ACS chemical biology

pubs.acs.org/acschemicalbiology

Articles

Photoredox-Catalyzed Decarboxylative *C*-Terminal Differentiation for Bulk- and Single-Molecule Proteomics

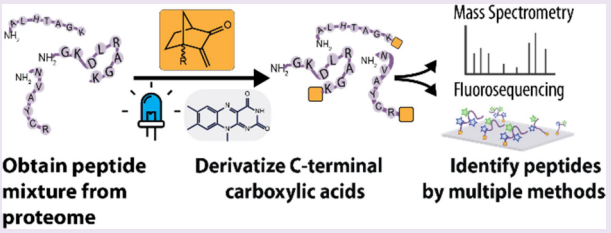
Le Zhang,[⊥] Brendan M. Floyd,[⊥] Maheshwerreddy Chilamari, James Mapes, Jagannath Swaminathan, Steven Bloom,* Edward M. Marcotte,* and Eric V. Anslyn*

Cite This: *ACS Chem. Biol.* 2021, 16, 2595−2603

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Methods for the selective labeling of biogenic functional groups on peptides are being developed and used in the workflow of both current and emerging proteomics technologies, such as single-molecule fluorosequencing. To achieve successful labeling with any one method requires that the peptide fragments contain the functional group for which labeling chemistry is designed. In practice, only two functional groups are present in every peptide fragment regardless of the protein cleavage site, namely, an *N*-terminal amine and a C-terminal carboxylic acid. Developing a global-labeling technology, therefore, requires one to specifically target the *N*- and/or C-terminus of peptides. In this work, we showcase the first successful application of photocatalyzed C-terminal decarboxylative alkylation for peptide mass spectrometry and single-molecule protein sequencing that can be broadly applied to any proteome. We demonstrate that peptides in complex mixtures generated from enzymatic digests from bovine serum albumin, as well as protein mixtures from yeast and human cell extracts, can be site-specifically labeled at their C-terminal residue with a Michael acceptor. Using two distinct analytical approaches, we characterize C-terminal labeling efficiencies of greater than 50% across complete proteomes and document the proclivity of various C-terminal amino-acid residues for decarboxylative labeling, showing histidine and tryptophan to be the most disfavored. Finally, we combine C-terminal decarboxylative labeling with an orthogonal carboxylic acid-labeling technology in tandem to establish a new platform for fluorosequencing.



INTRODUCTION

Progress in protein mass spectrometry (MS), along with the development of single-molecule protein-sequencing technologies, and other advancements in protein detection and quantification have yielded several new opportunities for proteomics.[1,2] Examples include (1) the discovery of novel microproteins,[3] (2) the deployment of single-cell proteomics,[4,5] and (3) the discovery of new diagnostic tools for the clinic.[6] The vast majority of existing applications take advantage of covalent labeling technologies for peptides and proteins, many of which are designed to target specific functional groups found in biogenic amino-acid (AA) residues. While extant conjugation methods have certainly opened the door to new possibilities for protein MS, still more methods are needed to mine this nascent area of research. For MS, the need for innovative cross-linking strategies to study protein−protein interactions and methods to label peptides with isotopic residues for quantitation and multiplexing is at the top of the list.[7,8] The development of single-molecule protein-sequencing methods has also led to a pronounced need for labeling chemistries that can discriminate between specific AA residues found in peptides,[12,13] particularly those that share a common functional group (e.g., the amines of lysine and *N*-terminal residues, and the carboxylic acids of aspartic acid, glutamic acid, and the C-terminal residue).[2] One way to address these topics would be to develop and/or to make use of labeling technologies that specially target the C- or *N*-terminus of peptides. These termini are (1) naturally present in almost every peptide; (2) often removed from the active binding site of the peptide, meaning that labeling these positions would be less likely to interfere with the endogenous activity of the peptide; and (3) generally accessible for modifications, free from any steric or conformational constraints.[9−11]

A number of techniques have been developed for the selective and efficient labeling of the *N*-terminal amino group of peptides and proteins (as opposed to the amine of lysine residues).[14−16] Corresponding methods for attaching a synthetic reporter to the C-terminus of peptides and proteins that also keep alternate side-chain acidic residues (i.e., aspartate and glutamate) pristine are few and seldom well

Received: August 10, 2021
Accepted: October 21, 2021
Published: November 4, 2021

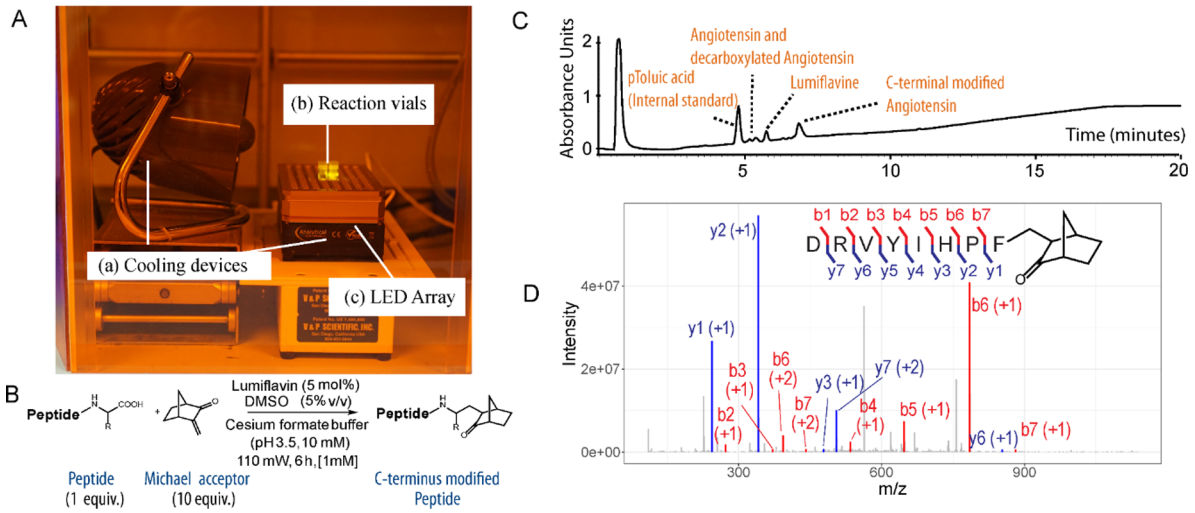


Figure 1. Demonstrating the selective C-terminal carboxylic acid labeling of angiotensin. (A) Reaction setup for performing the photoredox chemistry. The three main components as indicated in the picture are (a) a tabletop fan and an active cooling unit, (b) a glass vial containing the reaction mixture, including the photocatalyst lumiflavin, peptide, and Michael acceptor NB in acidic buffer, and (c) the Lumidox Gen II LED array with an aluminum block (Analytical Sales and Services, NJ, USA). (B) Reaction scheme illustrating the derivatization of the terminal carboxylic acid with NB.(C) UPLC-chromatogram of the reaction products. (D) Tandem (MS/MS) mass spectrum confirming the modification of angiotensin terminal carboxylic acid.

optimized.[17−19] Such discrimination is intrinsically challenging as the reactivities of carboxylic acid side chains and C-terminal carboxylates are very similar, and at least for full-length proteins, the acidic side chains are considerably more abundant than free C-termini. Hence, there are several innate chemical challenges to specifically derivatize protein or peptide C-termini, although the benefits from doing so are significant, especially for proteomics (e.g., isotopic quantitation, multiplexing, and sequencing). It is known that the C-terminal region of proteins can be very diverse within copies of the same protein with different levels of modifications[20,21] and proteolytic events[22] occurring to these regions that regulate a variety of processes such as RNA transcription[23] and protein aggregation.[24] A labeling and enrichment strategy for C-terminal peptides could serve as a discovery method for novel C-terminal variants that could be of biological importance.[25] A C-terminal derivatization strategy could be used to label the termini of proteins with an affinity tag such as biotin which can then serve as a handle for enrichment and subsequent MS analysis. Additionally, techniques that require a method of C-terminal selective labeling of peptides and proteins can help target them to imaging slides (for the fluorosequencing method) or providing the direction for translocation through nanopores.[26]

The selective labeling of C-terminal carboxylic acids of peptides in complex peptide mixtures for shotgun proteomics has been reported in the literature.[25] One such method involves oxazolone formation via the nucleophilic addition of C-terminal carboxylic acids to acetic anhydride.[27,28] Another method involves enzymatic coupling with carboxypeptidase, wherein the enzyme ligates a nucleophilic molecule to the C-terminal carboxylic acid at high pH.[18] However, in our experience, these methods lack the necessary C-terminal selectivity and mild reaction conditions that are required for applications in advanced proteomics. In our hands, we observed an indiscriminate cross-reaction of acetic anhydride with any number of peptide nucleophilic AA side chains, which blocked many ionizable groups, thus deterring subsequent conjugation reactions and hindering analysis by conventional MS. Similarly, the enzymatic method was synthetically laborious to implement—required time-consuming preparations prior to labeling, including esterification and lyophilization procedures—and the dramatic change in pH proved unsuitable for some peptides and proteins that were more pH-sensitive.

Recently, new methods have been described that offer the potential for faster, more generalized C-terminal modification via photoredox-catalyzed decarboxylation of C-terminal carboxylic acids.[29−31] These methods have been described for small peptides (most less than 10 AAs), but offer a starting point for applications to proteins, and the field of proteomics in general. Herein, we detail our studies to optimize and to assess the feasibility of photocatalyzed C-terminal labeling for bottom-up proteomics, simultaneously modifying and characterizing entire libraries of peptides from complex protein digests. We extensively describe the methods, instrumentation, and expectations for success when applying decarboxylative alkylation for bottom-up proteomics, such that it can be widely adopted by others.

RESULTS AND DISCUSSION

Optimizing the Conditions for Angiotensin Coupling.

We first sought to optimize the reaction conditions, starting from the procedure reported by Bloom et al.[29] The photoredox reactions are carried out in a commercially available instrument—Lumidox II system (Analytical Sales and Services, New Jersey)—fitted with an active cooling base (Analytical Sales and Services, NJ) and 445 nm blue light-emitting diodes (LEDs). A table fan was used to keep the reaction block cooled to room temperature for the duration of the experiment (6 h). (Note: Lumidox II allows for up to 96 photoredox reactions to be completed in parallel and is an optimal platform for small-scale reaction optimization and performance.)[32] A photograph of the setup is shown in Figure 1A. The small peptide angiotensin II (seq: H_2N-DRVYIHPF-CO_2H) and the Michael acceptor 3-methylene-2-norbornanone (herein termed NB for

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603

Table 1. Determination of Conversions of Peptides with Varying C-Terminal AAs

entry	S.M.	conversion[a]	entry	S.M.	conversion[a]
1	LYRAGA (Ala)	90%	12	LYRAGK (Lys)	83%
2	LYRAGR (Arg)	71%	13	LYRAGM (Met)	79%
3	LYRAGN (Asn)	81%	14	LYRAGF (Phe)	79%
4	LYRAGD (Asp)	87%	15	LYRAGP (Pro)	75%
5	LYRAGC (Cys)	N.D.[b]	16	LYRAGS (Ser)	79%
6	LYRAGE (Glu)	91%	17	LYRAGT (Thr)	75%
7	LYRAGQ (Gln)	89%	18	LYRAGW (Trp)	5.7%
8	LYRAGG (Gly)	51%	19	LYRAGY (Tyr)	65%
9	LYRAGH (His)	40%	20	LYRAGV (Val)	92%
10	LYRAGI (Ile)	94%	21	LYRAGP (Amide)	N.D.[b]
11	LYRAGL (Leu)	95%			

[a]Conversions were determined by the TIC peak area of the product divided by the TIC peak area of the product and S.M. [b]Not detected.

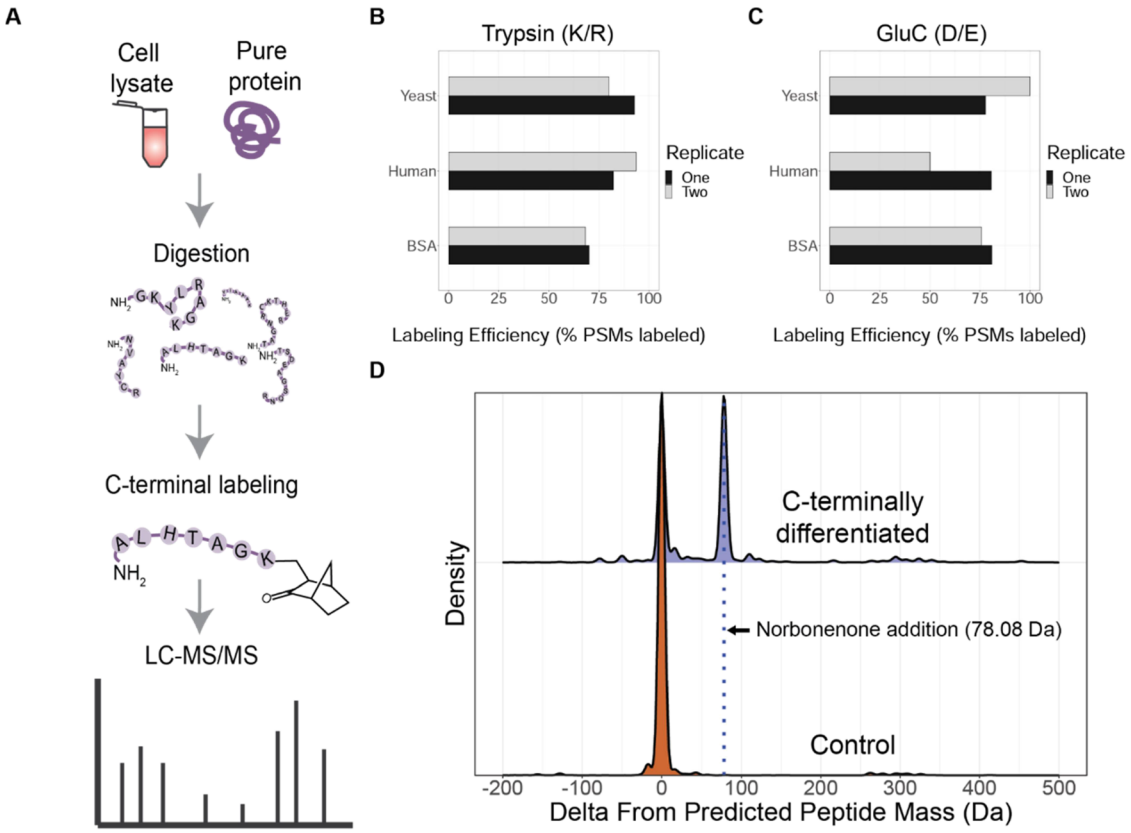


Figure 2. C-terminal carboxylic acid modifications occur with high efficiency on proteomic digests. (A) Workflow of the proteomics experiment with C-terminal labeling. C-terminal differentiation reaction labeling efficiency on (B) trypsin and (C) GluC-digested yeast, human, and BSA protein samples. Experiments performed in duplicate. (D) Delta mass values from the predicted peptide mass for the C-terminally differentiated sample (purple) and the control sample (orange). Blue line corresponds to the NB mass addition of 78.08 Da.

short) were used to screen experimental conditions for efficient conversion (reaction scheme shown in Figure 1B). The effect of aqueous buffers with various cosolvents, amounts of catalyst loading, equivalents of Michael acceptor, and angiotensin concentrations were all varied to optimize the reaction (summarized in ST1−T4). A representative liquid chromatography (LC) trace is shown in Figure 1C, which highlights the conversion efficiency to the angiotensin-NB adduct by comparison to the peaks corresponding to endogenous (unmodified) angiotensin and decarboxylated angiotensin peptide. The mass chromatogram of the corresponding trace is provided in Supplementary Figure SF1. The experiments with cesium formate buffer (pH 3.5) and dimethyl sulfoxide (DMSO) (5, 10, 20% v/v) resulted in high conversion. Considering that greater amounts of DMSO may diminish signals of modified and unmodified peptides, we elected to use only 5% DMSO (v:v) for further reactions.

Higher conversions (81%) could be achieved by increasing the catalyst loading to 10 mol % and using 10 equiv of the Michael acceptor (1 mM concentration). Finally, we optimized the reaction time and lamp power. We determined that the reaction performed the best with 110 mW/well and 6 h, giving an average conversion of ∼60%. High-resolution MS−MS results of the angiotensin-NB adduct indicated that the C-terminus of angiotensin was selectively modified (Figure 1D). The optimized conditions were reproducible at two

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603

independent locations (in setups at the University of Kansas and at UT-Austin).

Determining AA Biases in the Reaction Chemistry. Some C-terminal AAs are more susceptible to decarboxylative alkylation. Based on the mechanism proposed in the literature,[29] we rationalized that this bias could result from the stability of distinct α-amino radicals generated from different C-terminal residues (affecting both the rate of decarboxylation and the reactivity of the α-amino radical toward conjugate addition), as well as inherent differences in oxidation potentials between the various C-terminal residues. Thus, we investigated the variance in the yields of 21 short peptides, 20 of which have a unique biogenic AA at their C-terminus and 1 of which has its C-terminal acid replaced by an amide to serve as a negative control. The conversions are summarized in Table 1.

The data reveal that 13 of the 20 peptides gave greater than 75% conversion upon reaction with NB. The reaction of the C-terminal Cys-containing peptide with NB did not furnish the C-terminal modified product. Competitive thiol-conjugate addition was observed instead. Therefore, one will need to protect thiol groups to proceed with this C-terminal diversification strategy, as described below using iodoacetamide. It is worth noting that a tryptophan C-terminal peptide gave a much lower conversion of the modified product (~6%), one reason being the fast and reversible electron transfer between the electron-rich indole and the lumiflavin photocatalyst. The protonation of the imidazole ring of the C-terminal histidine peptide at pH 3.5 could deter single-electron transfer oxidation and might also diminish the nucleophilicity of the α-amino radical, and these issues could explain the lower conversions observed in this case (40%). As anticipated, a peptide lacking a C-terminal carboxylic acid gave no product.

After the initial optimization of the photoredox reaction, we suspected that there could be innate biases with respect to the C-terminal AAs.

Modifications Occur Efficiently on C-Terminal Carboxylic Acids in Heterogeneous Peptide Digests. To gain a better understanding of the applicability of the C-terminal differentiation reaction for proteomics experiments, we performed the reaction on heterogeneous peptide samples and probed the extent of the modifications using tandem MS (LC−MS/MS). We digested the protein bovine serum albumin (BSA), proteins from yeast lysate, and human cell lysate with endoproteases trypsin and GluC. In most cases, trypsin cleaves after lysine and arginine (K/R), and GluC cleaves after aspartate and glutamate (D/E). We then split these samples into control and experimental samples, where the only difference between the two samples was the presence or absence of the catalyst lumiflavin. After desalting and C-18 tip exchange, we analyzed the peptides using LC−MS/MS and searched for the NB modification in the analysis software as a C-terminal mass addition of 78.083 Daltons (Figure 2a). For each sample, we were able to estimate the C-terminal labeling efficiency as a ratio of peptide spectral matches (PSMs) with a C-terminal NB and the total PSMs observed in a sample. In the trypsin-digested samples, we observed an average labeling efficiency for each of our experimental digests ranging from 52% for the BSA sample to 76% for the human lysate (Figure 2b). Labeling efficiencies in GluC-digested samples ranged from 65% in the human lysate sample to 89% in the yeast lysate sample (Figure 2c). The difference in labeling efficiencies between these two sets of experiments may arise from the different terminal AAs generated by the two proteases, as suggested by the analysis shown in Table 1. However, the high extent of labeling should support most basic proteomics experiments.

Interestingly, when comparing the total peptide abundances between our negative control and experimental samples, we consistently saw, on average, a 76% reduction in the total high-confidence PSMs in our experimental samples (Figure SF2). We considered multiple hypotheses as to the cause of this decrease in peptide identifications. The first of these was that there were side reactions occurring in our experimental sample, resulting in unexpected modifications on peptides that would lower the confidence of identification. The second was that the proteomics software could not localize the C-terminal modification, resulting in a decrease in high-confidence PSMs. The third was that 445 nm irradiation in our experimental setup degraded some of our peptides in the presence of lumiflavin. To test the first of these hypotheses, we sought to identify what, if any, side reactions might have occurred during the reaction process with our previously collected shotgun proteomics data. Thus, we reanalyzed the mass spectra using the proteomics software MSFragger,[33] performing an open search (see the Methods section) to identify all chemical modifications—without specifying them in advance—that resulted in a change in mass from the predicted, unmodified peptide masses. Two distinct peaks were observed. One was centered on 0 Da, corresponding to the unmodified peptide, and the other was centered at 78.08 Da, closely corresponding to the expected mass addition of NB. In the control sample, we observed only a single large peak centered at 0 Da, as expected (Figure 2d). This analysis confirmed that the primary reaction observed was the expected C-terminal addition of NB and that our modification reaction only occurred in the presence of the lumiflavin photocatalyst. Based on the MSFragger analysis, we could thus rule out the presence of appreciable side reactions, leading to altered modification sites. Notably, even in the MSFragger open searches, we still observed a 73% average decrease in peptide identifications. These analyses thus suggest that some degradation due to irradiation in the presence of lumiflavin may cause a reduction in peptide abundances. Overcoming this limitation may call for the use of an alternative flavin photocatalyst, narrowed irradiation window, and other slight changes to the experimental protocol, which will serve as a starting point for follow-up studies. In spite of the reduced yield, the reaction proved to be highly selective when labeling the C-termini of peptides in a heterogeneous biological sample.

Minimum C-Terminal AA Bias Occurs on Peptides from Biological Samples. We next sought to determine if the C-terminal biases we observed with the synthetic peptides shown in Table 1 were also reflected in the biological samples. Any use of the decarboxylative alkylation chemistry on intact proteins and/or peptides or proteins whose terminal AAs vary will require minimal bias in the reaction chemistry for accurate quantification, particularly in fields such as C-terminomics.[25] To evaluate the C-terminal bias in complex mixtures, we again digested BSA, yeast, and human protein extract. To generate nonuniform C-terminal AAs, we used three proteases that cut specified AAs downstream of the *N*-terminus: LysN (K), AspN (D/E), and LysargiNase (K/R). Using these proteases, we were able to generate a mostly random population of C-terminal AAs and estimate the labeling efficiency of the C-terminal differentiation reaction. For most AAs, the median

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603

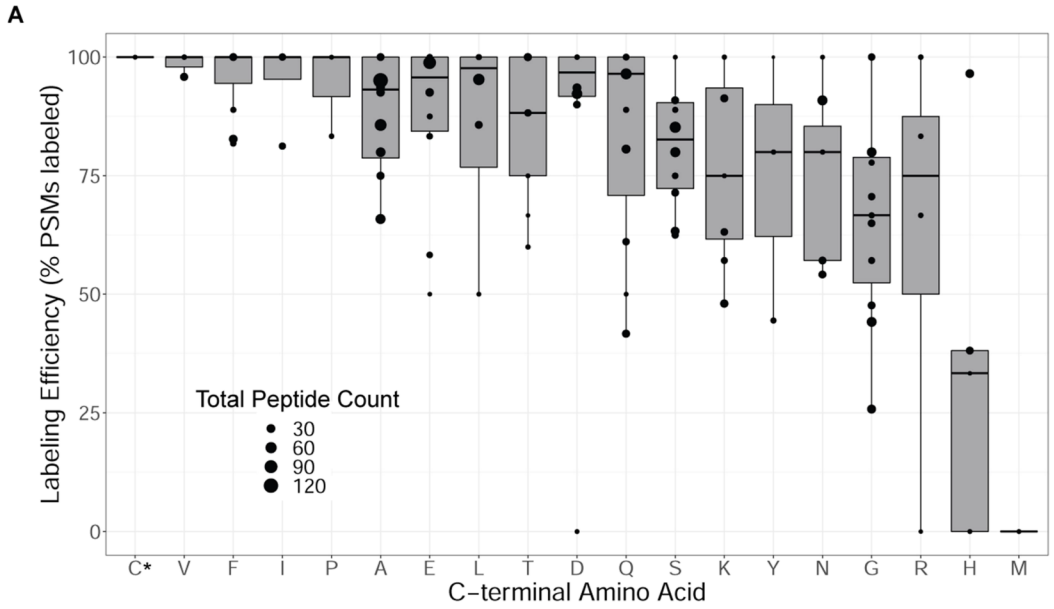


Figure 3. C-terminal labeling exhibits only moderate biases across peptides with varying C-terminal AA residues. The plot summarizes analyses of peptides generated from 15 independent experiments from digesting protein samples (BSA, yeast, and human cell protein extracts) using *N*-terminal cleaving proteases (AspN, lysN, and LysargiNase) that expose varying peptide C-terminal AAs. Peptides were identified by tandem MS, and the frequency of each C-terminal modified AA to its overall observed frequency was calculated to determine the labeling efficiency. Labeling efficiencies were determined for each of the 15 experiments indicated as separate data points. C* - Cysteine has been alkylated in advance using iodoacetamide.

labeling efficiency was in the range 75−100% (as seen in Figure 3). In certain cases, such as with histidine and methionine, the labeling efficiency was much lower than that expected based on the other AAs. For methionine, this is likely an artifact of low sampling rather than the actual labeling bias as the earlier experiments with the purified peptide showed an elevated labeling efficiency (Table 1). The histidine result corresponds with that observed in the purified peptide and has been discussed above. It should be noted that tryptophan was only observed once across all of the experiments and so was left out of this analysis, but it would have been expected to not be efficiently labeled based on Table 1. Because tryptophan indole is likely to affect the reaction conversion regardless of its position within a peptide, we analyzed the trypsin-digested samples to determine if any biases in the labeling efficiency could be observed for the 20 proteinogenic AAs when they occurred internally rather than on the C-terminus (Table 5). We observed that peptides containing an internal tryptophan exhibited a considerably lower labeling efficiency (45.5%) than any of the other AAs (84% on average). Cysteine and tyrosine also showed somewhat lower labeling efficiencies (67.4 and 62.7%, respectively). We do not expect the low labeling efficiency of His and Trp to affect potential applications, such as C-terminomics, based on their abundance in the human proteome, as together they make up the C-terminal AAs of only ∼4.5% of human proteins (3.33 and 1.18% for His and Trp, respectively). In comparison to previous studies using C-terminal derivatization, these biases are minimal. While labeling efficiencies were not reported from the use of the oxazolone chemistry on the thermophilic bacterium *T. tengocongensis*, oxazolone-derivatized C-terminal His and Trp were not well observed, and derivatized Cys was not observed at all.[27] The use of carboxypeptidase Y also showed a range of labeling efficiencies, with inefficient labeling of Cys, His, and Ile, and negligible reaction efficiency for Asn.[19] The high labeling efficiency across most of the C-termini demonstrates the robustness of the differentiation reaction. The largely unbiased nature of the reaction chemistry across peptides makes it highly attractive to methods such as C-terminal peptide enrichment strategies and single-molecule protein sequencing.

Adapting the Method for Single-Molecule Protein Sequencing. Our group recently developed a new proteomics technology, termed fluorosequencing, for identifying peptides in a complex mixture at a single-molecule resolution.[9] In this method, a researcher selectively labels AAs on peptides with fluorophores, immobilizes the peptides on a glass slide, subjects them to rounds of Edman degradation chemistry, and monitors the fluorescence of individual peptides after the loss of each consecutive AA per Edman cycle. The Edman cycle coinciding with the loss in fluorescent intensity (occurring due to the removal of the fluorescently labeled AA) indicates the position of the labeled AA in the sequence. This pattern, termed as a fluorosequence, is mapped to a reference database to identify the likely source peptide and proteins. As can be observed in the workflow, selectively labeling AAs with fluorophores and immobilizing the peptides to a solid imaging surface are two critical steps in the process. The selective labeling of the C-terminal carboxylic acid not only provides a unique attachment point for immobilization, but its exclusivity also allows for subsequent labeling of internal acidic residues using alternate technologies, such as amide coupling, which can enhance the fluorosequencing method when used in tandem. We implemented the peptide-labeling process for angiotensin (illustrated in Figure 4A) by first conjugating the C-terminal carboxylic acid with a synthesized NB-PEG4-alkyne linker in solution using the above-described and optimized photoredox chemistry, then capturing the peptide on the solid support via its *N*-terminal amine using a pyridine carboxaldehyde reagent (PCA) (procedure adapted

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603

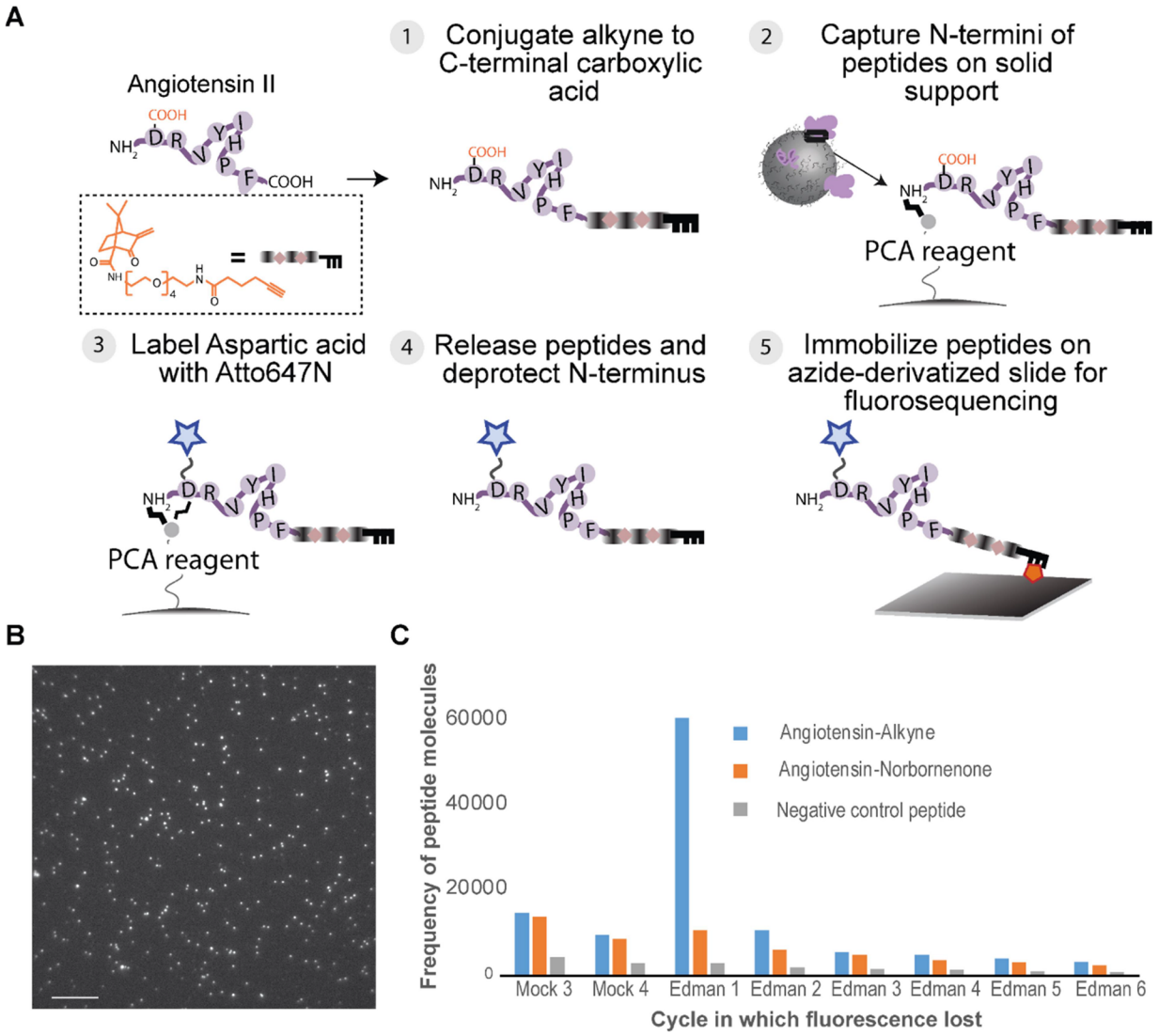


Figure 4. Application of C-terminal differentiation to immobilize peptides for fluorosequencing, enabling single-molecule determination of the sequence position of fluorophore-derivatized internal acidic AAs. The C-terminal carboxylic acid of angiotensin was labeled with NB-PEG4-alkyne, and its internal aspartic acid was labeled with fluorophore Atto647N prior to single-molecule peptide sequencing. (A) The steps involved in the process are illustrated. (B) A representative image of the numerous fluorescently labeled single peptide molecules of angiotensin labeled with Atto647N and immobilized on an azide-functionalized glass slide. (C) The bar chart shows the result of the fluorosequencing experiment performed on these labeled and immobilized angiotensin molecules, with (angiotensin-alkyne, blue bars) and without alkyne (red bars) and negative control (peptide-AGAGANGSNFGAN-(CO)NH_2, gray bar), which is incapable of C-terminal functionalization as well as labeling. The peptides were exposed to four cycles of "mock" Edman degradation (performing the full set of treatments, but omitting phenyl isothiocyanate)

from a previous study[15]), followed by a two-step labeling process to attach a fluorophore on glutamic acid. The two-step coupling reaction involves an amide coupling using 3-azidopropylamine, followed by installing an Atto647N-PEG4-DBCO using standard copper-free click chemistry. Following the labeling step and washing away the free uncoupled fluorophores from the solid support, we released and deprotected the peptide scarlessly from the support. After a serial dilution through four orders of magnitude, we immobilized the liberated and fluorescently labeled peptide on an azide-functionalized glass surface using copper-catalyzed click chemistry (Figure 4B, see the Methods section). Fluorosequencing this peptide (seen in the blue bar in Figure 4C) reveals the position of the glutamic acid residue on individual molecules of angiotensin, as seen by the preferential removal of the fluorophore after the first Edman cycle. As suitable negative controls, we also fluorosequenced (a) angiotensin, but without the alkyne linker to control for the nonspecific attachment of the labeled peptide on the azide slides and (b) a negative control peptide of sequence H_2N-AGAGANGSNFGAN-(CO)NH_2, with an amidated C-terminal carboxylic acid to serve as a control for excess fluorophores carried along through the reaction. The 20-fold increase in the counts of fluorescence loss after the first Edman cycle for angiotensin when compared with the controls confirms the utility of the selective C-terminal decarboxylative alkylation chemistry for enabling single-molecule sequencing of internal carboxylic acid-containing AA residues.

CONCLUSIONS

Photoredox-catalyzed decarboxylative alkylation has been reported to label the C-terminal carboxylic acid residue of

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603

peptides with high efficiency, but the vast applications of this methodology are untapped. With the goal of applying this approach to the field of bottom-up proteomics and in other emerging single-molecule proteomics technologies, we first optimized the photocatalyzed procedure using endogenous angiotensin as our target substrate. We then tested this method with 21 synthetic peptides, each having a chemically unique C-terminal residue, and compared the labeling efficiencies. Following this, we applied the decarboxylation protocol to biological samples and estimated the labeling efficiency in these complex mixtures. A search of possible alternative modifications verified that there were no prominent side reactions. We then used a diverse set of proteases to generate and to characterize a mostly random pool of C-terminal peptides in a biological sample, in terms of labeling efficiency and bias. These results both showed minimal AA bias in complex solutions and corresponded closely to the labeling biases observed with purified peptides. Finally, we demonstrated the utility of decarboxylative alkylation for single-molecule proteomics (namely, fluorosequencing) by using this chemistry to append an alkyne-containing linker to the C-terminus of angiotensin II, and then using orthogonal chemistries to couple and to sequence the peptide on-slides using total internal reflection fluorescence (TIRF) microscopy. We anticipate that this differentiation reaction will be broadly useful for proteomics applications.

METHODS

Materials. The main materials purchased were angiotensin (Sigma, Cat #05-23-0101), 3-methylene-2-norbornanone (Sigma; Cat# M46055), lumiflavin (Cayman Chemical, Cat# 20645), and cesium formate (Alfa Aesar, Cat #B2132714). All synthetic peptides were synthesized by Genscript (NJ, USA). Intact protein and trypsin/LysC-digested peptides for BSA (Cat# 88341, Thermo), yeast (*Saccharomyces cerevisiae*), cell protein extract (Cat# V7341, V7461), and human K562 cell protein extract (Cat# V6951, V6941) were purchased from Promega Inc. (WI, USA). The NB-PEG4-alkyne linker was custom-synthesized and purified by Comminex Inc. (Budapest, Hungary) based on the procedure described by Bloom et al.[29] Proteases Pierce trypsin (Cat #90058), Pierce GluC (Cat #90054), and Pierce LysN (Cat#90300) were purchased from Thermo Scientific. LysArginase (Cat #EMS0008) was purchased from Millipore Sigma (WI, USA). AspN (Cat#V1621) was purchased from Promega Inc. (Cat #V1621). 2,2,2-trifluoroethanol (Cat #T63002) and tris(2-carboxyethyl)phosphine (Cat #77720) were purchased from Sigma-Aldrich. Pierce C18 tips (Cat #87784) were purchased from Thermo Scientific.

NB Conjugation. *Preparation of Stock Solutions.* Peptide solution (4.2 mM): 1 mg of angiotensin dissolved in 227 μL of ddH_2O, lumiflavin solution (0.78 mM): 1 mg of lumiflavin dissolved in 5 mL of ddH_2O, and 3-methylene-2-norbornanone (0.2 M): 2.052 μL in 84 μL of DMSO.

Procedure for Angiotensin. To a 1 mL vial, peptide (0.42 μmol, 1 equiv) in 100 μL of ddH_2O and lumiflavin (42 nmol, 10 mol %) in 53 μL of ddH_2O were added from the stock solutions. To this reaction vial, 3-methylene-2-norbornanone (4.2 μmol, 10 equiv) in 21 μL of DMSO, 42 μL of cesium formate buffer (pH 3.5, 0.1 M), and 204 μL of ddH_2O were added. The resulting solution was degassed by sparging with nitrogen for 3 min, parafilmed, and irradiated using second-generation Analytica blue LED lamps. The sample was irradiated under 110 mW for 6 h with stirring at 550 rpm and fan cooling. After 6 h, the sample was removed, diluted with 523 μL of methanol, and filtered. To this solution, 1 equiv of *p*-toluic acid (in 57 μL of methanol) was added as an internal standard and subjected to LC−MS.

Calculation of Reaction Conversion. Initial optimization with angiotensin was studied using a Waters Acquity ultrahigh-performance liquid chromatography (UPLC) (Empower Software) connected to an Advion Expression CMS mass spectrometer. Method: Solvent A: acetonitrile, Solvent B: water, (5% A and 95% B, 0−2 min; 10% A and 90%, 2−3 min; 20% A and 80% B, 3−10 min; 30% A and 70% B, 10−12 min; 95% A and 5% B, 12−16.5 min; 95% A and 5% B, 16.5−18 min; 5% A and 95% B, 18−20 min). For the further optimization of angiotensin at UT-Austin, LC/MS were recorded on an Agilent Technologies 6120 Single Quadrupole or 6130 Single Quadrupole interfaced with an Agilent 1200 series LC system equipped with a diode-array detector. The resulting spectra were analyzed using Agilent LC/MSD ChemStation. All LC experiments were run with a 5−95% gradient elution (methanol/water) over 15 min. The conversion was calculated by the ratio of the peak areas of C-terminal-modified angiotensin and *p*-toluic acid (internal standard) under UV trace for angiotensin photoredox reaction optimization reaction (Supporting Information Table 1−4) For the determination of conversions of peptides with varying C-terminal AAs (Table 1), the value was obtained by dividing the peak area of the C-terminal modified product by the addition of peak areas of the C-terminal modified product and unreacted starting materials under the total ion chromatogram (TIC).

Protein Digestion. Mass spec-compatible yeast (100 μg), human K562 cell protein extracts, and BSA were denatured in 2,2,2-trifluoroethanol (TFE) and 5 mM tris(2-carboxyethyl)phosphine (TCEP) at 55 °C for 45 min. Proteins were alkylated in the dark with 5.5 mM iodoacetamide, and then, the remaining iodoacetamide was quenched in 100 mM dithiothreitol (DTT). MS-grade GluC, LysargiNase, LysN, or AspN was then added to the solution at enzyme:protein ratios of 1:50, 1:50, 1:50, and 1:25, respectively, and the digestion reaction was incubated at 37 °C for 24 h. Samples were then frozen at −20 °C, thawed, and the volume was reduced to 500 μL in a vacuum centrifuge. Samples were then filtered using a 10 kDa Amicon filter and desalted using Pierce C18 tips (Thermo Scientific). The samples were resuspended in 95% water, 5% acetonitrile, and 0.1% formic acid prior to MS.

Procedure for the Enzyme-Digested Sample. Following the reaction procedure of the model peptide, to a 1 mL vial, 100 μg of cell lysate was dissolved in 200 μL of ddH_2O, and the sample solution was separated into two 1 mL vials (one for reaction, the other as control). To the reaction vial, lumiflavin (42 μmol, 10 mol %) in 53 μL of ddH_2O was added from the stock solutions. To this reaction vial, 3-methylene-2-norbornanone (4.2 mmol, 10 equiv) in 21 μL of DMSO, 42 μL of cesium formate buffer (pH 3.5, 0.1 M), and 204 μL ddH_2O were added. The resulting solution was degassed by sparging with nitrogen for 3 min, parafilmed, and irradiated using the second-generation Analytica blue LED lamps. The sample was irradiated under 110 mW for 6 h with stirring at 550 rpm and fan cooling. After 6 h, the sample was removed, diluted with 523 μL of methanol, and filtered. To this solution, 1 equiv of *p*-toluic acid (in 57 μL of methanol) was added as an internal standard and subjected to LC−MS. For the control experiment, instead of the addition of lumiflavin (42 μmol, 10 mol %) in 53 μL of ddH_2O, 53 μL of ddH_2O only was added, and the rest procedures described above were followed for the reaction vial.

Mass Spectrometry. Peptides were separated on a 75 μM × 25 cm Acclaim PepMap100 C-18 column (Thermo Scientific) using a 5−50% acetonitrile + 0.1% formic acid gradient over 60 or 120 min (for BSA and the lysate samples, respectively) and analyzed online by nanoelectrospray-ionization tandem MS on an Orbitrap Fusion (Thermo Scientific). Data-dependent acquisition was activated with parent-ion (MS1) scans collected at a high resolution (120,000). Ions with charge +1 were selected for collision-induced dissociation fragmentation spectrum acquisition (MS2) in the ion trap using a Top Speed acquisition time of 3 s. Dynamic exclusion was activated with a 60 s exclusion time for ions selected more than once. MS proteomics data were acquired in the UT-Austin Proteomics Facility and have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with data set identifiers PXD026393.

Protein Identification. Proteins were identified with Proteome Discoverer 2.3 (Thermo Scientific), searching against the UniProt

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603

human, yeast, or bovine reference proteome. Modifications of the C-terminus were added with the expected NB mass addition of [+78.083 Da], in addition to methionine oxidation [+15.995 Da], *N*-terminal acetylation [+42.011 Da], *N*-terminal methionine loss [−131.04 Da], and *N*-terminal methionine loss with the addition of acetylation [−89.03 Da]. Peptide and protein identifications were thresholded at a 1% false discovery rate.

MSFragger Analysis. Raw files from MS experiments were converted to the mzML format using MSConvert with the peak-picking filter selected. The files were split into their respective biological directories and analyzed in two separate batches (one for yeast and one for human protein). Modifications were identified using MSFragger 2.4 with the yeast proteome UP000002311 (accessed 2/3/2021) and the human proteome UP000005640 (accessed 3/25/2020) containing reversed protein sequences as a decoy database. We performed an open search with a 75 ppm fragment mass tolerance and the upper and lower precursor mass windows set to 500 and − 200, respectively. *N*-terminal acetylation [+42.01060 Da] and methionine sulfoxidation [+15.99490 Da] were considered as variable modifications. Spectra were assigned to peptides and processed using PeptideProphet with default settings. Protein identification was performed using ProteinProphet with default settings. A false discovery rate for all identification steps was set at 1%.

Fluorophore Labeling of Angiotensin. Angiotensin (1 mg) (∼1 μmole) (in 100 μL) was first labeled with 50 eq of NB-PEG4-alkyne along with lumiflavin (10% mol/mol of angiotensin) in 0.1 M cesium formate buffer using the setup as optimized and described earlier. The negative controls were set up with (a) 1 mg of angiotensin and 50 eq NB (lacks alkyne handle) and (b) 1 μmole of inert peptide (amide terminal and no internal acidic residue) with sequence AGA-GANGSNFGAN-(CO)NH_2. After the reaction, we adjusted the pH of the solution to 8.5 using 100 μL of HEPES buffer (pH 8.5) and made up the volume to 600 μL with water. We then incubated these peptide solutions with PCA-functionalized synphase lanterns (Mimotopes) for 16 h at 37 °C to capture the peptides onto the lanterns. We then performed a series of washes of the lanterns with acetonitrile/water (1:1 v/v), acetonitrile/water with 0.1% formic acid (1:1 v/v), acetonitrile, and dimethylformamide by soaking the lantern for 5 min in each solvent. To label the internal aspartic acid, in the first step, we performed amide coupling to conjugate the azide group on the AA side chain using 50 eq 3-azidopropylamine (Cat#A2738, TCI), 50 eq HCTU, and 50 eq DIEA base in 600 μL of dimethylformamide (DMF) and incubating at 37 °C for 2 h. Following the washing of lanterns, we added 1 eq of Atto647N-PEG4-DBCO in a 600 μL of 1:1 v/v of DMF/water and incubated the lantern for 16 h at room temperature. We again washed the lanterns to remove the excess dyes. The labeled peptides were cleaved by reacting the lanterns with 600 μL of trifluoroacetic acid (TFA) cocktail (95% TFA, 2.5% water, and 2.5% triisopropylsilane) at room temperature for 2 h. After blowing dry N_2 to remove the TFA and cold ether precipitation, we solubilized the fluorophore-labeled peptides with 50 μL of the acetonitrile/water mixture. We then deprotected the PCA-modified N-terminus of the peptide by incubating in hydrazine (dimethylaminoethylhydrazine dihydrochloride) at 60 °C for 16 h.[15]

Peptide Surface Immobilization. For single-molecule peptide sequencing, a 40 mm German Desag 263 borosilicate glass coverslip (Bioptechs) surface was first cleaned by UV/ozone (Jelight Company) and then functionalized by soaking for 30 min in methanol containing 0.01% azidopropyltriethoxysilane (Gelest, Cat # SIA0777.0) and 4 mM acetic acid. Weakly attached silane was removed by gentle agitation for 10 min in a bath of methanol and a subsequent 10 min in water. The coverslip with immobilized peptides was dried under a nitrogen gas stream and baked in a vacuum oven for 20 min at 110 °C.

Fluorosequencing. Angiotensin peptides were covalently coupled to the coverslip surface *via* copper-catalyzed click chemistry between the alkyne-modified C-terminal AA residue and the azido silane. A fresh solution of 2 mM copper sulfate, 1 mM tris(3-hydroxypropyltriazolylmethyl)amine (Sigma, Cat # 762342), 20 mM HEPES (pH 8.0), and 5 mM sodium ascorbate with fluorescently labeled angiotensin was incubated for 30 min at room temperature on the coverslip, washed with water to remove unbound peptides, and dried under a nitrogen gas stream. Single-molecule sequencing was performed as described.[9] Fluorosequencing datasets were analyzed using the *SigProc* software tool, available as part of the Plaster package at https://github.com/erisyon/plaster. The raw image files are uploaded to Zenodo (doi: 10.5281/zenodo.4989148).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.1c00631.

Additional experimental details and data including LC−MS results, tandem MS analyses, and reaction optimization conditions. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Steven Bloom − *School of Pharmacy - Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66045-7572, United States;* orcid.org/0000-0002-4205-4459; Email: spbloom@ku.edu

Edward M. Marcotte − *Department of Molecular Biosciences, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0001-8808-180X; Email: marcotte@utexas.edu

Eric V. Anslyn − *Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0002-5137-8797; Email: anslyn@austin.utexas.edu

Authors

Le Zhang − *Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States*

Brendan M. Floyd − *Department of Molecular Biosciences, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0001-7412-0492

Maheshwerreddy Chilamari − *School of Pharmacy - Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66045-7572, United States*

James Mapes − *Erisyon, Inc., Austin, Texas 78701, United States*

Jagannath Swaminathan − *Department of Molecular Biosciences, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas 78712, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.1c00631

Author Contributions

⊥L.Z. and B.M.F. contributed equally.

Notes

The authors declare the following competing financial interest(s): J.S., E.M.M., and E.V.A. are co-founders and shareholders of Erisyon, Inc., and E.M.M. and E.V.A. serve on the scientific advisory board. J.S, E.M.M, E.V.A, B.M.F and L.Z are co-inventors on a patent application relevant to this work.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the UT Mass Spectrometry Facility for their assistance. This work was supported by the Welch Regents Chair (F-0046) to E.V.A. and by Erisyon, Inc., to E.M.M. and E.V.A. E.M.M. acknowledges additional support from the Welch Foundation (F-1515), NIH (R35 GM122480, R01 HD085901, R01 DK110520) and NSF STTR Grant (#1938726) with Erisyon Inc. Mass spectrometry data collection was supported by CPRIT grant RP110782 to Maria Person. J.S., E.M.M., and E.V.A. are the co-founders and shareholders of Erisyon, Inc., and E.M.M. and E.V.A. serve on the scientific advisory board. J.S., E.M.M., E.V.A., B.M.F., and L.Z. are co-inventors on a patent application relevant to this work.

■ REFERENCES

(1) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present. *Anal. Bioanal. Chem.* **2012**, *404*, 939−965.

(2) Restrepo-Pérez, L.; Joo, C.; Dekker, C. Paving the Way to Single-Molecule Protein Sequencing. *Nat. Nanotechnol.* **2018**, *13*, 786−796.

(3) Chen, J.; Brunner, A.-D.; Cogan, J. Z.; Nuñez, J. K.; Fields, A. P.; Adamson, B.; Itzhak, D. N.; Li, J. Y.; Mann, M.; Leonetti, M. D. Pervasive Functional Translation of Noncanonical Human Open Reading Frames. *Science* **2020**, *367*, 1140−1146.

(4) Marx, V. A Dream of Single-Cell Proteomics. *Nat. Methods* **2019**, *16*, 809−812.

(5) Kelly, R. T. Single-Cell Proteomics: Progress and Prospects. *Mol. Cell. Proteomics* **2020**, *19*, 1739−1748.

(6) Banerjee, S. Empowering Clinical Diagnostics with Mass Spectrometry. *ACS Omega* **2020**, *5*, 2041−2048.

(7) O'Reilly, F. J.; Rappsilber, J. Cross-Linking Mass Spectrometry: Methods and Applications in Structural, Molecular and Systems Biology. *Nat. Struct. Mol. Biol.* **2018**, *25*, 1000−1008.

(8) Pappireddi, N.; Martin, L.; Wühr, M. A Review on Quantitative Multiplexed Proteomics. *ChemBioChem* **2019**, *20*, 1210−1224.

(9) Swaminathan, J.; Boulgakov, A. A.; Hernandez, E. T.; Bardo, A. M.; Bachman, J. L.; Marotta, J.; Johnson, A. M.; Anslyn, E. V.; Marcotte, E. M. Highly Parallel Single-Molecule Identification of Proteins in Zeptomole-Scale Mixtures. *Nat. Biotechnol.* **2018**, *36*, 1076−1082.

(10) Ouldali, H.; Sarthak, K.; Ensslen, T.; Piguet, F.; Manivet, P.; Pelta, J.; Behrends, J. C.; Aksimentiev, A.; Oukhaled, A. Electrical Recognition of the Twenty Proteinogenic Amino Acids Using an Aerolysin Nanopore. *Nat. Biotechnol.* **2020**, *38*, 176−181.

(11) Restrepo-Pérez, L.; Wong, C. H.; Maglia, G.; Dekker, C.; Joo, C. Label-Free Detection of Post-Translational Modifications with a Nanopore. *Nano Lett.* **2019**, *19*, 7957−7964.

(12) Hernandez, E. T.; Swaminathan, J.; Marcotte, E. M.; Anslyn, E. V. Solution-Phase and Solid-Phase Sequential, Selective Modification of Side Chains in KDYWEC and KDYWE as Models for Usage in Single-Molecule Protein Sequencing. *New J. Chem.* **2017**, *41*, 462−469.

(13) Chen, X.; Wu, Y.-W. Selective Chemical Labeling of Proteins. *Org. Biomol. Chem.* **2016**, *14*, 5417−5439.

(14) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-Step Site-Specific Modification of Native Proteins with 2-Pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, *11*, 326−331.

(15) Howard, C. J.; Floyd, B. M.; Bardo, A. M.; Swaminathan, J.; Marcotte, E. M.; Anslyn, E. V. Solid-Phase Peptide Capture and Release for Bulk and Single-Molecule Proteomics. *ACS Chem. Biol.* **2020**, *15*, 1401−1407.

(16) Rosen, C. B.; Francis, M. B. Targeting the N terminus for Site-Selective Protein Modifications. *Nat. Chem. Biol.* **2017**, *13*, 697−705.

(17) Peschke, B.; Bak, S. Controlled Coupling of Peptides at Their C-Termini. *Peptides* **2009**, *30*, 689−698.

(18) Duan, W.; Xu, G. ProC-TEL: Profiling of Protein C-Termini by Enzymatic Labeling. *Methods Mol. Biol.* **2017**, *1574*, 135−144.

(19) Xu, G.; Shin, S. B. Y.; Jaffrey, S. R. Chemoenzymatic Labeling of Protein C-Termini for Positive Selection of C-Terminal Peptides. *ACS Chem. Biol.* **2011**, *6*, 1015−1020.

(20) Hori, Y.; Kikuchi, A.; Isomura, M.; Katayama, M.; Miura, Y.; Fujioka, H.; Kaibuchi, K.; Takai, Y. Post-Translational Modifications of the C-Terminal Region of the Rho Protein Are Important for its Interaction with Membranes and the Stimulatory and Inhibitory GDP/GTP Exchange Proteins. *Oncogene* **1991**, *6*, 515−522.

(21) Crosbie, R. H.; Miller, C.; Cheung, T.; Goodnight, T.; Muhlrad, A.; Reisler, E. Structural Connectivity in Actin: Effect of C-Terminal Modifications on the Properties of Actin. *Biophys. J.* **1994**, *67*, 1957−1964.

(22) Sarnovsky, R.; Rea, J.; Makowski, M.; Hertle, R.; Kelly, C.; Antignani, A.; Pastrana, D. V.; Fitzgerald, D. J. Proteolytic Cleavage of a C-terminal Prosequence, Leading to Autoprocessing at the *N* Terminus, Activates Leucine Aminopeptidase from *Pseudomonas aeruginosa*. *J. Biol. Chem.* **2009**, *284*, 10243−10253.

(23) Phatnani, H. P.; Greenleaf, A. L. Phosphorylation and Functions of the RNA Polymerase II CTD. *Genes Dev.* **2006**, *20*, 2922−2936.

(24) Afitska, K.; Fucikova, A.; Shvadchak, V. V.; Yushchenko, D. A. Modification of C Terminus Provides New Insights into the Mechanism of α-Synuclein Aggregation. *Biophys. J.* **2017**, *113*, 2182−2191.

(25) Tanco, S.; Gevaert, K.; Van Damme, P. C-terminomics: Targeted Analysis of Natural and Posttranslationally Modified Protein and Peptide C-Termini. *Proteomics* **2015**, *15*, 903−914.

(26) Alfaro, J. A.; Bohländer, P.; Dai, M.; Filius, M.; Howard, C. J.; van Kooten, X. F.; Ohayon, S.; Pomorski, A.; Schmid, S.; Aksimentiev, A.; et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **2021**, *18*, 604−617.

(27) Liu, M.; Fang, C.; Pan, X.; Jiang, H.; Zhang, L.; Zhang, L.; Zhang, Y.; Yang, P.; Lu, H. Positive Enrichment of C-Terminal Peptides Using Oxazolone Chemistry and Biotinylation. *Anal. Chem.* **2015**, *87*, 9916−9922.

(28) Kim, J.-S.; Shin, M.; Song, J.-S.; An, S.; Kim, H.-J. C-Terminal de Novo Sequencing of Peptides Using Oxazolone-Based Derivatization with Bromine Signature. *Anal. Biochem.* **2011**, *419*, 211−216.

(29) Bloom, S.; Liu, C.; Kölmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; Mac Millan, D. W. C. Decarboxylative Alkylation for Site-Selective Bioconjugation of Native Proteins via Oxidation Potentials. *Nat. Chem.* **2018**, *10*, 205−211.

(30) Garreau, M.; Le Vaillant, F.; Waser, J. C-Terminal Bioconjugation of Peptides through Photoredox Catalyzed Decarboxylative Alkynylation. *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 8182−8186.

(31) Du, E. L.; Garreau, M.; Waser, J. Small Peptide Diversification through Photoredox-Catalyzed Oxidative C-Terminal Modification. *Chem. Sci.* **2021**, *12*, 2467−2473.

(32) Immel, J.; Chilamari, M.; Bloom, S. Combining Flavin Photocatalysis with Parallel Synthesis: A General Platform to Optimize Peptides with Non-Proteinogenic Amino Acids. *Chem. Sci.* **2021**, *12*, 10083.

(33) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry−Based Proteomics. *Nat. Methods* **2017**, *14*, 513−520.

https://doi.org/10.1021/acschembio.1c00631
ACS Chem. Biol. 2021, 16, 2595−2603