This is an open access article published under a Creative Commons Non-Commercial No Derivative Works (CC-BY-NC-ND) Attribution License, which permits copying and redistribution of the article, and creation of adaptations, all for non-commercial purposes.

ACS AUTHORCHOICE

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JACS

Article

# Universal Single-Residue Terminal Labels for Fluorescent Live Cell Imaging of Microproteins

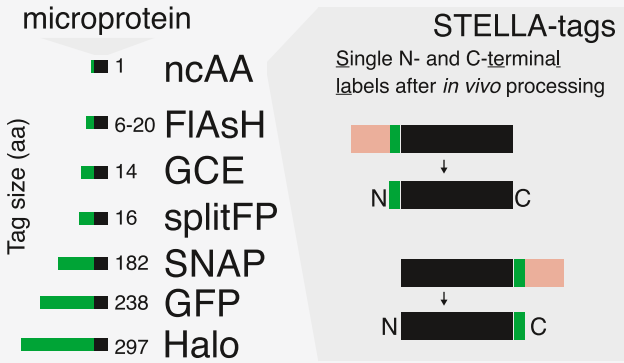
Lorenzo Lafranchi, Dörte Schlesinger, Kyle J. Kimler, and Simon J. Elsässer*

Cite This: *J. Am. Chem. Soc.* 2020, 142, 20080−20087

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Genetically encoded fluorescent tags for visualization of proteins in living cells add six to several hundred amino acids to the protein of interest. While suitable for most proteins, common tags easily match and exceed the size of microproteins of 60 amino acids or less. The added molecular weight and structure of such fluorescent tag may thus significantly affect *in vivo* biophysical and biochemical properties of microproteins. Here, we develop single-residue terminal labeling (STELLA) tags that introduce a single noncanonical amino acid either at the N- or C-terminus of a protein or microprotein of interest for subsequent specific fluorescent labeling. Efficient terminal noncanonical amino acid mutagenesis is achieved using a precursor tag that is tracelessly cleaved. Subsequent selective bioorthogonal reaction with a cell-permeable organic dye enables live cell imaging of microproteins with minimal perturbation of their native sequence. The use of terminal residues for labeling provides a universally applicable and easily scalable strategy, which avoids alteration of the core sequence of the microprotein.

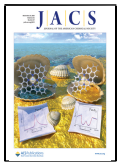



## ■ INTRODUCTION

Across genomes of all kingdoms of life, short open reading frames (sORFs) of 300 bp or less exceed canonical ORFs in number by orders of magnitude. sORFs can be present in mRNA molecules, translated alternatively to the main ORF, but also in otherwise noncoding transcripts, such as long noncoding RNAs, microRNAs, or antisense transcripts.[1−3] Despite their protein-coding potential, sORFs remained relatively understudied in the absence of direct evidence of translation and function on the peptide level. In the past decade, novel sequencing techniques, in particular RNA-sequencing and ribosome profiling, have provided a wealth of new insights into the coding potential of genomes.[4,5] It has become clear that sORF-encoded peptides (SEPs), also called microproteins or micropeptides, are abundantly produced in prokaryotic and eukaryotic cells. Dedicated proteomic methods have provided further direct evidence for the existence of thousands of microproteins in mammalian cells.[6−10] A rapidly increasing number of microproteins are validated through detailed functional studies. For example, MOTS-c, a signaling peptide encoded by the mitochondrial genome, has a systemic role in regulating insulin sensitivity and metabolic homeostasis,[11] whereas the primate-specific microprotein ORF0 is actively shaping our genomes by enhancing the mobility of LINE-1 retrotransposons.[12] Finally, several research groups recently showed that the microprotein Myomixer/Minion/Myomerger is essential for skeletal muscle formation during embryogenesis.[13−15]

Obstacles for systematically characterizing microproteins exist because the classic toolbox for studying protein function *in vivo* and *vitro* has historically been tailored to canonical proteins. Few dedicated techniques have been developed to advance the study of microproteins. Commonly used tags that are fused to a protein of interest for microscopy or biochemical applications range in size from 6 to 250 amino acids. Given the small size of microproteins, the size and biophysical properties of tags become a relevant variable influencing their function or localization.[16−18] Peptide tags, such as HA, FLAG, Strep, are relatively small but not applicable to live cell microscopy. Fluorescent proteins and enzymatic labeling domains (e.g., SNAP-tag[19]) are large. Split fluorescent proteins provide a more versatile alternative[20] in which a ∼16 amino acid peptide tag forms a fluorescent protein with a separately expressed ∼24 kDa fragment. The resulting fluorescent protein assembly however is stable and has the same size as its nonsplit variant.[21] Developed 20 years ago, the tetracysteine-containing 6−20 amino acid FlAsH-tag remains the shortest fluorescent tag that can be used to visualize proteins in living cells.[22,23]

**Received:** September 6, 2020
**Published:** November 11, 2020



Downloaded via 65.21.193.86 on February 24, 2022 at 17:07:03 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.


© 2020 American Chemical Society

Amber suppression is an approach for genetic code expansion, converting an amber stop codon (TAG) into an elongation codon for a noncanonical amino acid (ncAA) and allows for specific labeling of a protein of interest with a minimal, bioorthogonal group.[24] Bioorthogonal tags can subsequently be exploited to chemoselectively install an organic fluorophore on the protein of interest.[25,26] Therefore, fluorescent labeling via a site-specifically installed ncAA promises to be a strategy for visualizing and studying microproteins in living cells with minimal perturbation to their function. While selective fluorescent labeling via ncAAs has been demonstrated for a number of proteins, application of this technology to microproteins remains a challenge. Usually, the amber stop codon is inserted at different positions in the coding sequence of a protein of interest and mutants are screened to evaluate the best-possible labeling site.[27,28] In fact, suppression efficiency is highly context-dependent.[29] To ensure that ncAA insertion does not disrupt the functionality of the studied protein and that the ncAA is located in an exposed region, previous knowledge of the protein's functional regions is required. Overall, this approach is highly laborious, time-consuming, and does not fit the purpose of studying uncharacterized microproteins. The recently developed 14 amino acid long GCE-tag, bearing an HA tag followed by a linker and an amber stop codon, has been shown to allow fluorescent tagging of various cellular proteins,[30] but it shares the same size constraints as the tags discussed above.

In order to facilitate the systematic characterization of putative microproteins in their cellular environment, we developed STELLA (single-residue terminal labeling): An universal approach to incorporate ncAAs at the amino- (N-) or carboxyl- (C-) terminus of a protein of interest.

## ■ RESULTS

Here, we set out to establish approaches to express fusion proteins that would post-translationally yield the native microproteins with a single N- or C-terminal ncAA in a traceless manner thus requiring no additional amino acid overhangs.

ncAAs cannot be trivially installed at the N-terminus of a coding sequence: Placement of a stop codon following an ATG start codon will shift translation to a downstream initiation site.[31] In order to tracelessly express N-terminal-labeled microproteins, we designed a fusion protein akin to ubiquitin precursor proteins.[32] The coding sequence of ubiquitin was positioned directly upstream of the microprotein sequence (Figure 1A). This placement mimics the setting of the UBA52 and UBA80 genes, from which endogenous ubiquitin is transcribed and translated as a precursor fused to the amino terminus of two ribosomal proteins, RPL40 and RPS27A.[33,34] After translation, ubiquitin is tracelessly cleaved from fusion proteins by the action of cellular deubiquitinases.[35−37] The first codon of the microprotein was then mutated to an amber stop codon. Thus, we expected that a microprotein bearing the ncAA at its N-terminus would be post-translationally released from a ubiquitin-microprotein intermediate.

To test our design, we generated a series of plasmids expressing GFP with and without a ubiquitin leader. An amber codon was either placed N-terminally or at position 150 of the GFP coding sequence (Figure 1B). The constructs were transiently expressed for 24 h in HEK293T cells that were engineered to stably express the AF mutant of the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (PylRS)

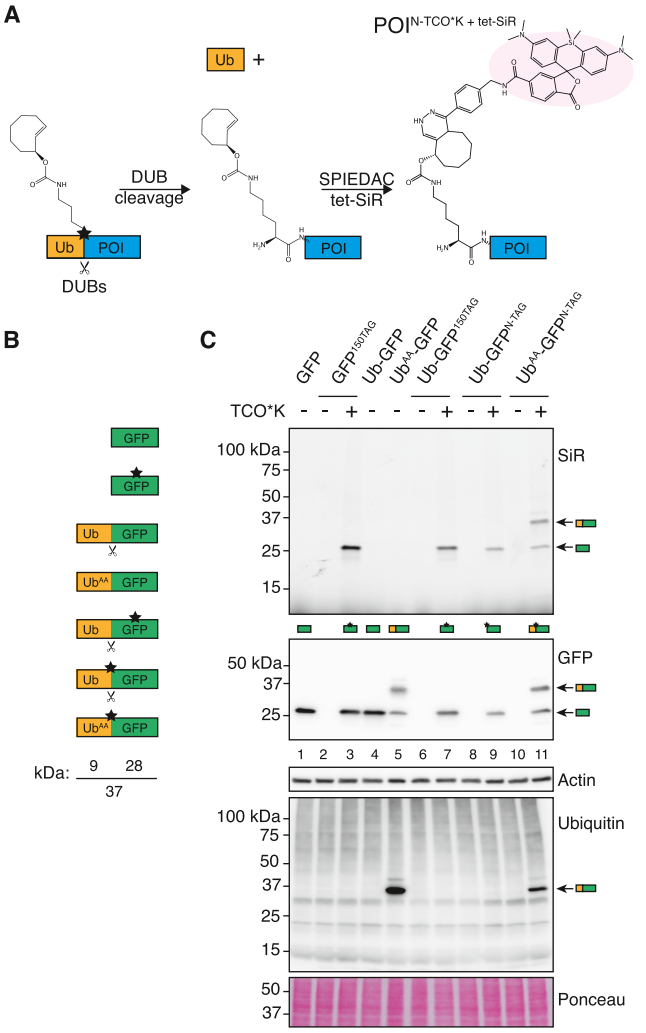



**Figure 1.** N-terminal single residue labeling (STELLA) tag. (A) Schema of the strategy adopted to incorporate noncanonical amino acids (ncAAs, here TCO*K) at the amino-terminus of a protein of interest (POI), in particular, a microprotein. Proteins of interest, whose start codon is replaced by an amber stop codon, are cloned downstream of ubiquitin (Ub). Cellular deubiquitinases cleave Ub, generating the N-terminally tagged POI, which can be fluorescently labeled using a tetrazine dye (here tet-SiR). (B) Summary of the constructs tested in panel C using GFP as a proof-of-principle POI. The position of the ncAA is indicated as star. (C) The different constructs presented in panel B were transiently expressed in HEK293T$^{PylRS\text{-}AF}$ in the presence of 50 $\mu$M TCO*K. 24 h after transfection, cells were lysed in RIPA buffer. Thereafter, lysates were labeled with tet-SiR and analyzed by WB.

along with 4 copies of tRNA$^{Pyl}_{CUA}$ (PylT) (hereafter HEK293T$^{PylRS\text{-}AF}$, see SI Figure 1A for characterization).[27,38,39] We incorporated axial trans-cyclooct-2-ene-L-lysine (TCO*K) with PylRS-AF and used well-established strain-promoted inverse electron-demand Diels−Alder cycloaddition (SPIEDAC) reactions to couple tetrazine-Silicon Rhodamine (tet-SiR) or 6-Methyl-Tetrazine-BODIPY-FL (me-tet-BDP-FL) fluorescent dyes (SI Figure 2).

As hypothesized, expression of a ubiquitin-GFP fusion (Ub-GFP) resulted in a single GFP band slightly above 25 kDa, corresponding to the size of GFP (Figure 1C, compare lane 1 to lane 4). Additional ubiquitin, freed from the ectopically expressed ubiquitin-GFP fusion, did not alter the cellular pool


https://dx.doi.org/10.1021/jacs.0c09574
J. Am. Chem. Soc. 2020, 142, 20080−20087

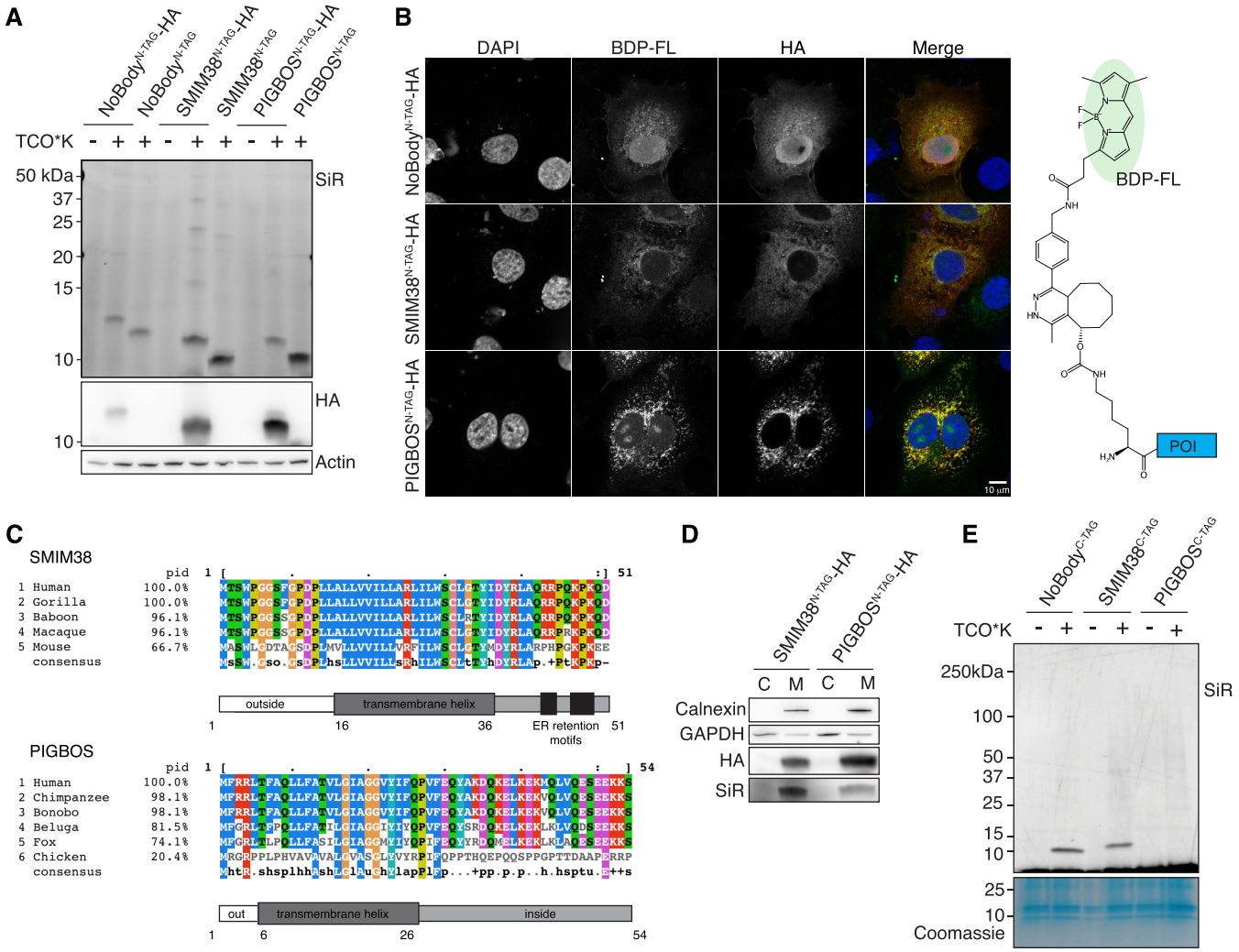



**Figure 2.** N-terminal single-residue labeling of microproteins. (A) Indicated constructs were expressed in HEK293T$^{PylRS\text{-}AF}$ cells. 48 h after transfection, cells were lysed in RIPA buffer, labeled using tet-SIR, and analyzed by WB. Minimally modified microproteins and HA-tagged microproteins were run side-by-side. (B) The microproteins presented in panel A were analyzed by confocal microscopy to assess their subcellular localization. Microproteins were transfected in COS-7$^{PylRS\text{-}AF}$ cells, and the medium was simultaneously supplemented with 50 $\mu$M TCO*K. Cells were fixed 24 h post-transfection. Colocalization analysis and negative controls are shown in SI Figures 9 and 10. (C) Multiple sequence alignments and sketches of SMIM38 and PIGBOS microproteins. Both microproteins are predicted to contain a single-pass transmembrane domain. Putative ER-retention motifs are present at the C-terminus of SMIM38. (D) The microproteins SMIM38 and PIGBOS were expressed in HEK293T$^{PylRS\text{-}AF}$ cells. After 48 h, cells were fractionated to separate membrane-associated proteins from cytoplasmic proteins and analyzed by WB. (E) The panel of microproteins presented in panel A was expressed in HEK293T$^{PylRS\text{-}AF}$ cells using the intein-based system. 48 h after transfection, cells were lysed in RIPA buffer, labeled using tet-SIR and analyzed by SDS-PAGE.

of ubiquitin or noticeably increase overall substrate ubiquitylation (Figure 1C). To evaluate the efficiency of ubiquitin processing, we mutated the two ultimate residues (GG → AA) of ubiquitin, inhibiting cleavage by deubiquitinases.[40] Indeed, Ub$^{G75,76A}$-GFP was processed inefficiently, resulting in an additional higher band corresponding to a ubiquitin-GFP fusion product (Figure 1C, lane 5).

Next, we compared suppression and labeling efficiencies of the N-terminal versus an internal amber codon. Without TCO*K, no GFP was produced with either construct. In the presence of TCO*K, the N-terminal stop codon in GFP$^{N\text{-}TAG}$ was suppressed at a somewhat lower level compared to the internal GFP$^{150TAG}$. Both precursors, Ub-GFP$^{N\text{-}TCO*K}$ and Ub-GFP$^{150TCO*K}$, were quantitatively processed, yielding GFP$^{N\text{-}TCO*K}$ and GFP$^{150TCO*K}$, respectively (Figure 1C, compare lanes 7 and 9). Native and tandem mass spectrometry (SI Figure 3A−C) further confirmed the product GFP$^{N\text{-}TCO*K}$ corresponded to GFP carrying a N-terminal 239 Da TCO*K moiety. Recording GFP fluorescence in live cells over 24 h, we observed similar expression kinetics across different ncAA concentrations, suggesting no major difference in ncAA incorporation efficiencies and similar stabilities of GFP$^{N\text{-}TCO*K}$ and GFP$^{150TCO*K}$ (SI Figure 4). Thus, in agreement with the observation made in *E. coli*, that most bulky ncAAs behave like other stabilizing residues according to the N-end rule,[41] an N-terminal TCO*K did not alter GFP stability in cells.

In order to label the C-terminus of microproteins, an amber codon could in principle be placed as the penultimate codon before the endogenous stop codon. However, suppression efficiency in such a configuration is expected to be low.[42] Thus, we designed a C-terminal fusion construct with a self-cleaving intein that would embed the C-terminal amber codon


https://dx.doi.org/10.1021/jacs.0c09574
J. Am. Chem. Soc. 2020, 142, 20080−20087

approximately in the middle of the ORF while subsequently producing a C-terminally tagged microprotein (SI Figure 5A). Inteins are autocatalytic splicing elements that self-excise from a protein precursor and have been widely employed for mediating precise protein processing.[43] In our construct, we adopted the Ssp DnaE mini-intein N159A mutant of the dual-intein-based polyprotein vector system engineered by Zhang and colleagues.[44] The mini-intein domain mediates cis-splicing at a leading cysteine, cleaving the peptide bond with the preceding residue. Processing at the scissile peptide bond has been shown to be compatible with any preceding amino acid, including ncAAs.[43,45,46] Thus, we sought to express the microprotein sequence including a C-terminal amber codon followed by Ssp DnaE. When the amber codon is suppressed, the microprotein-Ssp DnaE fusion is produced as an intermediate product, from which a microprotein bearing a C-terminal ncAA is released (SI Figure 5A). To ensure that no unspliced product remained in the cell, a FKBP12 degradation domain was introduced after the Ssp DnaE.[47,48]

Using GFP as proof-of-principle, we tested a number of fusion constructs including controls in which the intein's autocatalytic activity was disrupted by mutating the N-terminal catalytic cysteine of the Ssp DnaE domain (SI Figure 5B,C).[44] The intein system enabled us to incorporate a ncAA at the C-terminus of GFP, but a limitation of C-terminal labeling remains that a fraction of the construct is always terminated at the amber codon.

Next, we applied N- and C-terminal STELLA labeling to a representative set of microproteins. For this purpose, we selected the previously characterized microprotein NoBody[49] and two putative microproteins which were discovered using a dedicated mass spectrometry pipeline[8] and had not been characterized at the time. NoBody is a 68 amino acid long microprotein, involved in mRNA turnover and nonsense-mediated decay.[49] The second microprotein is composed of 52 amino acids and was dubbed Small integral membrane protein 38 (SMIM38), because of a predicted transmembrane helix. SMIM38 appears to be conserved across mammals (Figure 2C). No experimental validation of the membrane localization or further characterization exists for SMIM38 to date. The third microprotein comprises 54 amino acids and is located on the opposite strand of the GPI mannosyltransferase 3 (also known as Phosphatidylinositol-glycan biosynthesis class B protein, PIGB). Due to its genomic location, this microprotein was named PIGBOS1/PIGBOS (PIGB opposite strand). PIGBOS is well conserved among tetrapods (Figure 2C) and was recently characterized in detail.[18] The microprotein is inserted in the mitochondrial outer membrane and interacts with the ER protein CLCC1 at ER−mitochondria contact sites. Loss of PIGBOS was shown to result in increased levels of unfolded protein response, culminating in cell death.[18] In order to tag these microproteins, we first turned to N-terminal labeling. The three microproteins, including a version carrying a C-terminal HA tag for each, were cloned as ubiquitin fusions with an amber codon replacing the native start codon. The constructs were transiently expressed for 48 h in HEK293T$^{PylRS\text{-}AF}$ cells in presence or absence of TCO*K. Ub processing was efficient and TCO*K was incorporated in the microproteins (Figure 2A and SI Figure 6), enabling their fluorescent labeling with tet-SiR. Importantly, for all three microproteins, selective labeling was well above the background signal stemming from nonspecific labeling of endogenous proteins carrying an amber stop codon and/or side-reactions of the tetrazine dye. Moreover, when an HA tag was included, microproteins were also detected at the corresponding size using an HA-specific antibody (Figure 2A). We confirmed that TCO*K was correctly incorporated as a single residue at the N-terminus of NoBody by performing a tryptic digest and tandem mass spectrometry on purified NoBody$^{N\text{-}TCO^*K}$-HA (SI Figure 3D).

Encouraged by these results, we moved on to validating STELLA labeling by microscopy. We engineered COS-7 cells to express PylRS-AF and four copies of PylT (COS-7$^{PylRS\text{-}AF}$), since integrating PylRS-AF in the genome allowed for more robust amber suppression than cotransfecting PylRS and the gene of interest in the parental COS-7 cell line (SI Figure 1). After confirming that STELLA labeling strategies were effective across different mammalian cell lines (SI Figure 7), we expressed the above-described microproteins in COS-7$^{PylRS\text{-}AF}$ cells and compared SPIEDAC labeling with immunofluorescent staining of the HA tag (Figure 2B). Following transfection, cells were incubated with 50 $\mu$M TCO*K for 24 h. After fixation, cells were labeled with 6-Methyl-Tetrazine-BODIPY-FL (me-tet-BDP-FL) and costained using an antibody against the HA tag. We chose me-tet-BDP-FL over tet-SiR for imaging experiments, because the latter exhibited strong nonspecific nuclear staining in our hands (SI Figure 8). Confocal microscopy of NoBody and PIGBOS was in agreement with their known localization in the cytoplasm[49] and mitochondria[18] respectively. The specificity of bioorthogonal labeling was further confirmed by a series of control experiments (SI Figure 9A,B).

By performing an analogous transfection and labeling in HEK293T$^{PylRS\text{-}AF}$, we confirmed that the microproteins exhibited the same localization across different cell lines (SI Figure 9C). Colocalization analysis confirmed that BDP-FL and HA signals broadly colocalize (Pearson correlations 0.68−0.90; Mander's co-occurrence coefficients 0.62−0.86), agreeing on specific subcellular localizations (Figure 2B, SI Figure 10). Imperfect colocalization, e.g., more nuclear signal for BDP-FL, may reflect unspecific background[50] but also shortcomings of antibody staining. For example, some systematic discrepancies between antibody staining and the underlying protein localization as determined by a fluorescent protein tag have been described before.[51]

Analysis of the amino acid sequences of SMIM38 and PIGBOS using TOPCONS predicted that both microproteins possess a central transmembrane helical region.[52] The transmembrane region of SMIM38 spans amino acids 16 to 36, followed by predicted ER retention motifs (Figure 2C).[53] Thus, we sought to experimentally test if SMIM38$^{N\text{-}TCO^*K}$-HA and PIGBOS$^{N\text{-}TCO^*K}$-HA indeed insert into membranes. 48 h after transfection, we fractionated cells to separate membrane-inserted proteins from cytoplasmic proteins. Resulting fractions were then analyzed by in-gel fluorescence and western blotting. Both microproteins were exclusively found in the membrane fraction as was the positive control, the ER-associated transmembrane protein calnexin (Figure 2D). Thus, N-terminal labeling did not appear to interfere with membrane insertion of either microprotein.

While N-terminal labeling performed well on microproteins, we also tested our C-terminal labeling strategy for NoBody, SMIM38, and PIGBOS. The microproteins were cloned into the intein-based vector and transiently expressed in HEK293T$^{PylRS\text{-}AF}$ cells in the presence or absence of


https://dx.doi.org/10.1021/jacs.0c09574
J. Am. Chem. Soc. 2020, 142, 20080−20087

TCO*K. After 48 h, cell lysates were prepared, reacted with tet-SiR, and analyzed by SDS-PAGE (Figure 2E and SI Figure 6). While we did not achieve C-terminal labeling of PIGBOS$^{C\text{-}TAG}$, both NoBody$^{C\text{-}TAG}$ and SMIM38$^{C\text{-}TAG}$ yielded good and selective labeling, indicating that this method could be useful for the study of particular microproteins or proteins that cannot be labeled on the N-terminus, e.g., those that bear amino-terminal signal peptides or are processed in other ways at the N-terminus.

Having established the STELLA tags, we sought to test the capability of labeling microproteins or small proteins in different subcellular compartments (Figure 3). We chose some of the smallest proteins annotated in UniProt[54] with known distinctive localization. Caveolin-3 is a 151 amino acid protein localized to the plasma membrane and Golgi. CENP-A is a 140 amino acid histone variant defining centromeres. VAMP8 is a 100 amino acid vesicle-associated membrane protein. ATP5F1E is the smallest core subunit of mitochondrial ATP Synthase with just 50 amino acids.[55] We also included the 222 amino acid SARS-CoV-2 Membrane (M) protein for its known exclusive localization to the Golgi apparatus.[56] For all these proteins, N-terminal fluorescent labeling reproduced the expected subcellular patterns, demonstrating that ncAA incorporation and labeling is universally applicable (Figure 3A). Lysate labeling further confirmed the specificity of the labeling (Figure 3B).

Since both me-tet-BDP-FL and tet-SiR are cell permeable and SPIEDAC is a biocompatible and nontoxic reaction,[57] we next tested whether our labeling scheme was amenable for visualizing microproteins in living cells. We first confirmed that SPIEDAC labeling on microproteins proceeds as efficiently and selective in living cells as demonstrated in the lysates above. Indeed, me-tet-BDP-FL or tet-SiR incubation of live cells selectively labeled the ncAA-bearing microproteins as evident from in-gel fluorescence assessed after SDS-PAGE (SI Figure 11 A,B). Further, we carried out live cell imaging experiments in COS-7$^{PylRS\text{-}AF}$ cells (Figure 4). Transfected cells were grown in the presence of TCO*K for 24 h. Subsequently, the free ncAA was washed out and cells were grown for 1 h in ncAA-free medium. ncAA-bearing target microproteins were labeled using me-tet-BDP-FL and simultaneously cells were counterstained with either Hoechst, ER-tracker, or MitoTracker and immediately imaged by widefield microscopy.

Colocalization with MitoTracker confirmed the previously described localization of PIGBOS to mitochondria.[18] Co-staining with ER-Tracker revealed that SMIM38 is mainly localized to the ER (Figure 4; see also Figure 2B). To further confirm the universality of our labeling approach, we included two additional transmembrane-spanning microproteins with known subcellular localization. Sarcolipin is a 31 amino acid transmembrane microprotein localized in the ER of muscle cells,[58] whereas NDUFB1 is a 57 amino acid subunit of the mitochondrial complex I (NADH:ubiquinone oxidoreductase) resident in the inner mitochondrial membrane.[59] As expected, Sarcolipin$^{N\text{-}TCO*K}$ and NDUFB1$^{N\text{-}TCO*K}$ colocalized with ER-Tracker and MitoTracker, respectively, in live cells (Figure 4). Our N-terminal labeling approach also visualized CENP-A faithfully in the nucleus of live cells akin to our result in fixed cells (SI Figure 11C and Figure 3, respectively). Additional controls showed negligible background fluorescence for live cell microscopy experiments in the absence of TCO*K or me-tet-BDP-FL (SI Figure 11 C). Importantly, live cell imaging of

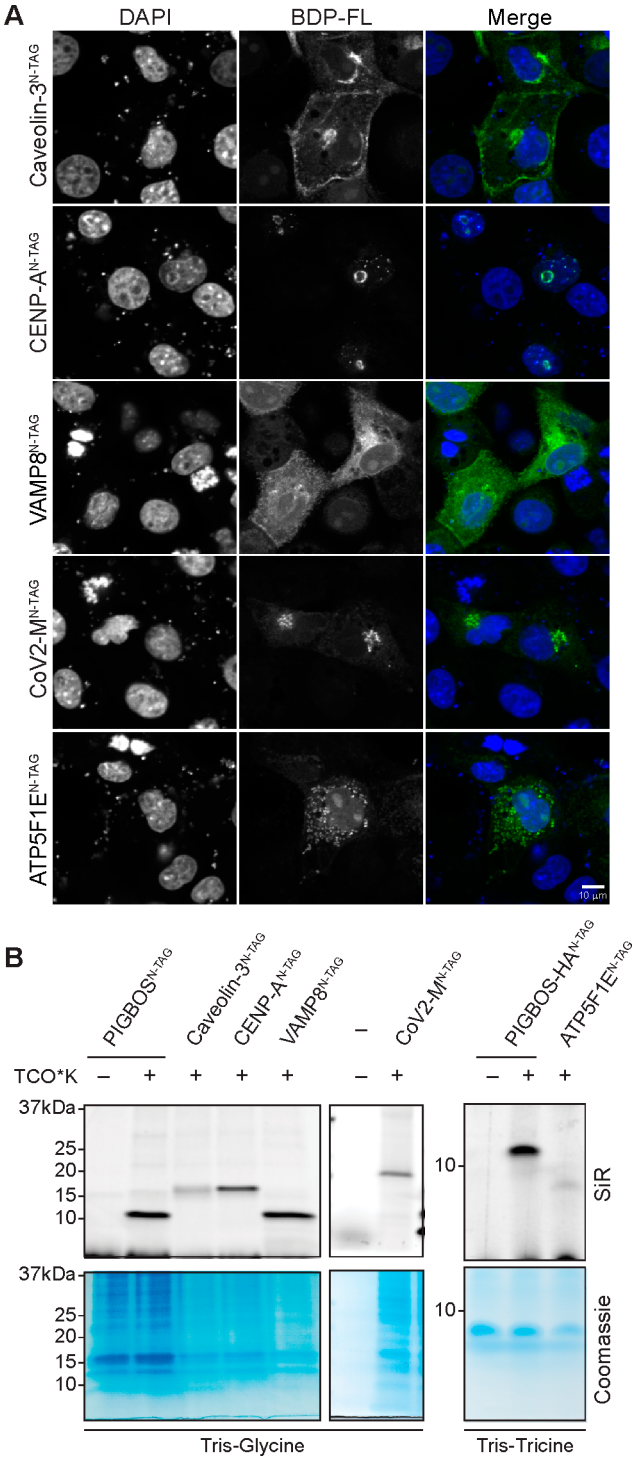



**Figure 3.** N-terminal single-residue labeling of a panel of small proteins with known subcellular localization. (A) Confocal microscopy images of Caveolin-3, CENP-A, VAMP8, SARS-CoV2-M, and ATP5F1E proteins, N-terminally labeled using STELLA. Constructs were transfected in COS-7$^{PylRS\text{-}AF}$ cells and incubated with 50 $\mu$M TCO*K. Cells were fixed 24 h post-transfection and prepared for imaging. (B) The constructs presented in (A) were expressed in HEK293T$^{PylRS\text{-}AF}$ cells in the presence of 50 $\mu$M TCO*K. 48 h after transfection, cells were lysed in RIPA buffer and labeled using tet-SIR. Proteins were separated by Tris-Glycine (left) or Tris-Tricine SDS-PAGE (right) and imaged in-gel.


https://dx.doi.org/10.1021/jacs.0c09574
J. Am. Chem. Soc. 2020, 142, 20080−20087

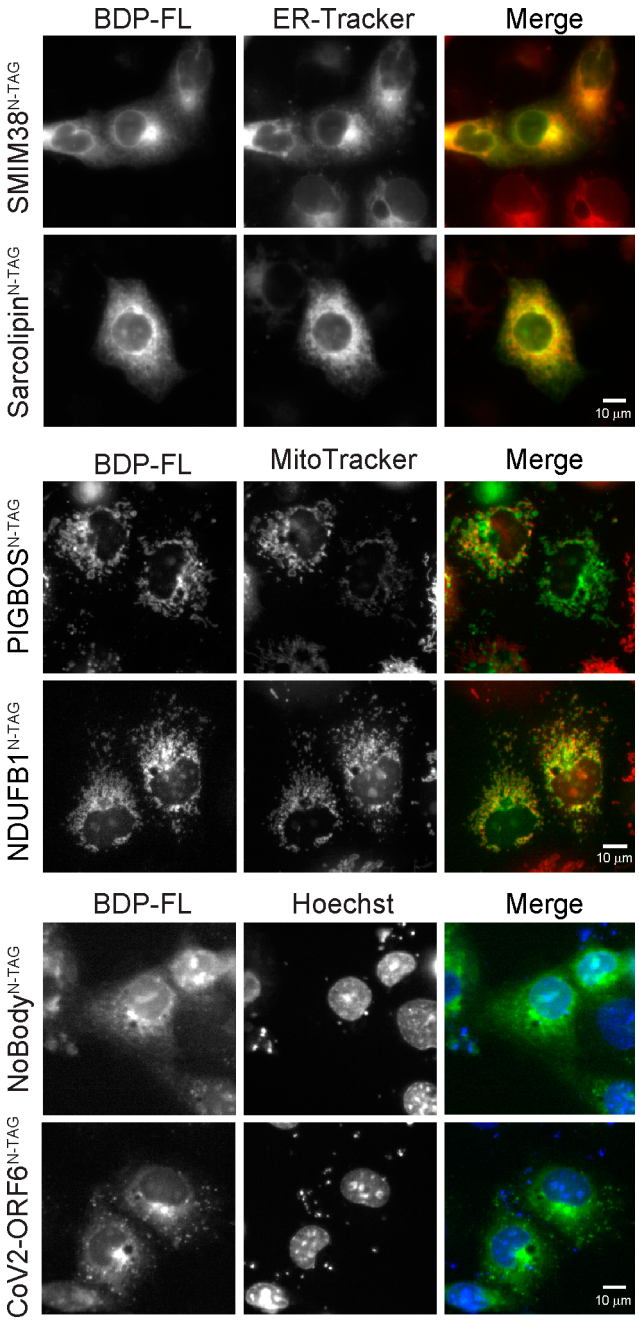



**Figure 4.** Live cell imaging of microproteins. Subcellular localization of SMIM38, Sarcolipin, PIGBOS, NDUFB1, NoBody, and SARS-CoV-2 ORF6 was assessed by widefield live cell microscopy. 24 h post-transfection, COS-7$^{PylRS-AF}$ cells were exposed to 500 nM me-tet-BDP-FL in combination with either ER-tracker, Mitotracker, or Hoechst for 30 min. Cells were imaged directly after withdrawal of the dyes. Gels showing selective labeling for the same proteins are shown in SI Figures 11 and 12.

minimally tagged, native microproteins allows to study their cellular dynamics, as exemplified by time lapse movies showing NDUFB1 moving with mitochondria (SI Movie 1 and SI Figure 13). In summary, N-terminal STELLA labeling enabled us to visualize different microproteins in living cells without perturbing their physiological localization.

Beyond cellular microproteins, viruses encode an unproportionally large number of alternative and short ORFs in their compact genome.[60] The novel coronavirus SARS-CoV-2 encodes a number of conserved sORFs of which ORF3b, ORF7b, ORF9b, ORF9c are alternative reading frames in a larger ORF. Evidence exists that some of these are translated and interact with cellular proteins.[61] We expressed these sORFs using our N-terminal STELLA tag and were able to label and visualize all except ORF10 (SI Figure 12), the translation of which remained elusive also in a previous study.[61] Notably, all expressed sORFs, including the only 22 amino acid ORF3b,[62] appeared to associate with various membranous compartments in the cytoplasm (SI Figure 12A). SARS-CoV-2 ORF6 has been suggested to suppress the cellular antiviral response by sequestering nuclear import factors to the ER.[63] Intriguingly, using live cell imaging we found SARS-CoV-2 ORF6$^{N-TCO*K}$ to feature fast moving speckles, suggesting that ORF6 associates with trafficking vesicles of the host cell. This illustrates the feasibility of studying dynamic processes with STELLA (Figure 4, SI Movie 2, SI Figure 13).

In conclusion, we show that STELLA tags have a unique utility in studying microproteins. We demonstrate the selective labeling and visualization of microproteins, as short as 22 amino acids, by SDS-PAGE and microscopy. Our labeling strategy is scalable and compatible with live cell experiments, enabling the study and characterization of a large number of microproteins and providing universal and predictable means to label larger proteins as well.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.0c09574.

Materials and Methods and uncropped gel images (PDF)

Movie relating to Figure 4 (AVI)

Movie relating to Figure 4 (AVI)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Simon J. Elsässer** − *Science for Life Laboratory, Department of Medical Biochemistry and Biophysics, Division of Genome Biology and Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm 17165, Sweden;* orcid.org/0000-0001-8724-4849; Email: simon.elsasser@scilifelab.se

### Authors

**Lorenzo Lafranchi** − *Science for Life Laboratory, Department of Medical Biochemistry and Biophysics, Division of Genome Biology and Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm 17165, Sweden*

**Dörte Schlesinger** − *Science for Life Laboratory, Department of Medical Biochemistry and Biophysics, Division of Genome Biology and Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm 17165, Sweden*

**Kyle J. Kimler** − *Science for Life Laboratory, Department of Medical Biochemistry and Biophysics, Division of Genome Biology and Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm 17165, Sweden*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.0c09574

### Author Contributions

All authors have given approval to the final version of the manuscript.

### Funding

Research was funded by Karolinska Institutet SFO Molecular Biosciences, Ming Wai Lau Center for Reparative Medicine, Ragnar Söderbergs Stiftelse, Stiftelsen för Strategisk Forskning (S.J.E.) and Boehringer Ingelheim Fonds (D.S.)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Rozina Caridha, Birthe Meineke, and Johannes Heimgärtner for experimental assistance. We thank Michael Landreh for intact mass determination and the Karolinska Institutet Proteomics Biomedicum facility for MS/MS service. We thank Johannes Heimgärtner for critical reading of the manuscript and all members of the Elsässer lab for input into the concept, experimental design, and manuscript. We thank the other groups in the Division of Genome Biology for their support, specifically the J. Bartek lab for access to the Nikon Eclipse Ti2 Microscope, and the O. Fernandez-Capetillo lab for access to the GE AI600 Gel Imager.

## ■ ABBREVIATIONS

STELLA, single-residue terminal labeling; TCO*K, axial trans-cyclooct-2-ene-L-lysine; SPIEDAC, strain-promoted inverse electron-demand Diels–Alder cycloaddition; tet-SiR, tetrazine-Silicon Rhodamine; me-tet-BDP-FL, 6-Methyl-Tetrazine-BODIPY-FL; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; GFP, Green Fluorescent Protein; ER, Endoplasmic Reticulum; ORF, open reading frame

## ■ REFERENCES

(1) Couso, J.-P.; Patraquim, P. Classification and Function of Small Open Reading Frames. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 575–589.

(2) Plaza, S.; Menschaert, G.; Payre, F. In Search of Lost Small Peptides. *Annu. Rev. Cell Dev. Biol.* **2017**, *33*, 391–416.

(3) Makarewich, C. A.; Olson, E. N. Mining for Micropeptides. *Trends Cell Biol.* **2017**, *27*, 685–696.

(4) Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63.

(5) Brar, G. A.; Weissman, J. S. Ribosome Profiling Reveals the What, When, Where and How of Protein Synthesis. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 651–664.

(6) Slavoff, S. A.; Mitchell, A. J.; Schwaid, A. G.; Cabili, M. N.; Ma, J.; Levin, J. Z.; Karger, A. D.; Budnik, B. A.; Rinn, J. L.; Saghatelian, A. Peptidomic Discovery of Short Open Reading Frame-Encoded Peptides in Human Cells. *Nat. Chem. Biol.* **2013**, *9*, 59–64.

(7) Ma, J.; Ward, C. C.; Jungreis, I.; Slavoff, S. A.; Schwaid, A. G.; Neveu, J.; Budnik, B. A.; Kellis, M.; Saghatelian, A. Discovery of Human SORF-Encoded Polypeptides (SEPs) in Cell Lines and Tissue. *J. Proteome Res.* **2014**, *13*, 1757–1765.

(8) Ma, J.; Diedrich, J. K.; Jungreis, I.; Donaldson, C.; Vaughan, J.; Kellis, M.; Yates, J. R.; Saghatelian, A. Improved Identification and Analysis of Small Open Reading Frame Encoded Polypeptides. *Anal. Chem.* **2016**, *88*, 3967–3975.

(9) Vanderperre, B.; Lucier, J.-F.; Bissonnette, C.; Motard, J.; Tremblay, G.; Vanderperre, S.; Wisztorski, M.; Salzet, M.; Boisvert, F.-M.; Roucou, X. Direct Detection of Alternative Open Reading Frames Translation Products in Human Significantly Expands the Proteome. *PLoS One* **2013**, *8*, e70698.

(10) Samandi, S.; Roy, A. V.; Delcourt, V.; Lucier, J.-F.; Gagnon, J.; Beaudoin, M. C.; Vanderperre, B.; Breton, M.-A.; Motard, J.; Jacques, J.-F.; Brunelle, M.; Gagnon-Arsenault, I.; Fournier, I.; Ouangraoua, A.; Hunting, D. J.; Cohen, A. A.; Landry, C. R.; Scott, M. S.; Roucou, X. Deep Transcriptome Annotation Enables the Discovery and Functional Characterization of Cryptic Small Proteins. *eLife* **2017**, *6*, e27860.

(11) Lee, C.; Zeng, J.; Drew, B. G.; Sallam, T.; Martin-Montalvo, A.; Wan, J.; Kim, S.-J.; Mehta, H.; Hevener, A. L.; de Cabo, R.; Cohen, P. The Mitochondrial-Derived Peptide MOTS-c Promotes Metabolic Homeostasis and Reduces Obesity and Insulin Resistance. *Cell Metab.* **2015**, *21*, 443–454.

(12) Denli, A. M.; Narvaiza, I.; Kerman, B. E.; Pena, M.; Benner, C.; Marchetto, M. C. N.; Diedrich, J. K.; Aslanian, A.; Ma, J.; Moresco, J. J.; Moore, L.; Hunter, T.; Saghatelian, A.; Gage, F. H. Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* **2015**, *163*, 583–593.

(13) Bi, P.; Ramirez-Martinez, A.; Li, H.; Cannavino, J.; McAnally, J. R.; Shelton, J. M.; Sánchez-Ortiz, E.; Bassel-Duby, R.; Olson, E. N. Control of Muscle Formation by the Fusogenic Micropeptide Myomixer. *Science* **2017**, *356*, 323–327.

(14) Zhang, Q.; Vashisht, A. A.; O'Rourke, J.; Corbel, S. Y.; Moran, R.; Romero, A.; Miraglia, L.; Zhang, J.; Durrant, E.; Schmedt, C.; Sampath, S. C.; Sampath, S. C. The Microprotein Minion Controls Cell Fusion and Muscle Formation. *Nat. Commun.* **2017**, *8*, 15664.

(15) Quinn, M. E.; Goh, Q.; Kurosaka, M.; Gamage, D. G.; Petrany, M. J.; Prasad, V.; Millay, D. P. Myomerger Induces Fusion of Non-Fusogenic Cells and Is Required for Skeletal Muscle Development. *Nat. Commun.* **2017**, *8*, 15665.

(16) Lightfoot, J. W.; Wilecki, M.; Rödelsperger, C.; Moreno, E.; Susoy, V.; Witte, H.; Sommer, R. J. Small Peptide-Mediated Self-Recognition Prevents Cannibalism in Predatory Nematodes. *Science* **2019**, *364*, 86–89.

(17) Pauli, A.; Norris, M. L.; Valen, E.; Chew, G.-L.; Gagnon, J. A.; Zimmerman, S.; Mitchell, A.; Ma, J.; Dubrulle, J.; Reyon, D.; Tsai, S. Q.; Joung, J. K.; Saghatelian, A.; Schier, A. F. Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* **2014**, *343*, 1248636.

(18) Chu, Q.; Martinez, T. F.; Novak, S. W.; Donaldson, C. J.; Tan, D.; Vaughan, J. M.; Chang, T.; Diedrich, J. K.; Andrade, L.; Kim, A.; Zhang, T.; Manor, U.; Saghatelian, A. Regulation of the ER Stress Response by a Mitochondrial Microprotein. *Nat. Commun.* **2019**, *10*, 4883.

(19) Mollwitz, B.; Brunk, E.; Schmitt, S.; Pojer, F.; Bannwarth, M.; Schiltz, M.; Rothlisberger, U.; Johnsson, K. Directed Evolution of the Suicide Protein O6-Alkylguanine-DNA Alkyltransferase for Increased Reactivity Results in an Alkylated Protein with Exceptional Stability. *Biochemistry* **2012**, *51*, 986–994.

(20) Kamiyama, D.; Sekine, S.; Barsi-Rhyne, B.; Hu, J.; Chen, B.; Gilbert, L. A.; Ishikawa, H.; Leonetti, M. D.; Marshall, W. F.; Weissman, J. S.; Huang, B. Versatile Protein Tagging in Cells with Split Fluorescent Protein. *Nat. Commun.* **2016**, *7*, 11046.

(21) Cabantous, S.; Terwilliger, T. C.; Waldo, G. S. Protein Tagging and Detection with Engineered Self-Assembling Fragments of Green Fluorescent Protein. *Nat. Biotechnol.* **2005**, *23*, 102–107.

(22) Hoffmann, C.; Gaietta, G.; Zürn, A.; Adams, S. R.; Terrillon, S.; Ellisman, M. H.; Tsien, R. Y.; Lohse, M. J. Fluorescent Labeling of Tetracysteine-Tagged Proteins in Intact Cells. *Nat. Protoc.* **2010**, *5*, 1666–1677.

(23) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific Covalent Labeling of Recombinant Protein Molecules inside Live Cells. *Science* **1998**, *281*, 269–272.

(24) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550*, 53–60.

(25) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

(26) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974–6998.

(27) Nikić, I.; Plass, T.; Schraidt, O.; Szymański, J.; Briggs, J. A. G.; Schultz, C.; Lemke, E. A. Minimal Tags for Rapid Dual-Color Live-

Cell Labeling and Super-Resolution Microscopy. *Angew. Chem., Int. Ed.* **2014**, *53*, 2245−2249.
(28) Peng, T.; Hang, H. C. Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* **2016**, *138*, 14423−14433.
(29) Chemla, Y.; Ozer, E.; Algov, I.; Alfonta, L. Context Effects of Genetic Code Expansion by Stop Codon Suppression. *Curr. Opin. Chem. Biol.* **2018**, *46*, 146−155.
(30) Segal, I.; Nachmias, D.; Konig, A.; Alon, A.; Arbely, E.; Elia, N. A Straightforward Approach for Bioorthogonal Labeling of Proteins and Organelles in Live Mammalian Cells. *BMC Biol.* **2020**, *18*, 5.
(31) Kalstrup, T.; Blunck, R. Reinitiation at Non-Canonical Start Codons Leads to Leak Expression When Incorporating Unnatural Amino Acids. *Sci. Rep.* **2015**, *5*, 11866.
(32) Komander, D.; Clague, M. J.; Urbé, S. Breaking the Chains: Structure and Function of the Deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 550−563.
(33) Baker, R. T.; Board, P. G. The Human Ubiquitin-52 Amino Acid Fusion Protein Gene Shares Several Structural Features with Mammalian Ribosomal Protein Genes. *Nucleic Acids Res.* **1991**, *19*, 1035−1040.
(34) Redman, K. L.; Rechsteiner, M. Identification of the Long Ubiquitin Extension as Ribosomal Protein S27a. *Nature* **1989**, *338*, 438−440.
(35) Larsen, C. N.; Krantz, B. A.; Wilkinson, K. D. Substrate Specificity of Deubiquitinating Enzymes: Ubiquitin C-Terminal Hydrolases. *Biochemistry* **1998**, *37*, 3358−3368.
(36) Grou, C. P.; Pinto, M. P.; Mendes, A. V.; Domingues, P.; Azevedo, J. E. The de Novo Synthesis of Ubiquitin: Identification of Deubiquitinases Acting on Ubiquitin Precursors. *Sci. Rep.* **2015**, *5*, 12836.
(37) Matentzoglu, K.; Scheffner, M. Ubiquitin-Fusion Protein System: A Powerful Tool for Ectopic Protein Expression in Mammalian Cells. *BioTechniques* **2009**, *46*, 21−22.
(38) Elsässer, S. J. Generation of Stable Amber Suppression Cell Lines. *Methods Mol. Biol.* **2018**, *1728*, 237−245.
(39) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep Engineering of Pyrrolysyl-TRNA Synthetase to Genetically Encode N(Epsilon)-(o-Azidobenzyloxycarbonyl) Lysine for Site-Specific Protein Modification. *Chem. Biol.* **2008**, *15*, 1187−1197.
(40) Han, X. J.; Lee, M. J.; Yu, G. R.; Lee, Z. W.; Bae, J. Y.; Bae, Y. C.; Kang, S. H.; Kim, D. G. Altered Dynamics of Ubiquitin Hybrid Proteins during Tumor Cell Apoptosis. *Cell Death Dis.* **2012**, *3*, e255.
(41) Kunjapur, A. M.; Stork, D. A.; Kuru, E.; Vargas-Rodriguez, O.; Landon, M.; Söll, D.; Church, G. M. Engineering Posttranslational Proofreading to Discriminate Nonstandard Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 619−624.
(42) Brar, G. A. Beyond the Triplet Code: Context Cues Transform Translation. *Cell* **2016**, *167*, 1681−1692.
(43) Vila-Perelló, M.; Muir, T. W. Biological Applications of Protein Splicing. *Cell* **2010**, *143*, 191−200.
(44) Zhang, B.; Rapolu, M.; Liang, Z.; Han, Z.; Williams, P. G.; Su, W. W. A Dual-Intein Autoprocessing Domain That Directs Synchronized Protein Co-Expression in Both Prokaryotes and Eukaryotes. *Sci. Rep.* **2015**, *5*, 8541.
(45) Batjargal, S.; Walters, C. R.; Petersson, E. J. Inteins as Traceless Purification Tags for Unnatural Amino Acid Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 1734−1737.
(46) Qianzhu, H.; Welegedara, A. P.; Williamson, H.; McGrath, A. E.; Mahawaththa, M. C.; Dixon, N. E.; Otting, G.; Huber, T. Genetic Encoding of Para-Pentafluorosulfanyl Phenylalanine: A Highly Hydrophobic and Strongly Electronegative Group for Stable Protein Interactions. *J. Am. Chem. Soc.* **2020**, *142*, 17277.
(47) Banaszynski, L. A.; Chen, L.-C.; Maynard-Smith, L. A.; Ooi, A. G. L.; Wandless, T. J. A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell* **2006**, *126*, 995−1004.
(48) Chu, B. W.; Banaszynski, L. A.; Chen, L.; Wandless, T. J. Recent Progress with FKBP-Derived Destabilizing Domains. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5941−5944.
(49) D'Lima, N. G.; Ma, J.; Winkler, L.; Chu, Q.; Loh, K. H.; Corpuz, E. O.; Budnik, B. A.; Lykke-Andersen, J.; Saghatelian, A.; Slavoff, S. A. A Human Microprotein That Interacts with the MRNA Decapping Complex. *Nat. Chem. Biol.* **2017**, *13*, 174−180.
(50) Nikić, I.; Estrada Girona, G.; Kang, J. H.; Paci, G.; Mikhaleva, S.; Koehler, C.; Shymanska, N. V.; Ventura Santos, C.; Spitz, D.; Lemke, E. A. Debugging Eukaryotic Genetic Code Expansion for Site-Specific Click-PAINT Super-Resolution Microscopy. *Angew. Chem., Int. Ed.* **2016**, *55*, 16172−16176.
(51) Stadler, C.; Rexhepaj, E.; Singan, V. R.; Murphy, R. F.; Pepperkok, R.; Uhlén, M.; Simpson, J. C.; Lundberg, E. Immunofluorescence and Fluorescent-Protein Tagging Show High Correlation for Protein Localization in Mammalian Cells. *Nat. Methods* **2013**, *10*, 315−323.
(52) Tsirigos, K. D.; Peters, C.; Shu, N.; Käll, L.; Elofsson, A. The TOPCONS Web Server for Consensus Prediction of Membrane Protein Topology and Signal Peptides. *Nucleic Acids Res.* **2015**, *43*, W401−7.
(53) Gouw, M.; Michael, S.; Sámano-Sánchez, H.; Kumar, M.; Zeke, A.; Lang, B.; Bely, B.; Chemes, L. B.; Davey, N. E.; Deng, Z.; Diella, F.; Gürth, C.-M.; Huber, A.-K.; Kleinsorg, S.; Schlegel, L. S.; Palopoli, N.; Roey, K. V.; Altenberg, B.; Reményi, A.; Dinkel, H.; Gibson, T. J. The Eukaryotic Linear Motif Resource - 2018 Update. *Nucleic Acids Res.* **2018**, *46*, D428−D434.
(54) UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506−D515.
(55) Walker, J. E.; Fearnley, I. M.; Gay, N. J.; Gibson, B. W.; Northrop, F. D.; Powell, S. J.; Runswick, M. J.; Saraste, M.; Tybulewicz, V. L. Primary Structure and Subunit Stoichiometry of F1-ATPase from Bovine Mitochondria. *J. Mol. Biol.* **1985**, *184*, 677−701.
(56) Klumperman, J.; Locker, J. K.; Meijer, A.; Horzinek, M. C.; Geuze, H. J.; Rottier, P. J. Coronavirus M Proteins Accumulate in the Golgi Complex beyond the Site of Virion Budding. *J. Virol.* **1994**, *68*, 6523−6534.
(57) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004−2021.
(58) Odermatt, A.; Becker, S.; Khanna, V. K.; Kurzydlowski, K.; Leisner, E.; Pette, D.; MacLennan, D. H. Sarcolipin Regulates the Activity of SERCA1, the Fast-Twitch Skeletal Muscle Sarcoplasmic Reticulum Ca2+-ATPase. *J. Biol. Chem.* **1998**, *273*, 12360−12369.
(59) Stroud, D. A.; Surgenor, E. E.; Formosa, L. E.; Reljic, B.; Frazier, A. E.; Dibley, M. G.; Osellame, L. D.; Stait, T.; Beilharz, T. H.; Thorburn, D. R.; Salim, A.; Ryan, M. T. Accessory Subunits Are Integral for Assembly and Function of Human Mitochondrial Complex I. *Nature* **2016**, *538*, 123−126.
(60) DiMaio, D. Viral Miniproteins. *Annu. Rev. Microbiol.* **2014**, *68*, 21−43.
(61) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; Tummino, T. A.; Hüttenhain, R.; Kaake, R. M.; Richards, A. L.; Tutuncuoglu, B.; Foussard, H.; Batra, J.; Haas, K.; Modak, M.; Kim, M.; Krogan, N. J. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* **2020**, *583*, 459−468.
(62) Konno, Y.; Kimura, I.; Uriu, K.; Fukushi, M.; Irie, T.; Koyanagi, Y.; Sauter, D.; Gifford, R. J.; USFQ-COVID19 Consortium; Nakagawa, S.; Sato, K. SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Rep.* **2020**, *32*, 108185.
(63) Frieman, M.; Yount, B.; Heise, M.; Kopecky-Bromberg, S. A.; Palese, P.; Baric, R. S. Severe Acute Respiratory Syndrome Coronavirus ORF6 Antagonizes STAT1 Function by Sequestering Nuclear Import Factors on the Rough Endoplasmic Reticulum/Golgi Membrane. *J. Virol.* **2007**, *81*, 9812−9824.


https://dx.doi.org/10.1021/jacs.0c09574
J. Am. Chem. Soc. 2020, 142, 20080−20087