



Published in final edited form as:

ACS Chem Biol. 2020 June 19; 15(6): 1401–1407. doi:10.1021/acscchembio.0c00040.

## Solid-Phase Peptide Capture and Release for Bulk and Single-Molecule Proteomics

Cecil J Howard<sup>[a]</sup>, Brendan M. Floyd<sup>[b]</sup>, Angela M. Bardo<sup>[b]</sup>, Jagannath Swaminathan<sup>[b]</sup>, Edward M. Marcotte<sup>[b]</sup>, Eric V. Anslyn<sup>[a]</sup>

<sup>[a]</sup>Department of Chemistry, University of Texas at Austin, 100 E 24<sup>th</sup> St Austin, TX 78712

<sup>[b]</sup>Department of Molecular Biosciences, University of Texas at Austin, 2500 Speedway, Austin, TX 78712

### Abstract

The field of proteomics has expanded recently with more sensitive techniques for the bulk measurement of peptides as well as single-molecule techniques. One limiting factor for some of these methods is the need for multiple chemical derivatizations and highly pure proteins free of contaminants. We demonstrate a solid-phase capture-release strategy suitable for the proteolysis, purification, and subsequent chemical modification of peptides. We use this resin on an HEK293T cell lysate and perform one-pot proteolysis, capture, and derivatization to survey peptide capture biases from over 40,000 unique peptides from a cellular proteome. We also show that this capture can be reversed in a traceless manner, such that it is amenable for single-molecule proteomics techniques. With this technique, we perform a fluorescent labeling and C-terminal derivatization on a peptide and subject it to fluorosequencing, demonstrating that washing the resin is sufficient to remove excess dyes and other reagents prior to single-molecule protein sequencing.

### Keywords

peptides; protein derivatization; proteomics; single molecule protein sequencing; solid-phase peptide capture

### Introduction

With the increasing sensitivity of proteomic methods, a number of new proteins, protein isoforms<sup>1</sup>, and post-translational modifications<sup>3</sup> have been discovered. The increase in sensitivity is due both to improvements of the mass spectrometers and to the ability to generate high quality protein/peptide samples that are often highly derivatized<sup>4</sup>. In addition, new single-molecule protein sequencing techniques are being developed, including by the use of modified nanopore sequencing pores<sup>5-6</sup> and fluorosequencing<sup>7</sup>. These techniques may rely only on conventional bottom-up proteomic preparations using proteases, and may also employ selective detection of site-specific fluorescent labelling of amino acids. Simulation studies suggest two or more separate fluorescent labels are required for adequate wide

coverage of the proteome, and currently this would be extremely difficult for standard solution-phase labelling of samples<sup>8-10</sup>. Moreover, purification techniques are prone to sample losses, and the inclusion of multiple derivatization / purification cycles can lead to low abundance peptides dropping below the detection thresholds<sup>11</sup>. This can lead in turn to a bias against rare or low abundance peptides that may be biologically important<sup>12</sup>.

One method that we envisioned could be used to improve sample preparation is to bind the proteins/peptides to a bead or other solid matrix. While such an approach has been done before, immobilization typically relies on either the addition of non-natural amino acids as purification handles<sup>13</sup>, relies on non-specific hydrophobic precipitation<sup>14</sup>, charge based non-covalent bonding, or in some cases metal chelation strategies<sup>15</sup>. These properties make them unattractive for studies in which installing non-natural amino acids *via* amber codon suppression is difficult, such as in mammalian cell cultures<sup>16</sup>, and/or due to the limitations in solvent/buffer conditions that are compatible with non-covalent immobilization techniques,<sup>15</sup> or to the loss of peptides due to non-covalent and weak attachment to the resins.

A method that allows for the binding of peptides on resin support in a covalent and reversible manner would enable complex manipulations with higher overall yields. As demonstrated below in the context of fluorosequencing, the method introduced herein allows for the capture, purification and derivatization of peptides, and in theory, detection of low abundance peptides. Importantly, this procedure allows for derivatization schemes that could otherwise not be performed because of the use of excess reagents and washing steps, analogous to peptide synthesis on resin, where experimental procedures are optimized to impart high yield and speed rather than the need for purification<sup>17</sup>.

In order for such a procedure to be applicable to all peptides generated during a proteolytic cleavage, the ideal capture reagent would only react with the N- or C-terminus of a peptide/protein and not with any of the side chains that may be present. A specific reaction with the N-terminus is an attractive choice, as all peptides generated *via* protease action possess a free amino terminus, and the protease generally imparts knowledge of either the N- or C-terminal amino acids due to its specificity of cleavage<sup>18</sup>. The N-terminus is also a good target due to the difference in its conjugate acid's  $pK_a$ , as well as nucleophilicity, when comparing it to the sidechain amine on lysine<sup>19</sup>. While there are methods that allow for the selective modification of the N-terminal amine over lysine amines, most of them rely on permanently modifying the N-terminus, such as creating a new amide bond<sup>20</sup> or converting the N-terminal amino acid into an aldehyde<sup>21</sup>. It is also possible to use the C-terminus as a handle for peptide purification. However, due to the similarities between the side chain carboxylic acids of Asp and Glu and the C-terminus this is a less attractive option, as the  $pK_a$  of these three range from 3.3–4.2<sup>22</sup>.

Here we describe the use of imidazolinone formation from the N-terminus of peptides using 2-pyridinyl carboxaldehyde (2PCA) (Scheme 1) as first introduced by MacDonald *et al.*<sup>2a</sup> This was an attractive option for a next-generation capture resin that would have no cross-reactivity with lysine residues, unlike current styrene-aldehyde resins that rely on unstable imine formation with Lys and the N-terminus<sup>23</sup>. However, in order to be a tractable method for purification, a chemical trigger to reverse the stable imidazolinone ring in a traceless

manner is needed. Thus, as described below, we herein report a method for the covalent, reversible, and efficient peptide capture and release<sup>24</sup>.

## RESULTS AND DISCUSSIONS

### Optimization of Aldehyde Linker

To understand substituent effects on the peptide capture we screened aromatic and heteroaromatic aldehydes possessing different rings, heteroatoms, and regiochemical placement of the aldehyde. In total, 30 aldehydes were tested and ranked in order of the amount of N-terminally capped product, as an imidazolinone, formed on the model peptide Ser-Gly-Trp in a six-hour reaction at 37 °C in 50 mM sodium phosphate buffer pH 7.5 (Figure 1). Each reaction was analyzed by LC/MS, and iminal formation was confirmed by the presence of two distinct peaks in the 218 nm HPLC trace that had masses corresponding to the imidazolinone capped product. The two peaks are indicative of the two diastereomers<sup>2b</sup> formed during the ring closure and were observed in all reactions tested.

We found that compounds that contain strongly electron-withdrawing groups (Fig 1A and F) do not lead to imine-intermediates, the step required prior to ring closure. We postulate that this is due to the aldehyde being largely hydrated, which does not reverse to allow imine formation. However, less electron withdrawing character facilitates some product formation, but remains at unacceptable yields (J, L, N, O, Q, R, W). Imidazolinone formation is also disfavored when the aldehyde is on an electron rich aromatic ring, such as a thiazole/pyrrole (C, D, E, G, H, K), or has a substituent with a large negative Hammett sigma-value (M, albeit a Hammett plot was not generated). Aldehydes that are known to promote the formation of the imine complex through intramolecular hydrogen bonding or through a general-acid catalyzed mechanism<sup>25</sup>, albeit having a negative Hammett value, can promote product formation (V).

Clearly some electronic withdrawing character is necessary to facilitate nucleophilic attack of the N-terminal amine and ring closure with the adjacent amide, but not so much as to favor hydration. Thus, it appears that electron withdrawing heteroatoms adjacent to the aldehydes, exemplified as pyridines, triazoles, imidazoles, and furans (Z, AA, BB, CC, and DD) promote the imidazolinone formation.

### Traceless Reversal of Imidazolinone Cap

The highly efficient peptide capture suggested that it might prove useful for preparation of single-molecule proteomics samples, in particular by allowing for more effective fluorescent labeling of peptides in preparation for single molecule sequencing using a new proteomics technique known as fluorosequencing. For this approach, peptides with fluorescently labeled amino acids are covalently tethered to a microscope cover slip and imaged at the single-molecule level using total internal reflection fluorescence microscopy (TIRF) following consecutive removals of N-terminal amino acids by Edman degradation. Edman cycles in which fluorescently labeled amino acids are removed reveal the amino acid sequence positions of the labeled amino acids, thus providing a peptide fingerprint on a molecule-by-molecule basis.

In order for the FPCA resin to be useful for fluorosequencing, the N-terminus of the peptide must be protected during the initial labeling of amino acids to minimize side reactions when Lys is labeled. Another requirement is that the N-terminus must be free to allow for Edman degradation. Very recently, the Francis lab showed that when peptides attached to a resin bound PCA are treated with hydroxylamine the peptides slowly are released into solution over 6–10 days<sup>26</sup> (Figure 2A). Concurrently, we had optimized a similar reaction and found it was amenable to analysis of peptides. To do this, we took inspiration from the field of Native Chemical Ligation<sup>27</sup> and its use of thiazolidine, a protected variant of Cys, where a formaldehyde is condensed with the N-terminus and side chain to form a 5-membered ring.

To reveal the Cys, it is common to utilize 300–400 mM methoxyamine at 37–42 °C, which quantitatively deprotects Cys in 3–4 hours<sup>28</sup>. With this as a starting point, we purified peptides that were N-terminally capped with either 4-nitrobenzaldehyde, PCA, or 3-formylisoquinoline. The status of the N-terminal cap was determined by <sup>1</sup>H-NMR (Supplemental Figure 2–4) and no presence of the imine proton could be detected. Next, the peptides were treated with 300 mM methoxyamine at pH 3 and 60 °C for 24 hours (Figure 2B). We observed a range of N-terminal deprotection levels, ranging from ~45% – 70% for the three different aldehydes tested. To optimize this reaction for speed and quantitative conversion, we utilized the more reactive dimethylaminoethyl hydrazine (DMAEH), which is extremely rapid in the formation of hydrazones<sup>29</sup>. After a 24-hour treatment of 300 mM DMAEH at pH 3 and 60 °C we observed nearly quantitative deprotection for all three aldehydes (Figure 2C). This deprotection condition was used for subsequent experiments.

### Solid-Phase Peptide Capture

Next, we moved on to use the N-terminal specific capture of peptides as a method for the efficient preparation of samples for mass spectrometry-based proteomics. To do this, we used the water swellable PEG amine resin and performed amide bond formation to generate a resin that contained an acid cleavable Rink linker to which we added 6-formyl picolinic acid (FPCA) (Figure 3A). A PCA derivative was used over one of the three derivatives found to perform more efficient capture as FPCA is available commercially and this will ensure this method can be used by the proteomics community more easily.

To evaluate the extent of capture and release by the aldehyde resin, we used angiotensin I as a test peptide. Capture of the peptide was determined by comparing the integrated peaks corresponding to the peptide during RP-HPLC analysis of the initial solution and the peptide resulting from the TFA cleavage. The captured peptide was released from the resin using a TFA cocktail (95 % TFA, 2.5 % H<sub>2</sub>O, 2.5% triisopropylsilane) to free the capped peptide. The resulting pellet was resuspended and analyzed with high-resolution mass spectrometry, and by comparing the input and cleavage sample we see an overall 70% yield of peptide. The capped peptide exists as diastereomers and the two peaks were found in the chromatogram (Figure 3C).

### Determining Bias of Resin Capture

While the FPCA modified resin is able to efficiently capture a single pure peptide, we wanted to ensure that there was no bias in regards to what N-terminal amino acids were

captured. In order to test a wide range of peptides, human embryonic kidney (HEK293T) cells were lysed by sonication, the proteins denatured by trifluoroethanol, and treated with iodoacetamide to alkylate all Cys. Next, this mixture was split to allow for replicate studies, added onto PEGA-PCA resin, and N-methylated trypsin was added to digest the protein mixture. This was incubated for 16 hours at 37 °C to perform simultaneous protein digestion and peptide capture. Afterwards the resin was washed extensively, and this step removed the small molecules (iodoacetamide and DTT), salts (phosphate buffer), and unbound trypsin. Finally, the peptides were released using the TFA cocktail, precipitated, and analyzed by high-resolution tandem mass spectrometry to identify the bound peptides.

Mass spectrometry analysis of the released peptides identified 42,171 unique bead-bound peptides common to both replicates. More than half of the peptides observed did not possess a PCA adduct, attributed to non-specific adsorption to the resin as is commonly seen in resin-based purification from total cell lysates<sup>30</sup>. To determine the bias between PCA-captured vs non-specifically adsorbed peptides, we sorted peptides by their N-terminal amino acid, and compared the relative frequencies of these found with an N-terminal imidazolinone cap. From this analysis, we found only minor bias for N-terminal Ala (~2.5x) and against Met (~3x) (Figure 4). This is in agreement with solution-phase studies. Additionally, we found no peptides with a Pro at the second position, due to the fact that Pro contains a tertiary amide and is incapable of undergoing imidazolinone formation. We would like to point out that there are a number of reversible side products that can form (thioacetals forming with Cys residues for example), however these are minimized due to the irreversible nature of the imidazolinone product, effectively trapping the peptide as the desired product.

### Integration into Fluorosequencing

With a working capture and reversal protocol in place, we turned to applying this technique for fluorosequencing. We considered a scenario in which peptides would be captured by their N-termini on resin, their C-termini modified with an alkynyl linker, one or more of their internal amino acids fluorescently labeled, and then peptides released. This scenario would test a number of important chemical steps and the ability to remove free fluorescent dyes to a level they cannot be detected by single molecule proteomics. To test if such a strategy was feasible with the PCA resin, we captured a short peptide (H<sub>2</sub>N-Ala-Lys-Ala-Gly-Ala-Gly-Arg-Tyr-Gly-COOH) onto the resin. Note that this peptide lacks acidic residues to avoid double labeling them with the C-terminus. Peptides from native sources would require C-terminal differentiation or be generated by a protease such as GluC whose action leaves all acidic residues positioned at peptide C-termini. We labeled the C-terminus with propargyl amine using HCTU/DIEA chemistry in DMF. Next, the Lys was labeled using Atto647N NHS esters in DMF. Finally, the peptide was cleaved from the resin using the TFA cocktail, precipitated, and the pellet was allowed to air dry.

Fresh azide-silane slides were prepared suitable for fluorosequencing, and the fluorescent peptide attached to their surface by Cu(I)-Click (Figure 5A). The slide was washed extensively in ddH<sub>2</sub>O, and the PCA deprotected by submerging the slide in a bath of 500 mM DMAEH for 16 hours at 60 °C. Next, the slide was washed in water and assembled into the flow cell of the TIRF microscope and subjected to fluorosequencing as previously

described. We observed a strong loss of fluorescence after the second Edman cycle in 50% of the molecules, indicating that the Lys-Atto647N was removed at the appropriate Edman cycle as expected (Figure 5 B-D). The remaining peptide losses are accounted for by errors produced during the physico-chemical processes such as photobleaching, Edman inefficiency etc., and are similar to the results obtained with Fmoc protected peptides. These errors have been previously described and modeled<sup>7</sup>. Thus, this capture/label/release strategy offers a novel and elegant solution for proteomics experiments.

## Conclusion

Here, we have shown that the ability to covalently bind peptides from a complex mixture on a solid phase substrate, anchoring them by their N-termini and allowing for their efficient purification and handling for multiple types of proteomics experiments. Further, we show this method allows for the removal of salts, small molecules, and other unbound reagents used to prepare samples. We expect this methodology should also allow removal of non-peptide contaminants in the sample such as detergents.

Importantly, binding the peptides to the resin allows for the use of a number of organic solvents, as the PEG resin is compatible with many solvents that peptides generally are not fully soluble in. We utilized this feature to chemically modify the C-termini of peptides bound to the resin using standard peptide synthesis techniques and fluorescently label Lys residues of bound peptides using an NHS ester in DMF and organic bases.

For experimental techniques that require multiple chemical reactions such as fluorosequencing or nanopore sequencing<sup>31</sup>, we expect the resin will prove valuable for the analysis of proteomic samples with small sample sizes. Working with smaller protein concentrations will be important for these single-molecule techniques and helps advance proteomic techniques to be better able to analyze limiting samples, as for clinical biopsies.

## METHODS

### Materials:

Chemicals were purchased and used as received without any purification. 6-formyl picolinic acid was purchased from Enamine. Atto647N-NHS was purchased from Attotek. Azide functionalized slides were purchased from PolyAn. All other chemicals were purchased from Sigma Aldrich.

### Peptide Synthesis:

All peptides were synthesized using automated microwave-assisted solid-phase peptide synthesis (Liberty Blue Microwave Synthesizer, CEM Corporation). Synthesis was performed using standard Fmoc chemistry using DIC/Oxyma coupling strategies (1:1:1 ratio with amino acids). Coupling steps were performed at 90 °C for 120 seconds, and deprotection was performed using 20 % piperidine in DMF at 90 °C for 60 seconds. All peptides were cleaved from resin using trifluoroacetic acid (TFA), triisopropylsilane (TIS), and H<sub>2</sub>O (95:2.5:2.5) for 2.5 hours prior to the cleavage solution being concentrated under nitrogen stream. The peptide is precipitated with ice cold diethyl ether and collected by



centrifugation at 12,000 *g* for 10 minutes. Peptides were purified using a Grace-Vydac C18 column (Buffer A: H<sub>2</sub>O + 0.1 % formic acid; Buffer B: methanol + 0.1 % formic acid) over a 10–60 % gradient.

#### **In-solution N-terminal Peptide Capture:**

Peptides were mixed with four molar equivalents of aldehyde in 50 mM phosphate buffer pH 7.5, then incubated for 8–16 hours at 37 °C prior to purification or analysis. All aldehydes used were solubilized in DMF at 100 mM then diluted to final concentration. Samples were analysed by LC/MS. Amino formation was determined by quantitation of unreacted peptides at the zero hour and end point of the reaction using HPLC by measuring the area under the curve.

#### **Reversal of Amino Cap:**

Peptides were first allowed to react with 4-nitrobenzaldehyde, 2-pyridinylcarboxaldehyde, or 3-formylisoquinoline following standard in-solution reaction procedures (4 mM aldehyde and 1 mM peptide). Peptides were then purified using a Grace-Vydac C18 RP-HPLC column, analysed by LC/MS and lyophilized to dryness. Complete imidazolone conversion was checked by <sup>1</sup>H-NMR (MR400 MHz, Agilent) For the reversal tests, capped peptides were resuspended in either 0.3 M dimethylaminoethyl hydrazine or 0.3 M methoxyamine. Samples were incubated at 60 °C and then analysed by HPLC and mass spectrometry at each time point. Percentage released was determined by comparing the integration of the HPLC peak of the capped peptide over time.

#### **Aldehyde Capture Resin Preparation:**

Amino PEGA resin (Novabiochem) was used and was functionalized with Fmoc-Peg<sub>2</sub>-OH, Rink linker and 6-formylpyridine-2-carboxylic acid using HCTU/DIEA (1:1:1.1 ratio) chemistry coupling for 45 minutes. Deprotection was done using 20 % piperidine in DMF two times for five minutes each. Resin was stored in DMF at 4 °C prior to use.

#### **Resin-based Peptide Capture:**

Capture resin was washed in DMF, water, and 50 mM phosphate buffer pH 7.5. Each wash included a 5 minute incubation in the solvent. Peptide was then added to the resin in 50 mM phosphate buffer pH 7.5 and incubated at 37 °C for 16–24 hours. Next the resin was washed extensively in incubation buffer, water, and finally DMF. After derivatization, the resin was washed extensively in water, DMF, and finally DCM. The peptide was cleaved from resin in 95 % TFA, 2.5 % TIS, and 2.5 % H<sub>2</sub>O. The TFA was concentrated under N<sub>2</sub> stream and ether precipitated prior to mass spectrometry analysis.

#### **Cell Growth Conditions:**

HEK-293T cells were grown in Dulbecco's Modified Eagle Medium with 10 % Fetal Bovine Serum at 37 °C and 5 % CO<sub>2</sub>. Cells were passaged when between 70–80 % confluence.

### HEK Lysate Digestion and Capture:

Cells were grown to 80 % confluence and harvested in PBS and pelleted at 500 *g* for 3 minutes. Cells then suspended in hypotonic 50 mM Tris-HCl buffer pH 8 and placed on ice. Protease inhibitor (Mini cOmplete, EDTA Free protease inhibitor cocktail, Roche) was added to 1x concentration. Cells were sonicated (Branson 2510) for 1 minute at 42 kHz and placed on ice for an additional minute. This was repeated 3 times. The solution was then centrifuged at 17,000 *g* for 10 minutes at 4 °C and the supernatant was collected. Protein content was then measured using a Bradford Assay. 250 µg of protein was denatured in 2,2,2-trifluoroethanol (TFE) and 5 mM tris(2-carboxyethyl)phosphine (TCEP) at 45 °C for 45 minutes. Proteins were alkylated in the dark with 5.5 mM iodoacetamide, then the remaining iodoacetamide quenched in 100 mM dithiothreitol. MS-grade reductively methylated trypsin (Pierce) was then added to the solution in a ratio of 1:25. This mixture was added to PEGA-FPCA beads and diluted with 100 mM phosphate buffer pH 7.5. The peptide mixture was incubated at 37 °C for 18 hours. Next, the resin was washed extensively with H<sub>2</sub>O, DMF and DCM prior to cleaving the resin with a TFA cocktail (95 % trifluoroacetic acid, 2.5 % triisopropylsilane, and 2.5 % H<sub>2</sub>O) for 2.5 hours. The peptides were precipitated with ice cold ether and allowed to air dry.

### Mass Spectrometry:

Peptides were separated on a 75 µM x 25 cm Acclaim PepMap100 C-18 column (Thermo Scientific) using a 3–45 % acetonitrile + 0.1 % formic acid gradient over 120 min and analysed online by nanoelectrospray-ionization tandem mass spectrometry on an Orbitrap Fusion (Thermo Scientific). Data-dependent acquisition was activated, with parent ion (MS1) scans collected at 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 dataset identifier PXD016291 and [10.6019/PXD016291](https://doi.org/10.6019/PXD016291).

### Peptide Identification:

Peptides were identified with Proteome Discoverer 2.3 (Thermo Scientific), searching against the Uniprot human reference proteome. To distinguish PCA protected peptides, searches incorporated an optional peptide N-terminal dynamic modification (132.032 Da) corresponding to the PCA modified peptide, thresholding identifications with a false discovery rate of 1 %. A total of 133,793 tandem mass spectra were collected and assigned, corresponding to 39,581 (from the bound sample) and 25,049 (from the flowthrough) in replicate 1, and 43,986 (from bound) and 25,177 (from flowthrough) in replicate 2.

### On Bead Labeling of Peptides:

Peptides were captured on PCA resin as described. After rinsing, the C-terminus was first coupled with 100 mM propargylamine, 100 mM HCTU, and 100 mM DIEA in DMF for 2 hours at room temperature. The resin was extensively washed with DMF and the Lys



residues were labelled with 0.5 mM Atto647N-NHS (Attotec) and 1 mM DIEA overnight at room temperature. The resin was washed extensively in DMF and DCM and all of the peptides cleaved from the resin with a TFA cocktail (95 % TFA, 2.5 % H<sub>2</sub>O, and 2.5 % TIS) for 2.5 hours. The supernatant was collected and concentrated with a N<sub>2</sub> stream. Ice cold diethyl ether is added (10 vol) and the peptides collected by centrifugation for 10 minutes at 17,000 *g*. The peptide was analysed by high-resolution mass spectrometry to confirm the double labelling.

### Single Molecule Peptide Sequencing:

Approximately 200 pM of peptides were immobilized on an azide-silane slide (custom slides from PolyAn, Germany) using standard Cu(I)-Click chemistry. Briefly a 2 mL solution comprising peptide (200 pM), CuSO<sub>4</sub>/tris-hydroxypropyltriazolylmethylamine (THPTA) mix (1 mM/0.5 mM) and freshly prepared sodium L-ascorbate (5 mM) was incubated on the azide-silane slide at room temperature for 2 hours. Following the incubation, the slides were rinsed with water and fluorosequencing performed as previously described<sup>7</sup> with minor modifications. To deprotect the N-terminal PCA cap, the slides were bathed in 0.5 M DMAEH at 60 °C for 16 hours. Deprotection of the Fmoc group was performed by incubating slides with 20% Piperidine solution in DMF for 1 hour. The images were processed using custom developed scripts, available at <https://github.com/marcottelab/FluorosequencingImageAnalysis/>.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

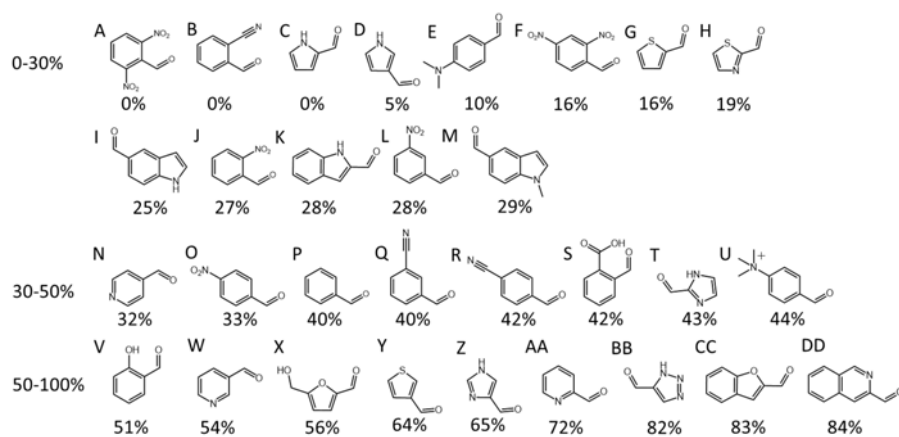
This work was supported by grants from DARPA (N66001-14-2-4051 to EVA and EMM), NIH (R35 GM122480, R01 HD085901, R01 DK110520 to EMM), Erisyon, Inc. (to EMM and EVA), and the Welch foundation (F-1515 to EMM and F-0046 to EVA). JS, EMM, AB and EVA are cofounders and shareholders of Erisyon, Inc. A patent application based on aspects of this work is pending.

### References

1. Hwang H; Jeong JE; Lee HK; Yun KN; An HJ; Lee B; Paik YK; Jeong TS; Yee GT; Kim JY; Yoo JS, Identification of Missing Proteins in Human Olfactory Epithelial Tissue by Liquid Chromatography-Tandem Mass Spectrometry. *J Proteome Res* 2018, 7 (17), 4320–4324.
2. MacDonald JI; Munch HK; Moore T; Francis MB, One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nature Chemical Biology* 2015, 11, 326–331. [PubMed: 25822913]
3. Schwammle V; Aspalter CM; Sidoli S; Jensen ON, Large Scale Analysis of Co-existing Post-translational Modifications in Histone Tails Reveals Global Fine Structure of Cross-talk. *Molecular & Cellular Proteomics* 2014, 13 (7), 1855–1865. [PubMed: 24741113]
4. Lin S; Garcia BA, Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. 2012, 512, 3–28.
5. Movileanu L; Schmittschmitt JP; Scholtz JM; Bayley H, Interactions of peptides with a protein pore. *Biophys J* 2005, 89 (2), 1030–45. [PubMed: 15923222]
6. van Ginkel J; Filius M; Szczepaniak M; Tulinski P; Meyer AS; Joo C, Single-molecule peptide fingerprinting. *Proceedings of the National Academy of Sciences of the United States of America* 2018, 115 (13), 3338–3343. [PubMed: 29531063]

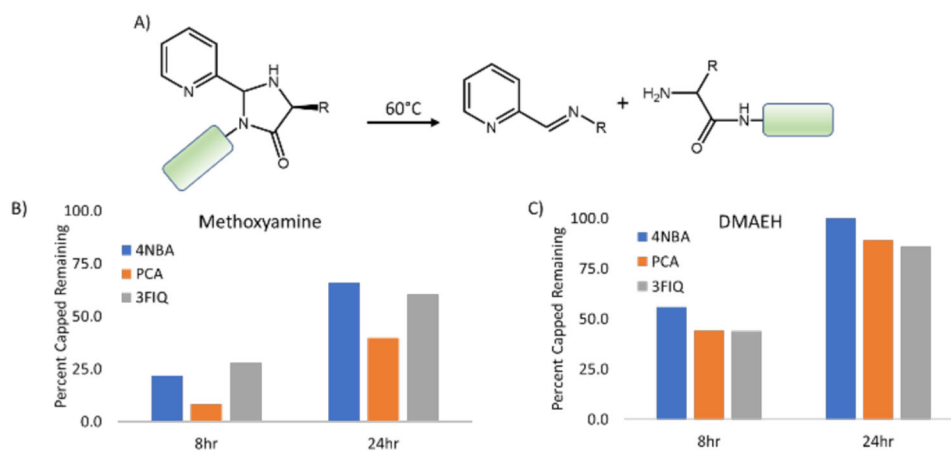
7. Swaminathan J; Boulgakov AA; Hernandez ET; Bardo AM; Bachman JL; Marotta J; Johnson AM; Anslyn EV; Marcotte EM, Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nature Biotechnology* 2018, 36, 1076–1082.
8. Ohayon S; Girsault A; Nasser M; Shen-Orr S; Meller A, Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. *PLoS Comput Biol* 2019, 15 (5), e1007067. [PubMed: 31145734]
9. Swaminathan J; Boulgakov AA; Marcotte EM, A theoretical justification for single molecule peptide sequencing. *PLoS Comput Biol* 2015, 11 (2), e1004080. [PubMed: 25714988]
10. Yao Y; Docter M; van Ginkel J; de Ridder D; Joo C, Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys Biol* 2015, 12 (5), 055003. [PubMed: 26266455]
11. Lee CH, A Simple Outline of Methods for Protein Isolation and Purification. *Endocrinol Metab (Seoul)* 2017, 32 (1), 18–22. [PubMed: 28256117]
12. Steen H; Jebanathirajah JA; Rush J; Morrice N; Kirschner MW, Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics* 2006, 5 (1), 172–81. [PubMed: 16204703]
13. Lang K; Chin JW, Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* 2014, 114 (9), 4764–4806. [PubMed: 24655057]
14. Batth T; Tollenaere M; Ruther P; Gonzalez-Franquesa A; Prabhakar B; Bekker-Jensen S; Deshmukh A; Olsen JV, Protein aggregation capture on microparticles enables multi-purpose proteomics sample preparation. *Molecular & Cellular Proteomics* 2019, 18 (5), 1027–1035. [PubMed: 30833379]
15. Dunn JD; Reid GE; Bruening ML, Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrometry Reviews* 2009, 29–54.
16. Lin X; Yu AC; Chan TF, Efforts and Challenges in Engineering the Genetic Code. *Life (Basel)* 2017, 7 (1).
17. Merrifield RB, Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc* 1963, 85 (14), 2149–2154.
18. Berg B. H. J. v. d.; Tholey A, Mass spectrometry - based proteomics strategies for protease cleavage site identification. *PROTEOMICS* 2012, 12 (4 - 5), 516–529. [PubMed: 22246699]
19. Oregioni A; Stieglitz B; Kelly G; Rittinger K; Frenkiel T, Determination of the pKa of the N-terminal amino group of ubiquitin by NMR. *Sci Rep* 2017, 7, 43748. [PubMed: 28252051]
20. Kitanaka A; Juichi H; Nihashi Y; Miyashita M; Miyagawa H, A Facile Method for Preferential Modification of the N-Terminal Amino Group of Peptides Using Triazine-Based Coupling Reagents. *Mass Spectrometry* 2017, 6 (1), A0059. [PubMed: 28785530]
21. Gilmore JM; Scheck RA; Esser - Kahn AP; Joshi NS; Francis MB, N - Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angewandte Chemie International Edition* 2006, 45 (32), 5307–5311. [PubMed: 16847857]
22. Grimsley GR; Scholtz JM; Pace CN, A summary of the measured pK values of the ionizable groups in folded proteins. *Protein science : a publication of the Protein Society* 2009, 18 (1), 247–251. [PubMed: 19177368]
23. Villain M; Vizzavona J; Rose K, Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. *Chemistry & Biology* 2001, 8 (7), 673–679. [PubMed: 11451668]
24. Jin Y; Yu C; Denman RJ; Zhang W, Recent advances in dynamic covalent chemistry. *Chem Soc Rev* 2013, 42 (16), 6634–54. [PubMed: 23749182]
25. Dagna JM; Pescitelli G; Tran L; Lynch VM; Anslyn EV; Di Bari L, In Situ Assembly of Octahedral Fe(II) Complexes for the Enantiomeric Excess Determination of Chiral Amines Using Circular Dichroism Spectroscopy. *Journal of the American Chemical Society* 2012, 134 (9), 4398–4407. [PubMed: 22272943]
26. Koo B; Dolan NS; Wucherer K; Munch HK; Francis MB, Site-Selective Protein Immobilization on Polymeric Supports through N-Terminal Imidazolidinone Formation. *Biomacromolecules* 2019, 20 (10), 3933–3939. [PubMed: 31448594]
27. Saiz C; Wipf P; Manta E; Mahler G, Reversible thiazolidine exchange: a new reaction suitable for dynamic combinatorial chemistry. *Organic letters* 2009, 11 (15), 3170–3173. [PubMed: 19585999]

28. Shimko JC; Howard CJ; Poirier MG; Ottesen JJ, Preparing Semisynthetic and Fully Synthetic Histones H3 and H4 to Modify the Nucleosome Core. *Methods in Molecular Biology* 2013, 981, 177–192. [PubMed: 23381862]
29. Kool ET; Crisalli P; Chan KM, Fast Alpha Nucleophiles: Structures that Undergo Rapid Hydrazone/Oxime Formation at Neutral pH. *Organic Letters* 2014, 16 (5), 1454–1457. [PubMed: 24559274]
30. Mellacheruvu D; Wright Z; Couzens AL; Lambert JP; St-Denis NA; Li T; Miteva YV; Hauri S; Sardu ME; Low TY; Halim VA; Bagshaw RD; Hubner NC; Al-Hakim A; Bouchard A; Faubert D; Fermin D; Dunham WH; Goudreault M; Lin ZY; Badillo BG; Pawson T; Durocher D; Coulombe B; Aebersold R; Superti-Furga G; Colinge J; Heck AJ; Choi H; Gstaiger M; Mohammed S; Cristea IM; Bennett KL; Washburn MP; Raught B; Ewing RM; Gingras AC; Nesvizhskii AI, The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* 2013, 10 (8), 730–6. [PubMed: 23921808]
31. Restrepo-Perez L; Joo C; Dekker C, Paving the way to single-molecule protein sequencing. *Nat Nanotechnol* 2018, 13 (9), 786–796. [PubMed: 30190617]

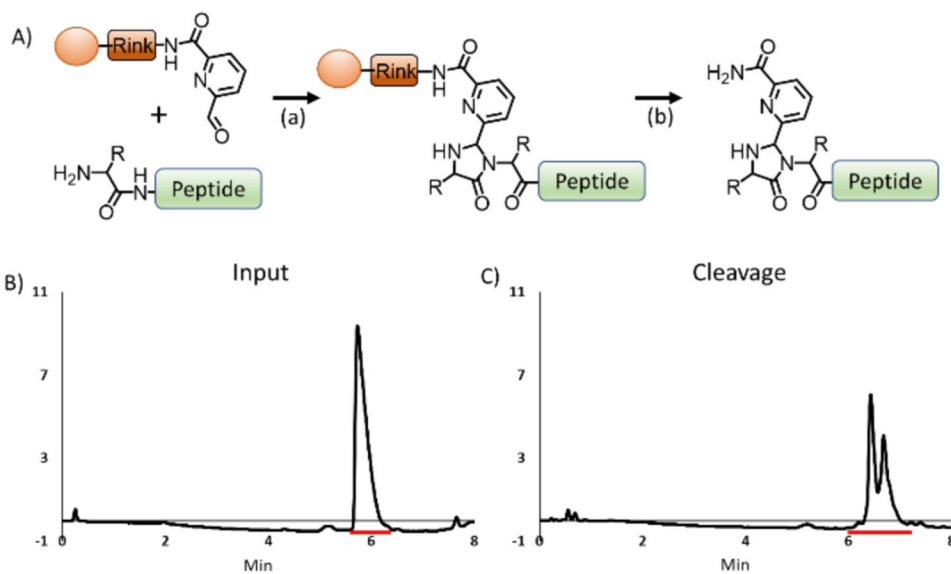


**Figure 1.**

A-DD) are the aldehydes screened to find compounds that undergo aminal formation with the N-terminus of a Ser-Gly-Trp peptide. Each shows the structure and the percent aminal formed based on the quantitation of the area under the curve from the HPLC of the reaction. Each reaction was performed in triplicate.



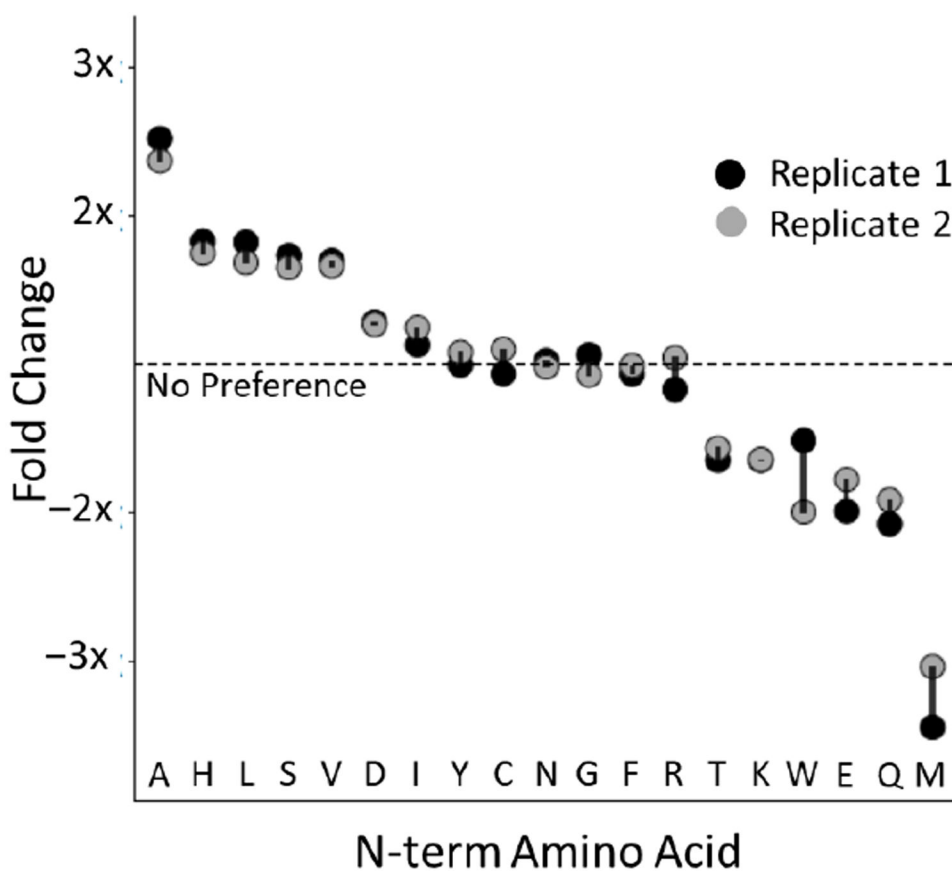
**Figure 2.** Reversal tests of N-terminally imidazolinone capped SGW peptide with 4-nitrobenzaldehyde (4NBA), 2PCA, or FIQ using either 0.3 M methoxyamine or 0.3 M dimethylaminoethyl hydrazine.



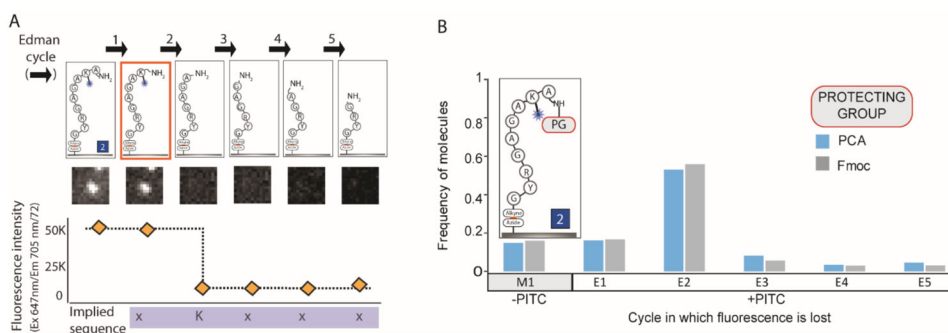
**Figure 3.**

A) Schematic showing the PEG-Rink-FPCA resin synthesized and the steps for coupling and releasing peptides: (a) peptide in 50 mM sodium phosphate pH 7.5 is added to the resin and incubated at 37 °C for 16 hours, (b) the peptide is liberated from the resin using 95% trifluoroacetic acid, 2.5% H<sub>2</sub>O, and 2.5% triisopropyl silane for 2.5 hours B-C) HPLC of angiotensin I input (280 nm trace) (A) and TFA cleavage (B) after capture on PEG-Rink-FPCA resin. Red line indicates area under curve used to quantitate percent of peptide captured.

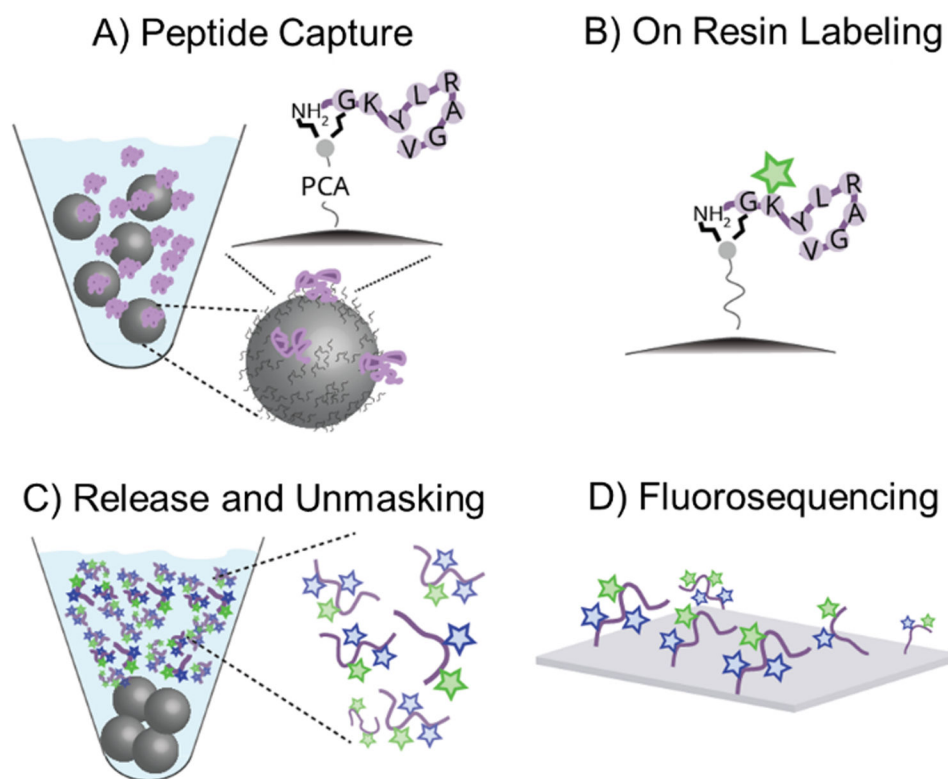




**Figure 4.** Bias of the PCA capture reagent as a function of differing N-terminal amino acids of captured peptides. Across 42,171 unique peptides identified with high confidence (FDR < 1 %) in both of two replicate HEK293T cell lysate proteomics experiments, we measured the fold enrichment in PCA-specific vs non-specifically bound peptides, based on a total of 59,612 tandem mass spectra (peptide-spectral matches) over the two replicates.

**Figure 5.**

A) Representative field of view of individual fluorescent peptide molecules from a fluorosequencing experiment. B) Light emitted from one peptide molecule across Edman cycles (labeled in panel D), with its subsequent loss of fluorescence after cycle 2. C) Measured fluorescence intensities of the peptide molecule in panel B. D) The frequency of molecules exhibiting a loss of fluorescence after each Edman cycle is plotted for both a PCA or Fmoc protected peptide, showing data from 103-671 and 59-720 molecules, respectively. M1 is a "mock" cycle where the slide is washed with all reagents used in fluorosequencing without the phenyl isothiocyanate (PITC) being present to remove non-specifically bound peptides or free fluorophores. E1-E5 represent consecutive Edman cycles.

**Scheme 1.**

Overall scheme of using an N-terminal capture resin for fluorosequencing. A) N-terminal capture using a pyridinyl carboxaldehyde derivative attached to a resin<sup>2</sup>.