Synthesis of Non-linear Protein Dimers through a Genetically Encoded Thiol-ene Reaction

Jessica Torres-Kolbus¹, Chungjung Chou¹ⁿ, Jihe Liu², Alexander Deiters¹,²

¹ Department of Chemistry, North Carolina State University, Raleigh, North Carolina, United States of America, ²Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

Abstract

Site-specific incorporation of bioorthogonal unnatural amino acids into proteins provides a useful tool for the installation of specific functionalities that will allow for the labeling of proteins with virtually any probe. We demonstrate the genetic encoding of a set of alkene lysines using the orthogonal PyrRS/PylTCUA pair in Escherichia coli. The installed double bond functionality was then applied in a photoinitiated thiol-ene reaction of the protein with a fluorescent thiol-bearing probe, as well as a cysteine residue of a second protein, showing the applicability of this approach in the formation of heterogeneous non-linear fused proteins.

Introduction

Covalent attachment of proteins to ligands, polymers, and surfaces creates macromolecules combining specific biological function with favorable physical and chemical properties. For example, studying biological processes in their native environment often requires the addition of reporter tags to proteins [1]. To date, the mainstay tagging strategy for imaging of proteins involves genetic fusions of fluorescent proteins [2,3]. However, the large size of fluorescent proteins can interfere with the folding and activity of the targeted protein [4]. Alternatively, tag-mediated labeling methods have been exploited, including self-labeling proteins, such as HaloTag, SNAP-tag, CLIP-tag, and enzyme-mediated labeling [5,6]. Although these methods allow for smaller reporter tags, limitations with regard to the position and the structure of the label remain and the presence of an enzyme is required.

An alternative strategy to label proteins is via the introduction of a single-residue modification, which is nearly non-perturbing. Site-specific protein labeling can be achieved by the installation of tags through bioconjugation reactions with reactive handles previously installed in a protein by using an orthogonal aminoacyl-tRNA synthetase/amiinoacyl-tRNA pair for unnatural amino acid (UAA) mutagenesis [7–10]. The bioorthogonal groups can be installed at virtually any position at the protein expressed in pro- and eukaryotic cells and the choice of probes is nearly limitless. Bioorthogonal chemical handles that have been genetically encoded for conjugation reactions include ketones [11–14], azides [15–20], alkenes [21–28], alkynes [24,29–32], tetrazines [33], aryl halides [34,35], and aryl boronates [36]. The alkene functionality is currently receiving considerable attention due to its versatility in organic transformations and it is rarely found in natural proteins [37,38], allowing for selective modification. Carbon-carbon double bonds have been exploited for protein modification in reactions including olefin-metathesis [26,39], photoaddition of tetrazoles [25,27,28], inverse electron demand Diels-Alder cycloadditions [23,24,30], and thiol-ene reactions [22,40–44].

Bioorthogonal reactions have been applied in a variety of site-specific modifications of proteins such as fluorescent labeling, PEGylation, biotinylation, post-translational modification mimics, and surface immobilization [7,9,45–47]. Another area of interest for which bioconjugation reactions have been explored is in the generation of non-linear protein fusions. In biological systems, proteins often bind to other proteins to gain stability, affinity and higher specificity to perform specific cellular functions such as signal transduction, transcriptional regulation, and DNA repair [48–50]. Elucidation of many of these processes have led to the generation of chemical and biosynthetic methods to create non-linear protein linkages post-translationally for the control and performance of a number of functions, as well as protein trafficking and isolation. Methods that have been explored include native chemical ligation [51–54], enzyme based strategies [55,56], and conjugation employing reactions with UAA residues [57–63].

The introduction of UAAs at a specific position allows for the labeling of proteins with virtually any probe. We demonstrate the genetic encoding of a set of alkene lysines using the orthogonal PyrRS/PylTCUA pair in Escherichia coli. The installed double bond functionality was then applied in a photoinitiated thiol-ene reaction of the protein with a fluorescent thiol-bearing probe, as well as a cysteine residue of a second protein, showing the applicability of this approach in the formation of heterogeneous non-linear fused proteins.


Copyright: © 2014 Torres-Kolbus et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Supporting Information files.

Competing Interests: The authors have declared that no competing interests exist.

Funding: This work was supported by the National Science Foundation (CHE-0848398). JTK received support from the National Science Foundation Graduate Research Fellowship (NSF-0750733). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

* Email: deiters@pitt.edu

Current address: Research and Development Center for Membrane Technology and Department of Chemical Engineering, Chung Yuan Christian University, Jhong-Li, Taoyuan, Taiwan
The thiol-ene reaction involves a radical-mediated addition of a thiol to an alkene that occurs upon UV irradiation (365–405 nm) [64,65]. The reaction offers the possibility of using light to control both in space and time the formation of a stable thioether bond. As a result of its specificity for alkenes and compatibility with aqueous environments, the thiol-ene reaction is a bioorthogonal reaction that has been applied in polymer and material synthesis [66–72], carbohydrate modification [73,74], and peptide and protein modification [21,22,29,40–44]. Recently, orthogonal thiol-ene bioconjugations applying alkanyl UAAs and synthetic organic reaction partners have been reported [21,22]. In order to expand the chemical diversity of these orthogonal handles, we demonstrate the synthesis, incorporation and protein heterodimer formation using alternative thiol-ene reaction conditions.

**Results and Discussion**

**Incorporation of alkene lysines into proteins**

The pyrrolysyl-tRNA synthetase (PylRS), found in certain methanogenic archa and bacteria, directly charges pyrrolysin (Pyl) onto its cognate tRNA that subsequently delivers it in response to an in-frame amber stop codon, TAG [75–77]. Furthermore, it has been demonstrated that the pyrrolysyl-rRNA synthetase/pyrrolysyl-tRNA_CUA pairs from Methanococcus marburgi (MbPylRS/PylTCUA) and M. mazei (MmPylRS/PylTCUA) are functional in Escherichia coli [78], Saccharomyces cerevisiae [79], mammalian cells [80], Caenorhabditis elegans [81,82], Drosophila melanogaster [83], and Xenopus laevis [84]. Furthermore, the wild-type PylRS is capable of accommodating a broad range of unnatural lysine derivatives based on a carbamate linkage at the e-amino group to which a variety of functional groups, including tert-butyl [85], azido, alkynyl [17], norbornene [23], and diazirine [86] have been accommodated. We first synthesized a small collection of aliphatic alkene-lysines to diversify the structure of bioconjugation handles and to explore the ability to accommodate long-chain alkynes and lysine-linkages other than carbamates by the MbPylRS (Figure 1A and Schemes S1–S4).

To investigate whether the synthesized alkene-lysines are substrates for the wild-type MbPylRS, the incorporation efficiencies of 1–9 into myoglobin were evaluated by protein expression in E. coli. Cells were grown in the absence of an UAA and in the presence of 1–9. The amino acids 1, 2 and 3 have been previously described and incorporated into proteins using wild-type PylRS and/or PylRS mutants [22,85]. Here we found that additional analogs can be efficiently incorporated into myoglobin by the MbPylRS. The obtained incorporation efficiencies and ESI-MS results are listed in Figure 1B and the corresponding SDS-PAGE analysis is shown in Figure S1.

Previous crystallographic studies of PyRS have indicated that the synthetase holds a large hydrophobic pocket, capable of accommodating bulky and hydrophobic moieties [87,88]. In addition, it has been found that the carbamate moiety at the lysine side-chain is an essential discriminator for substrate recognition. For instance, the oxygen atom adjacent to the side-chain carbonyl group in 1 interacts via a water-mediated hydrogen bond with the side-chain carbonyl group of Asn346, a key residue in establishing substrate recognition in PylRS [85,87]. We found that the amino acid binding pocket of MbPylRS exhibited flexibility to accommodate substrates 1, 2, 3, 6 and 7 with amino acids 1 and 7 showing the highest incorporation efficiency, which could be explained by their smaller size. While the amino acids 4 and 5 were not efficiently incorporated into protein due to their longer carbon chains.

The successful incorporation of 1 and 7 into protein together with the inefficient substrate recognition of 8 and 9 by MbPylRS suggests that the presence of an oxygen atom adjacent to the side-chain carbonyl group favors the hydrogen-bond network to be established more efficiently. We hypothesize that the recognition of 7 by MbPylRS may be possible by re-directing the necessary interactions of the synthetase’s binding pocket to the α-position. Moreover, we have previously observed a preference for the carbamate moiety over an amide group to drive the efficient genetic encoding of e-C-carbamate linkage (1–6), an inverted carbamate 7, an amide 8, and an urea 9. (B) Myoglobin comparative incorporation efficiencies (%) and ESI-MS results. doi:10.1371/journal.pone.0105467.g001

The thiol-ene reaction involves a radical-mediated addition of a thiol to an alkene that occurs upon UV irradiation (365–405 nm) [64,65]. The reaction offers the possibility of using light to control both in space and time the formation of a stable thioether bond. As a result of its specificity for alkenes and compatibility with aqueous environments, the thiol-ene reaction is a bioorthogonal reaction that has been applied in polymer and material synthesis [66–72], carbohydrate modification [73,74], and peptide and protein modification [21,22,29,40–44]. Recently, orthogonal thiol-ene bioconjugations applying alkanyl UAAs and synthetic organic reaction partners have been reported [21,22]. In order to expand the chemical diversity of these orthogonal handles, we demonstrate the synthesis, incorporation and protein heterodimer formation using alternative thiol-ene reaction conditions.

**Results and Discussion**

**Incorporation of alkene lysines into proteins**

The pyrrolysyl-tRNA synthetase (PylRS), found in certain methanogenic archa and bacteria, directly charges pyrrolysin (Pyl) onto its cognate tRNA that subsequently delivers it in response to an in-frame amber stop codon, TAG [75–77]. Furthermore, it has been demonstrated that the pyrrolysyl-rRNA synthetase/pyrrolysyl-tRNA_CUA pairs from Methanococcus marburgi (MbPylRS/PylTCUA) and M. mazei (MmPylRS/PylTCUA) are functional in Escherichia coli [78], Saccharomyces cerevisiae [79], mammalian cells [80], Caenorhabditis elegans [81,82], Drosophila melanogaster [83], and Xenopus laevis [84]. Furthermore, the wild-type PylRS is capable of accommodating a broad range of unnatural lysine derivatives based on a carbamate linkage at the e-amino group to which a variety of functional groups, including tert-butyl [85], azido, alkynyl [17], norbornene [23], and diazirine [86] have been accommodated. We first synthesized a small collection of aliphatic alkene-lysines to diversify the structure of bioconjugation handles and to explore the ability to accommodate long-chain alkynes and lysine-linkages other than carbamates by the MbPylRS (Figure 1A and Schemes S1–S4).

To investigate whether the synthesized alkene-lysines are substrates for the wild-type MbPylRS, the incorporation efficiencies of 1–9 into myoglobin were evaluated by protein expression in E. coli. Cells were grown in the absence of an UAA and in the presence of 1–9. The amino acids 1, 2 and 3 have been previously described and incorporated into proteins using wild-type PylRS and/or PylRS mutants [22,85]. Here we found that additional analogs can be efficiently incorporated into myoglobin by the MbPylRS. The obtained incorporation efficiencies and ESI-MS results are listed in Figure 1B and the corresponding SDS-PAGE analysis is shown in Figure S1.

Previous crystallographic studies of PyRS have indicated that the synthetase holds a large hydrophobic pocket, capable of accommodating bulky and hydrophobic moieties [87,88]. In addition, it has been found that the carbamate moiety at the lysine side-chain is an essential discriminator for substrate recognition. For instance, the oxygen atom adjacent to the side-chain carbonyl group in 1 interacts via a water-mediated hydrogen bond with the side-chain carbonyl group of Asn346, a key residue in establishing substrate recognition in PylRS [85,87]. We found that the amino acid binding pocket of MbPylRS exhibited flexibility to accommodate substrates 1, 2, 3, 6 and 7 with amino acids 1 and 7 showing the highest incorporation efficiency, which could be explained by their smaller size. While the amino acids 4 and 5 were not efficiently incorporated into protein due to their longer carbon chains.
With amino acids 1–3 showing good incorporation efficiency, we site-specifically incorporated these amino acids into superfolder Green Fluorescent Protein (sfGFP) as a second model protein in E. coli. We found that alkene lysines 1–3 were successfully introduced at position Y151 in sfGFP (Figure 2A) and that sfGFP yields were obtained at 32–70 mg/L, an approximately 10-fold increase of incorporation efficiency compared to myoglobin bearing the same amino acids 1–3. ESI-MS analysis of purified sfGFP shows molecular weights corresponding to the site-specific incorporation of 1, 2, and 3 (Figure 2B).

### sfGFP labeling via the thiol-ene reaction

To verify that the thiol-ene reaction is suitable for labeling the alkene-bearing sfGFP, dansyl-thiol (10) was used as a fluorescent probe (Figure 3A and Scheme S5). Wild-type sfGFP and modified sfGFPs carrying 1 or 2, which showed the highest incorporation efficiency, were subjected to a thiol-ene reaction with 10 by irradiating the reaction mixture with 365 nm UV light in the presence of the photoinitiator I2959 for 5 min. Both samples were then analyzed by SDS-PAGE gel and in-gel fluorescence imaging. Figure 3B shows that the alkene-containing sfGFPs modified with 1 and 2 were both selectively labeled with 10 after UV irradiation while the wild-type sfGFP was not fluorescently labeled. These results demonstrate that a thiol-containing fluorescence probe could be site-specifically conjugated to sfGFP bearing an alkene functional group.

In order to show the potential of the thiol-ene reaction in protein chemistry, we hypothesized that cysteine residues in another protein could also be used as a possible reaction partner, leading to the formation of a non-linear protein heterodimer (Figure 3A). Lysozyme is a small protein containing 8 cysteine residues within 129 amino acids [91]. The cysteines form 4 disulfide bonds and can be reduced to release free thiol groups. Analysis of bioconjugated proteins by SDS-PAGE revealed bands of expected molecular weight, as the bands corresponding to sfGFP increased from 28 kD to 44 kD via conjugation to lysozyme after UV exposure in the presence of the photoinitiator I2959 for 10 min (Figure 3C). This result indicates that the majority of the observed products are sfGFP-lysozyme heterodimers since lysozyme was supplied in 4 fold excess compared to alkynyl sfGFP. Without UV irradiation, no significant mobility shift was observed. As expected, wild-type sfGFP did not undergo a thiol-ene reaction with lysozyme. Overall, a successful protein-protein heterodimer formation via thiol-ene conjugation of an alkene-containing protein was achieved.

In both bioconjugation strategies we found that the addition of sodium dodecyl sulfate (SDS) was necessary for an efficient and specific conjugation reaction to alkene-labeled proteins within 5–10 min, in contrast to previously reported 1–2 h reaction times [22,29,43], thus significantly reducing UV exposure. We found that under our experimental conditions, lysozyme is (at least partially) denatured [92,93], as confirmed by circular dichroism (CD) spectroscopy (Figure S2). This may result in more accessible cysteine residues and facilitate the thiol-ene bioconjugation reaction. Moreover, as a well-known surfactant, SDS has been proposed to form micelles in thiol-ene reactions for water-based polymerization reactions [94,95]. It is possible that the association of the proteins with micelles may increase their local concentration, thus further facilitating the reaction.

### Conclusions

In conclusion, we have synthesized a collection of alkene lysines of varying length and α-linkages and demonstrated their site-specific, genetically encoded incorporation into proteins in E. coli by the wild-type MbpRS/PylTCUA pair. The alkene-containing amino acids 1–3 showed the highest incorporation efficiencies into myoglobin and protein yields decreased with increasing side-chain length, hinting the limitations of the wild-type synthetase’s binding pocket to accommodate sterically demanding amino acids. Among these amino acids, we also successfully incorporated the amino acid 7 with an inverted carbamate functionality at the ε-position of lysine. Replacement of the carbamate motif for an amide or urea failed to provide efficient incorporation into protein, once again suggesting that the carbamate moiety at the lysine side-chain can be an essential discriminator for substrate recognition by wild-type PylRS.

Next, the alkene amino acids 1, 2, and 3 were successfully incorporated into sfGFP, with 1 and 2 exhibiting the highest incorporation efficiency. Utilizing the thiol-ene reaction, alkene-bearing sfGFP was site-specifically bioconjugated to a dansyl-thiol fluorophore (10) upon irradiation with 365 nm of UV light in the presence of photoinitiator I2959 after only 5 min. In addition, we applied the site-specific genetic incorporation of alkene-bearing amino acids into proteins in the direct, spacer-free synthesis of a non-linear protein fusion of sfGFP and lysozyme. All components are recombinantly expressed and no post-translational introduction of functional groups was required. The work described herein demonstrates for the first time the assembly of a protein heterodimer by means of a light-induced thiol-ene ligation using genetically encoded alkene-bearing UAs. This approach may become a promising tool to create non-linear proteins directly, with minimal synthetic effort, by creating direct protein-to-protein conjugations.

### Materials and Methods

#### Synthesis of alkene lysines: general considerations

Unless otherwise stated, all reagents used were commercial reagents used without purification and reactions were performed under nitrogen using flame-dried glassware. The $^1$H NMR and $^{13}$C NMR spectra were recorded on a 300 MHz or 400 MHz
Varian NMR spectrometer. The amino acid 1 was purchased from Chem-Impex International, Inc. For synthesis schemes of 2–10, please refer to Schemes S1–S5.

Procedure for the synthesis of 2a–5a

But-3-en-1-yl (2,5-dioxopyrrolidin-1-yl) carbonate (2a). N,N'-Disuccinimidyl carbonate (353 mg, 1.38 mmol) was added to a solution of 3-buten-1-ol (60 μL, 0.69 mmol) and triethylamine (288 μL, 2.07 mmol) in dry acetonitrile (5 mL) at room temperature. The resulting mixture was stirred overnight and then concentrated under vacuum. The product was purified by column chromatography on SiO2 gel, eluted with 97:2:1 DCM/acetone/TEA to deliver 2a (96 mg, 65%) as a colorless oil. 

1H-NMR (400 MHz, CDCl3): δ 5.73 (m, 1 H), 5.15–5.08 (m, 2 H), 4.31 (t, J = 6.4 Hz, 2 H), 2.78 (s, 4 H), 2.46 (m, 2 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 168.8, 151.5, 132.4, 118.4, 70.2, 32.7, 25.4 ppm.

2,5-Dioxopyrrolidin-1-yl pent-4-en-1-yl carbonate (3a). Compound 3a (193 mg, 73%) was obtained as a colorless oil from 4-penten-1-ol (0.12 mL, 1.16 mmol) by following the procedure described above. 1H-NMR (400 MHz, CDCl3): δ 5.73 (m, 1 H), 5.15–5.08 (m, 2 H), 4.31 (t, J = 6.4 Hz, 2 H), 2.78 (s, 4 H), 2.04 (m, 2 H), 1.85 (m, 2 H), 1.45 (m, 2 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 168.6, 151.4, 136.5, 115.8, 70.6, 29.2, 27.3, 25.3 ppm.

2,5-Dioxopyrrolidin-1-yl hex-5-en-1-yl carbonate (4a). Compound 4a (232 mg, 77%) was obtained as a colorless oil from 5-hexen-1-ol (0.15 mL, 1.25 mmol) by following the procedure described above. 1H-NMR (400 MHz, CDCl3): δ 5.73 (m, 1 H), 5.05–4.91 (m, 2 H), 4.27 (t, J = 6.4 Hz, 2 H), 2.77 (s, 4 H), 2.04 (m, 2 H), 1.70 (m, 2 H), 1.45 (m, 2 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 169.2, 151.9, 138.2, 115.5, 71.7, 33.3, 28.0, 25.7, 24.9 ppm.

2,5-Dioxopyrrolidin-1-yl hept-6-en-1-yl carbonate (5a). Compound 5a (202 mg, 67%) was obtained as a colorless oil from 6-hepten-1-ol (0.16 mL, 1.17 mmol) by following the procedure described above. 1H-NMR (400 MHz, CDCl3): δ 5.77 (m, 1 H), 5.00–4.91 (m, 2 H), 4.28 (t, J = 6.8 Hz, 2 H), 2.80 (s, 4 H), 2.05 (m, 2 H), 1.72 (m, 2 H), 1.39 (m, 4 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 183.6, 168.9, 151.6, 138.5, 114.7, 71.6, 33.5, 28.3, 25.5, 24.9 ppm.

Procedure for the synthesis of 2b–5b

(S)-6-(((But-3-en-1-yloxy)carbonyl)amino)-2-((tert-butoxycarbonyl)amino)hexanoic acid (2b). Boc-L-Lys-OH (218 mg, 0.88 mmol) was added to a stirred solution of 2a (157 mg, 0.74 mmol) in dry DMF (2 mL). The reaction was allowed to continue overnight at room temperature. The mixture was diluted in water (10 mL) and extracted with EtOAc (3 × 10 mL). The combined organic layers were washed with water (3 × 20 mL) and brine (10 mL). The resulting oil was dried over Na2SO4, filtered, and concentrated in vacuo to furnish 2b (219 mg, 86%) as an off-white foam. 1H-NMR (400 MHz, CDCl3): δ 5.73 (m, 1 H), 5.15–5.08 (m, 2 H), 4.31 (t, J = 6.4 Hz, 2 H), 2.78 (s, 4 H), 2.34 (m, 2 H), 1.81–1.40 (m, 15 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 176.5, 151.4, 136.5, 115.8, 70.6, 29.2, 27.3, 25.5 ppm.

(S)-2-((tert-Butoxycarbonyl)amino)-6-(((pent-4-en-1-yloxy)carbonyl)amino)hexanoic acid (3b). Compound 3a (202 mg, 67%) was obtained as a colorless oil from 6-hepten-1-ol (0.16 mL, 1.17 mmol) by following the procedure described above. 1H-NMR (400 MHz, CDCl3): δ 5.73 (m, 1 H), 5.15–5.08 (m, 2 H), 4.31 (t, J = 6.4 Hz, 2 H), 2.78 (s, 4 H), 2.34 (m, 2 H), 1.81–1.40 (m, 15 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 176.5, 151.4, 136.5, 115.8, 70.6, 29.2, 27.3, 25.5 ppm.

(S)-6-(((But-3-en-1-yloxy)carbonyl)amino)-2-((tert-butoxycarbonyl)amino)hexanoic acid (2b). Boc-L-Lys-OH (218 mg, 0.88 mmol) was added to a stirred solution of 2a (157 mg, 0.74 mmol) in dry DMF (2 mL). The reaction was allowed to continue overnight at room temperature. The mixture was diluted in water (10 mL) and extracted with EtOAc (3 × 10 mL). The combined organic layers were washed with water (3 × 20 mL) and brine (10 mL). The resulting oil was dried over Na2SO4, filtered, and concentrated in vacuo to furnish 2b (219 mg, 86%) as an off-white foam. 1H-NMR (400 MHz, CDCl3): δ 5.73 (m, 1 H), 5.15–5.08 (m, 2 H), 4.31 (t, J = 6.4 Hz, 2 H), 2.78 (s, 4 H), 2.34 (m, 2 H), 1.81–1.40 (m, 15 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 176.5, 151.4, 136.5, 115.8, 70.6, 29.2, 27.3, 25.5 ppm.
Procedure for the Boc-deprotection, 2-5

(S)-2-Amino-6-(((but-3-en-1-yl)oxycarbonyl)amino)hexanoic acid HCl salt (2). To a solution of 2b (110 mg, 0.32 mmol) and Et3SiH (0.1 mL, 0.64 mmol) in dry DCM (4.5 mL), trifluoroacetic acid (0.24 mL, 3.2 mmol) was added dropwise, and the reaction mixture was allowed to stir at room temperature overnight. The volatiles were removed under reduced pressure and the residue was dissolved in a solution of 4 N HCl in 1:4-dioxane (0.25 mL) and DCM (0.75 mL), allowed to stir for 10 min at room temperature and then concentrated. The latter process was repeated two more times to ensure complete TFA to HCl salt exchange. The concentrated residue was dissolved in a minimal amount of MeOH and was precipitated into ice-cold Et2O. The precipitate was pelleted by centrifugation, the supernatant decanted, and the solid was washed with Et2O before drying under vacuum, affording the amino acid 2 (32.2 mg, 92%) as a white solid.1H-NMR (400 MHz, DMSO-d6): δ 8.45 (s, br, 3 H), 7.05 (s, br, 1 H), 5.75 (m, 1 H), 5.10–5.01 (m, 2 H), 3.51 (m, 2 H), 2.94 (m, 2 H), 2.87 (m, 2 H), 1.78 (m, 2 H); 13C-NMR (75 MHz, DMSO-d6): δ 110.5, 108.6, 104.9, 95.7, 78.2, 41.7, 37.7 (t, J = 6–9 Hz, 1 H), 2.92 (m, 2 H), 2.23 (m, 2 H), 1.75 (m, 2 H), 1.36–1.26 (m, 4 H); HRMS-ESI (m/z): [M+H]+ calcd for C16H20N2O4 287.1684, found 287.1686.

(S)-2-Amino-6-(((pent-4-en-1-yl)oxycarbonyl)amino)hexanoic acid HCl salt (3). Deprotection of 3b (0.5 g, 1.89 mmol) was performed as described above to obtain 3 (0.4 g, 97%) as a white solid.1H-NMR (400 MHz, D2O): δ 5.56 (m, 1 H), 5.07–4.98 (m, 2 H), 4.05–4.00 (m, 3 H), 3.11 (t, J = 5.2 Hz, 2 H), 2.10 (m, 2 H), 1.97–1.87 (m, 2 H), 1.69 (t, J = 6.4 Hz, 2 H), 1.56–1.39 (m, 4 H); 13C-NMR (75 MHz, D2O): δ 170.2, 158.9, 138.6, 114.9, 65.0, 52.7, 39.9, 29.6, 29.4, 28.5, 27.5, 21.6 ppm; HRMS-ESI (m/z): [M+H]+ calcd for C16H20N2O4 287.1684, found 287.1686.

Synthesis of Non-Linear Protein Dimers

(S)-2-Amino-6-(((hex-5-en-1-yl)oxycarbonyl)amino)hexanoic acid HCl salt (4). Deprotection of 4b (145 mg, 0.39 mmol) was performed as described above to obtain 4 (104.4 mg, 96%) as a white solid.1H-NMR (400 MHz, DMSO-d6): δ 8.28 (s, br, 3 H), 7.00 (s, br, 1 H), 5.79 (m, 1 H), 5.02–4.93 (m, 2 H), 3.91 (t, J = 6.8 Hz, 2 H), 3.67 (s, br, 2 H), 2.94 (m, 2 H), 2.03 (m, 2 H), 1.76 (m, br, 2 H), 1.52 (t, J = 6.0 Hz, 2 H), 1.30 (m, br, 6 H); 13C-NMR (100 MHz, DMSO-d6): δ 171.1, 156.3, 139.5, 115.0, 63.4, 52.3, 32.0, 29.6, 28.0, 24.6, 21.7 ppm; HRMS-ESI (m/z): [M+H]+ calcd for C17H18NO4 273.1809, found 273.1803.

(S)-2-Amino-6-(((hex-5-en-1-yloxy)carbonyl)amino)hexanoic acid HCl salt (5). Deprotection of 5b (130 mg, 0.336 mmol) was performed as described above to obtain 5 (104.4 mg, 96%) as a white solid.1H-NMR (400 MHz, DMSO-d6): δ 8.45 (s, br, 3 H), 7.09 (m, 1 H), 5.78 (m, 1 H), 5.01–4.92 (m, 2 H), 3.90 (t, J = 6.4 Hz, 2 H), 3.82 (s, br, 1 H), 2.93 (m, 2 H), 2.01 (m, 2 H), 1.77 (m, br, 2 H), 1.52 (m, 2 H), 1.45–1.28 (m, 10 H); 13C-NMR (100 MHz, DMSO-d6): δ 171.0, 156.3, 138.7, 114.6, 63.5, 51.8, 33.1, 29.6, 28.9, 28.6, 24.9, 21.6 ppm; HRMS-ESI (m/z): [M+H]+ calcd for C17H18NO4 273.1809, found 273.1803.
Procedure for the synthesis of 7

(S)-6-[(Allylcarbamoyl)oxy]-2-[(tert-butoxycarbonyl)amino]hexanoic acid (7a). 6-Hydroxy-Boc-L-norleucine-OH (25 mg, 0.10 mmol) was dissolved in dry DCM (1 mL) and DPEA (55 µL, 0.30 mmol). The solution was chilled to 0°C before the addition of allyl isocyanate (18 µL, 0.20 mmol) and the reaction was allowed to proceed at 40°C overnight. After cooling to room temperature, the mixture was diluted with DCM (3 mL) and 5% citric acid (4 mL) was added. The aqueous layer was extracted with DCM (3×4 mL) and the combined organic layers were washed with water (10 mL) and brine (5 mL). The resulting organic layer was dried over Na2SO4, filtered and concentrated in vacuo to dryness to furnish 7a (29 mg, 89% yield) as an off-white foam. 1H-NMR (400 MHz, CDCl3): δ 5.85 (m, 1 H), 5.24–5.07 (m, 2 H), 4.74 (m, br, 1 H), 4.29 (s, br, 1 H), 4.06 (t, J = 5.6 Hz, 2 H), 3.78 (m, 2 H), 1.93–1.25 (m, 15 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 176.0, 155.8, 135.0, 116.4, 80.3, 64.8, 53.3, 43.2, 32.2, 28.7, 28.5, 22.0 ppm; HRMS-ESI (m/z): [M+Na]+ calcd for C13H29N2O6 333.1689, found 333.1654.

(S)-6-[(Allylcarbamoyl)oxy]-2-amino-hexanoic acid TFA salt (7). Deprotection of 7a (28 mg, 0.085 mmol) was performed by following the procedure described for compound 6 to afford compound 7 (28.3 mg, 96% as a white solid. 1H-NMR (400 MHz, D2O): δ 7.53 (m, 1 H), 5.19–4.80 (m, 2 H), 4.08 (t, J = 6.0 Hz, 1 H), 3.94 (t, J = 5.6 Hz, 2 H), 3.72 (m, 2 H), 1.95 (m, 2 H), 1.69 (m, 2 H), 1.50 (m, 2 H) ppm; 13C-NMR (100 MHz, D2O): δ 173.2, 159.0, 135.4, 115.1, 63.1, 53.7, 42.1, 29.7, 28.0, 21.0 ppm; HRMS-ESI (m/z): [M+H]+ calcd for C13H29N2O6 353.1689, found 353.1654.

Procedure for the synthesis of 8

(S)-2-[(tert-Butyoxycarbonyl)amino]-6-[(pent-4-enamido)hexanoic acid (8a). Compound 8a (212 mg, 97%) was obtained as an off-white foam from 2,5-dioxopyrrolidin-1-yl pent-4-enoate [96] (132 mg, 0.67 mmol) by following the procedure described for compound 2. 1H-NMR (400 MHz, CDCl3): δ 8.56 (s, br, 1 H), 6.20 (s, br, 1 H), 5.73 (m, 1 H), 5.34 (m, 1 H), 5.06–4.97 (m, 2 H), 4.25 (m, br, 1 H), 4.46 (s, br, 1 H), 3.23 (m, 2 H), 2.34 (m, 2 H), 2.27 (m, 2 H), 1.82–1.66 (m, br, 2 H), 1.59–1.35 (m, 13 H) ppm; 13C-NMR (100 MHz, CDCl3): δ 175.4, 173.5, 156.0, 137.0, 115.0, 80.1, 53.2, 39.3, 35.8, 32.3, 29.8, 28.9, 22.5 ppm; HRMS-ESI (m/z): [M+H]+ calcd for C15H27N3O5 352.1848, found 352.1845.

(S)-2-Amino-6-(pent-4-enamido)hexanoic acid HCl salt (8). Deprotection of 8a (175 mg, 0.54 mmol) was performed by following the procedure described for compound 2 to afford compound 8 (142 mg, 99%) as a white solid. 1H-NMR (400 MHz, DMSO-d6): δ 8.45 (s, br, 3 H), 7.93 (m, br, 1 H), 5.77 (m, 1 H), 5.02–4.92 (m, 2 H), 3.95 (m, br, 1 H), 3.81 (s, br, 1 H), 3.00 (m, 2 H), 2.25–2.14 (m, 4 H), 1.79 (m, br, 2 H), 1.40–1.29 (m, br, 4 H) ppm; 13C-NMR (100 MHz, DMSO-d6): δ 171.3, 171.0, 157.8, 115.0, 51.9, 38.0, 34.5, 29.6, 29.3, 28.6, 21.7 ppm; HRMS-ESI (m/z): [M+H]+ calcd for C11H20N2O3 235.1430, found 235.1430.

Myoglobin Expression in E. coli

Plasmids, pMyo4TAGpylT and pBKpylS, were co-transformed into E. coli Top10 cells as previously described [98] and selected with 25 µg/mL tetracycline and 50 µg/mL kanamycin. A single colony was used to inoculate 2 mL LB medium containing the corresponding antibiotics and grown overnight. Next, 500 µL of culture was used to seed 30 mL of LB culture containing 1 mM of the corresponding UAs and antibiotics. The pH was adjusted to 7 with 10 M NaOH immediately before inoculation. Cells were then cultivated to OD600 = 0.6 and 100 µL of 20% arabinosine solution was supplemented to induce arabinose promoter driven expression. The cells were cultivated at 37°C shaker overnight and harvested by centrifugation at 3000 g in standard 50 mL conical tubes. Lysis of the cell was conducted by re-suspending the cell pellets with standard Ni-NTA phosphate lysis buffer with lysozyme and 0.1% Triton X-100. After 1 hour of incubation at 4°C, cells were sonicated on ice to release the soluble portion and debris was removed by centrifugation. The cleared lysates were incubated with NiNTA wash buffer. The protein was eluted with 400 mM NaCl. The eluted sample was mixed with SDS loading buffer, heated at 95°C for 5 min and loaded onto 10% SDS-PAGE gel with 1.5 mm thickness and ran at 150 V for 50 min. The gel was stained overnight with Coomassie blue solution (0.1% Coomassie blue, 10% acetic acid, 40% ethanol), then de-stained (10% acetic acid, 40% ethanol) and analyzed (Figure S1). The protein was dialyzed in 1 L of 20 mM ammonium acetate buffer for mass spectrum analysis.
sfGFP Expression in *E. coli*

The plasmid pMyo4TAGpylT [98] was modified by replacing the myoglobin coding sequence with the sfGFP gene with an amber stop codon mutation placed on Y151 position located on the outer beta sheet domain. The co-transformation was the same as described above but condensed culture protocol was used to maximize UAA yields. The 2 mL of overnight culture was scaled-up to 400 mL culture in 2×1 L Erlenmeyer flask and grown to OD<sub>600</sub> = 0.6. Cells were harvested in 4×50 mL conical tubes and re-suspended in 50 mL of LB medium containing 1 mM of the corresponding UAA, antibiotics, and 0.1% arabinose. Cells were re-suspended by incubating in a rotary shaker at 37°C for 10 min and collected in a 250 mL Erlenmeyer flask. The cells were induced for 4 h and harvested by centrifugation. Cell pellets were first suspended by 3.6 mL of 50 mM Tris-HCl pH 8.0, supplemented with 2.4 mL of 4 M ammonium sulfate and extracted by three-phase partitioning method [99] with 6 mL of t-butanol and vigorous shaking. The aqueous bottom layer containing sfGFP was removed and dialyzed against 1 L Ni-NTA lysis buffer for 1–2 h to remove most of the ammonium sulfate. The dialyzed samples were filtered through 0.45 μm disc filter before loading into Ni-NTA gravity column containing 0.5 mL bed volume. The proteins were bound and washed with 12 mL bed volume of lysis buffer, 6 mL bed volume of wash buffer containing 50 mM imidazole and eluted with Ni-NTA elution buffer. Samples were analyzed by SDS-PAGE, dialyzed against PBS pH 7.4 for subsequent labeling reaction and then dialyzed against 20 mM ammonium acetate for mass spectrum analysis.

**Protein MS Analysis**

Protein MS was measured at the Genomics and Proteomics Core Laboratories, University of Pittsburgh. The protein solution was adjusted to 5 pmol/μL in 80% acetonitrile and 0.1% aqueous formic acid. The sample was injected into a Bruker micTOF with an Ultimate 3000 HPLC. The results were deconvoluted to calculate the molecular weight using HyStar.

**Thiol-ene Reactions with sfGFP**

A reaction buffer containing 30 μL of 1 M TrisHCl pH 6.8 (120 mM), 50 μL of 10% SDS (2%), 50 μL of 10 mM TCEP (2 mM), and 120 μL of water was made. In another eppendorf tube, a solution of 10 μL photoinitiator I2959 was prepared. Then 62.5 μL of I2959 was added to 250 μL of reaction buffer just before labeling. Next, 2.5 μL of reaction buffer/photoinitiator mix were added to 20 μL of sfGFP (2400 ng/μL) and incubated at room temperature for 10 min. Dansyl-thiol (10) 50X substrate solution was prepared with 10 μL of 100 mM TCEP (10 mM), 20 μL of 25 mM dansyl-thiol in DMSO and 70 μL of deionized water. Next, 16.8 μL of this solution was added to the reaction mixture and incubated at room temperature for another 10 min. Subsequently, 250 μL of 1X SDS loading buffer containing 2-mercaptoethanol and additional 50 μL of 100 mM DTT were prepared for stopping the reaction. Then 12 μL of the reaction samples were aliquoted into 200 μL PCR tubes. Samples were placed on a standard UV transilluminator at 365 nm for 5 min and the reaction was stopped by adding 12 μL of 1X SDS loading buffer to the mixture and heated at 95°C for 5 min. Next, 5 μL of the samples were loaded onto a 10% SDS-PAGE gel with 1.5 mm thickness and ran at 150 V for 50 min. After electrophoration, gels were rinsed briefly with deionized water and imaged. Gels were stained with coomassie blue and scanned to visualize the protein bands.

For protein heterodimer formation, the thiol-ene conjugation was carried out with a denatured and reduced lysozyme solution. A reaction buffer containing 120 μL of 1 M TrisHCl pH 6.8, 200 μL of 10% SDS, and 1 mL of 10 mM TCEP was prepared. In 1.32 mL of the reaction buffer, 19 mg of lysozyme were dissolved (1 mM). The solution was sealed in a 2 mL micro-centrifugation tube with a rubber septum and purged with nitrogen for 30 min. The tube containing the protein solution was then heated at 75°C for 30 min. Photo-initiator I2959 was diluted to 10 mM in a solution of 50% DMSO in water. sfGFP solutions were adjusted to a concentration of 23 μL (650 ng/μL) and 20 μL of this solution was mixed with 2 μL of reduced lysozyme and 0.5 μL of I2959 in the dark. The PCR tube containing this mixture was then placed on a standard UV transilluminator at 365 nm for 10 min. A solution containing 120 μL of 1 M Tris-HCl pH 6.8, 20 μL of 10% SDS, 10 μL of 1 M TCEP and 200 μL of glycerol was prepared and 20 μL of it were immediately added to the reaction solution after irradiation. 15 μL of the resulted sample was loaded onto native or 10% SDS-PAGE gel following standard procedures.

**Circular dichroism analysis of lysozyme**

CD experiments were performed on an Olis Circular Dichroism Spectrophotometer using 0.1 cm quartz cuvettes. A solution containing 19 mg of lysozyme in 1.32 mL 10 mM phosphate buffer pH 7.4 was prepared. SDS was added to the final concentration of 2%. The lysozyme concentration was diluted to 20 μM for the CD experiment, and CD spectra were collected from 195 to 260 nm in 1 nm increments with an integration time of 5 s and a bandwidth of 2 nm. Increased intensity in the far-UV spectrum with the addition of SDS (Figure S2) is in agreement with previous observations [100].

**Supporting Information**

Figure S1 SDS-PAGE analysis for the incorporation of alkene-bearing lysines 1-9 into myoglobin.  -AA: no UAA was supplemented; +AA: positive control UAA (1 mM); 1-9: myoglobin expression in the presence of the corresponding UAA (1 mM).

(TIF)

Figure S2 Circular dichroism (CD) spectrum of lysozyme with and without SDS treatment. Blue: lysozyme with no SDS; Red: lysozyme with 2% SDS.

(TIF)

Scheme S1 Synthesis of alkene-bearing lysines 2–6.

(TIF)

Scheme S2 Synthesis of alkene-bearing lysine 7.

(TIF)

Scheme S3 Synthesis of alkene-bearing lysine 8.

(TIF)

Scheme S4 Synthesis of alkene-bearing lysine 9.

(TIF)

Scheme S5 Synthesis of dansyl-thiol, 10.

(TIF)

**Author Contributions**

Conceived and designed the experiments: CC AD. Performed the experiments: JTK CC JL. Analyzed the data: JTK CC JL. Contributed to the writing of the manuscript: JTK CC AD.


