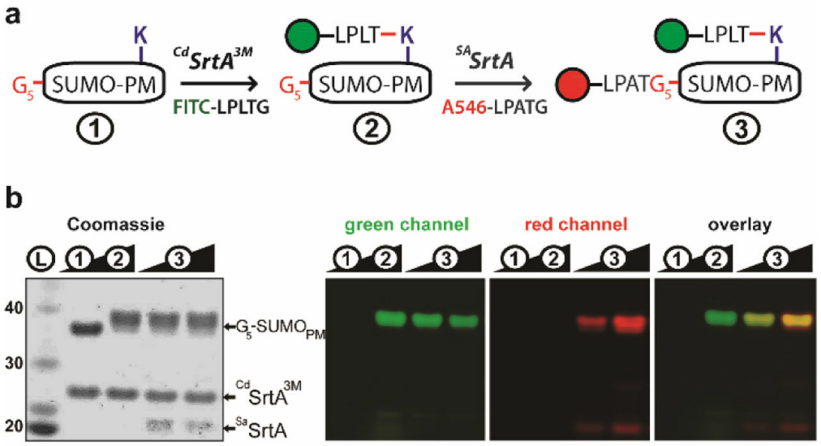
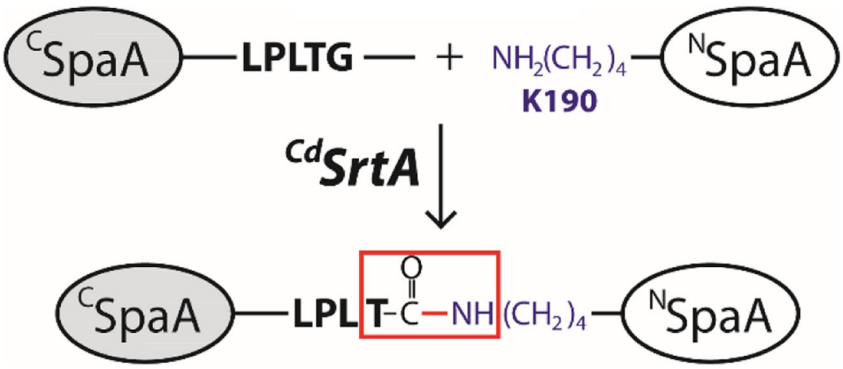


Figure 3.

Orthogonal protein labeling using $^{Cd}SrtA^{3M}$ and $^{Sa}SrtA$. (A) Sequential reaction scheme used to install fluorogenic peptides on a target protein via peptide- and isopeptide bonds. G_5-$SUMO_{PM}$ is a SUMO target protein that is fused to N- and C-terminal nucleophiles, pentaglycine (G_5) and the pilin motif (PM), respectively. (B) SDS-PAGE analysis of reaction mixture taken at different steps in the procedure. 1) prior to labeling, 2) after labeling with ^{FITC}LPLTG using $^{Cd}SrtA^{3M}$ and 3) after labeling with A546-LPATG using ^{Sa}SrtA (0.25/2hr incubations). Panels show as indicated fluorescence gel imaging to detect FITC and A546 fluorophores using 488/530 (green channel) and 532/605mn (red channel) wavelengths for excitation/emission, respectively, and the merged image of the gels demonstrating dual labeling. In the first panel, the same gel was visualized by coomassie staining.



Scheme 1.
^{Cd}SrtA-catalyzed isopeptide bond formation