Supporting information

Generating Cysteine-Trypsin Cleavage Sites with 2-Chloroacetamidine Capping

Samuel Ofori,[1] Heta Desai, [1,6] Flowreen Shikwana,[1,2 a] Lisa Boatner, [1,2] Emil Dominguez, [1,2] José O. Castellón[1,6] and Keriann M. Backus *[1,2,3,4,5,6]

1. Biological Chemistry Department, David Geffen School of Medicine, UCLA, Los Angeles, CA, 90095, United States

2. Department of Chemistry and Biochemistry, UCLA, CA, 90095, United States

3. DOE Institute for Genomics and Proteomics, UCLA, Los Angeles, CA 90095 (USA);

4. Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA 90095 (USA);

5. Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, UCLA, Los Angeles, CA 90095 (USA).

6. Molecular Biology Institute, UCLA, Los Angeles, California 90095, United States

* Corresponding Author: Keriann M. Backus, Biological Chemistry Department, David Geffen School of Medicine, UCLA, Los Angeles, CA, 90095, USA, E-mail: kbackus@mednet.ucla.edu.

Table of Contents

(A) Supplementary Figures

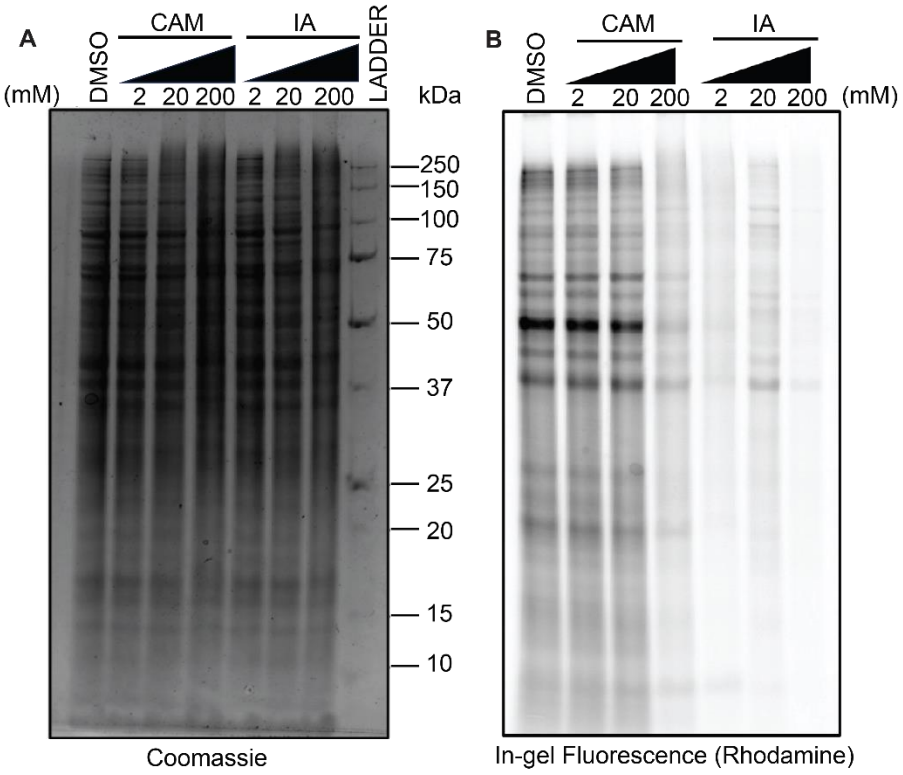

Figure S1. Full-length gels related to Figure 1 showing the in-gel-based concentration-dependent analysis of the various probes (2-chloroamidine – CAM; Iodoacetamide – IA) in cell lysates. HCT-15 cell lysates were treated with a dose range as specified on the gel (in mM) for 1 h., followed by IA-rhodamine (7 µM) for 1 h., quenching, and SDS-PAGE analysis. A) Coomassie InstantBlue visualization of protein loading B) Full-length gel for In-gel fluorescence (Rhodamine).

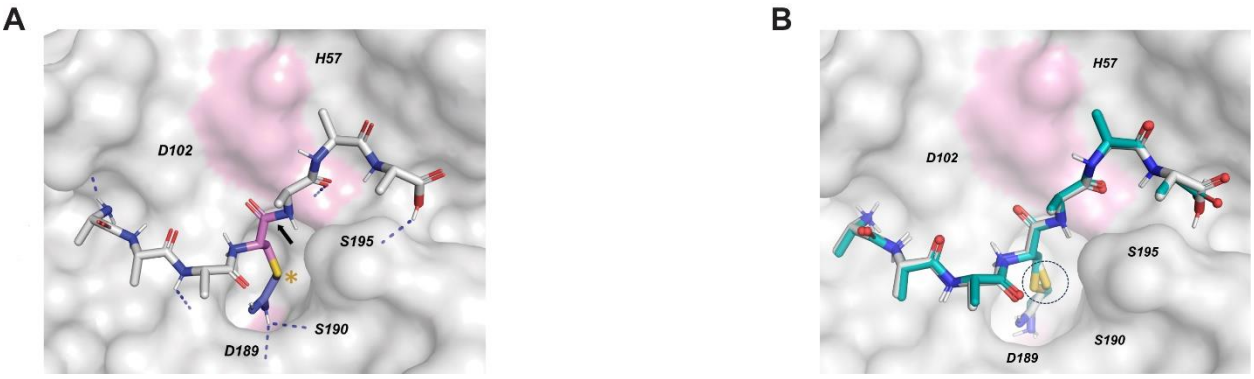

Figure S2. A) In silico docking of a CAM-capped peptide, AAA-CAM-AAA, near the catalytic triad of bovine trypsin (PDB: 1AZ8) B) Structural overlay of the highest scoring docking poses of CAM-capped peptide (AAA-CAM-AAA, -7.6 kcal/mol) and BEA-capped peptide (AAA-C(BEA)-AAA, -7.2 kcal/mol) near the catalytic triad of bovine trypsin (PDB: 1AZ8). Capping with BEA results in

a repositioning of the AAA-CEA-AAA resulting in a less favourable interaction with the key aspartate 189.

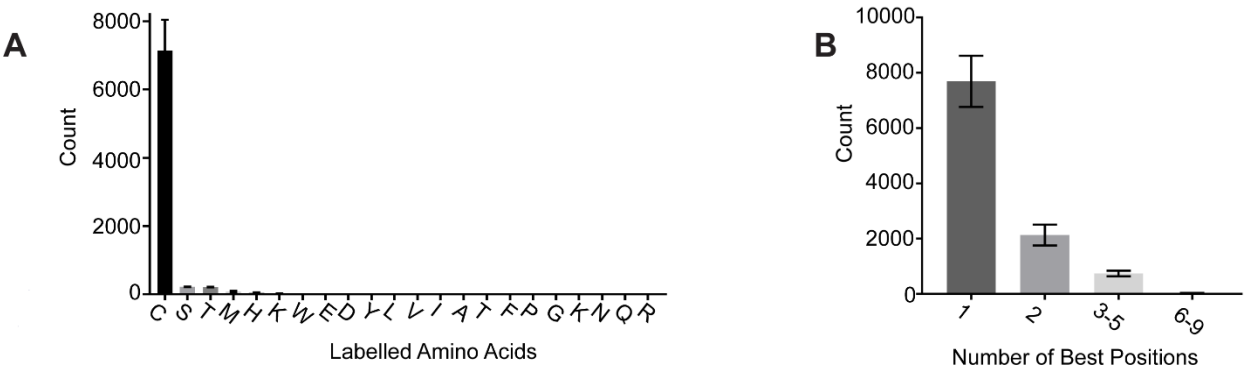

Figure S3. Mass offset search results using FragPipe[1-2]. A) Selectivity of CAM to Cysteines over other Amino Acids. B) Number of best positions for CAM localization. Data and analysis for both selectivity of CAMs to amino acids as well as the number of best position counts were obtained from FragPipe-based mass offset search. Details have been provided in the **Mass Offset Search Analysis for CAM Selectivity** section. All data is available in **Table S2**.

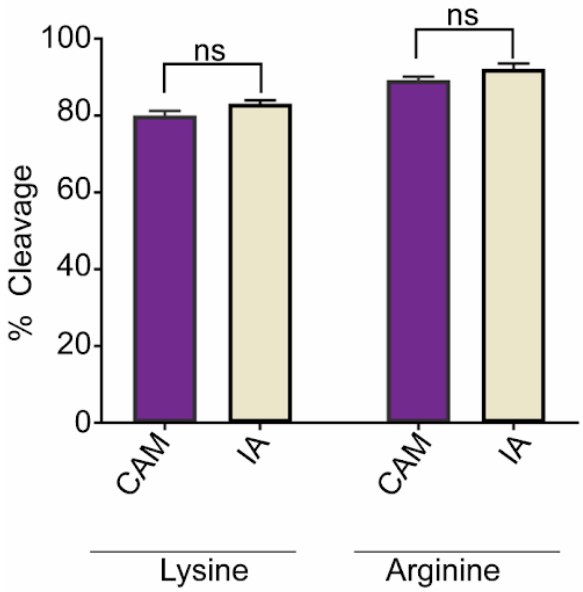

Figure S4. Cleavage at lysines and arginines. Percentage of Lysines (average n= 1648/2063 for CAM; n= 2916/3507 for IA) and Arginines (n= 1514/1697 for CAM; n= 2930/3193 for IA) Cleaved in CAM vs IA samples (n = 3 replicates). Statistics w/ student's t-test ****p<0.0001,ns p > 0.05. ns p > 0.05. All data is available in **Table S1**.

Figure S5. Selected examples of inferred cysteine cleavage sites. The identified peptides are shown in cyan and the P1 cysteines in magenta with the upstream tryptic peptide sequence in black. All data is available in data **Table S1.**

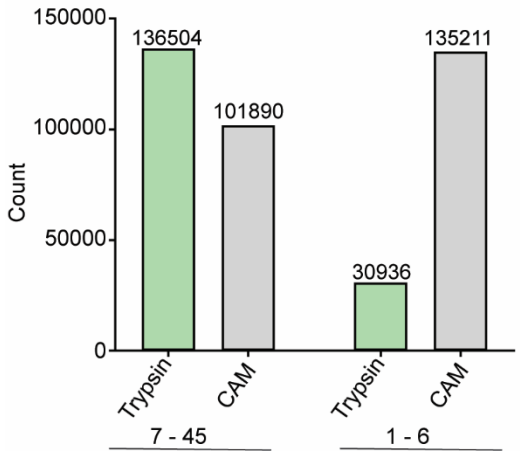

Figure S6. In silico digestion of the human reference proteome. CAM capping both increases total number of cysteine peptides and the number of short 1-6mer cysteine peptides. All data is available in **Table S5**.

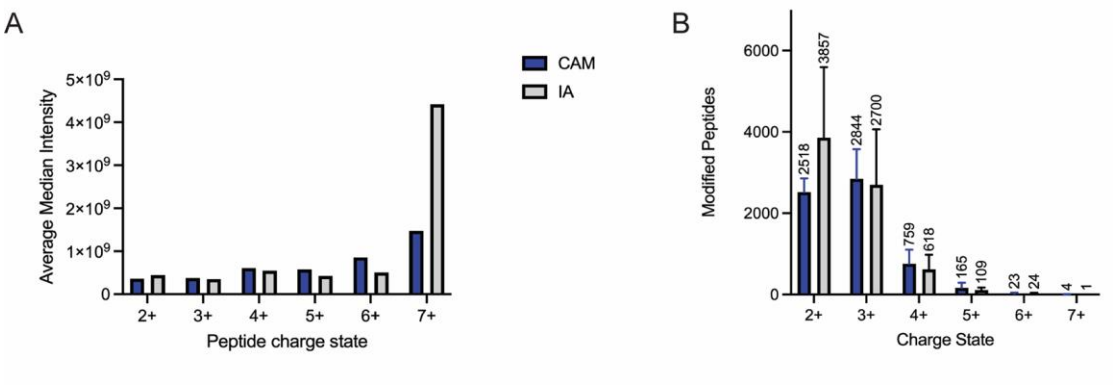

Figure S7. Charge State Analysis. A) Average Median Intensity and B) modified peptide counts. All data is available in **Table S3**.

(B) Supplementary Tables

Table S6. Files in Proteomics Identification Database (PRIDE) dataset (PXD047844)

Figure	Filename	Sample ID	Experiment	Fragpipe Analysis
Figs. 2 - 4	2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f1 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f2 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f3 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f4 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f5 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f6 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f7 2023_04_10_KB_noFAIMS_SO_1A_HighpH_II_f8	**CAM sample 1** (fractions 1 - 8)	Cyscapping*	LFQ, semi-tryptic

Figs. 2 - 4	2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f1 2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f2 2023_06_07_KB_NoFAIMS_SO_2A_fraction3_rerun 2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f4 2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f5 2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f6 2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f7 2023_04_21_KB_noFAIMS_SO_1B_HighpH_II_f8	**CAM sample 2** (fractions 1 - 8)	Cyscapping*	LFQ, semi-tryptic
Figs. 2 - 4	2023_06_01_KB_NoFAIMS_SO_3A_fraction1 2023_06_01_KB_NoFAIMS_SO_3A_fraction2 2023_06_01_KB_NoFAIMS_SO_3A_fraction3 2023_06_01_KB_NoFAIMS_SO_3A_fraction4 2023_06_01_KB_NoFAIMS_SO_3A_fraction5 2023_06_01_KB_NoFAIMS_SO_3A_fraction6 2023_06_01_KB_NoFAIMS_SO_3A_fraction7 2023_06_01_KB_NoFAIMS_SO_3A_fraction8	**CAM sample 3** (fractions 1 - 8)	Cyscapping	LFQ, semi-tryptic
Figs. 2	2023_05_18_KB_noFAIMS_SO_2A	**IA sample 1**	Cyscapping*	LFQ, semi-

- 4	_20mMIA_HighpH_II_f1 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f2 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f3 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f4 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f5 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f6 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f7 2023_05_18_KB_noFAIMS_SO_2A_20mMIA_HighpH_II_f8	(fractions 1 - 8)		tryptic
Figs. 2 - 4	2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f1 2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f2 2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f3 2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f4 2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f5 2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f6	**IA sample 2** (fractions 1 - 8)	Cyscapping*	LFQ, semi-tryptic
	2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f7 2023_04_21_KB_noFAIMS_SO_2B_20mMIA_HighpH_II_f8			
Figs. 2 - 4	2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction1 2023_06_01_KB_NoFAIMS_SO_3	**IA sample 3** fraction 1	Cyscapping*	LFQ, semi-tryptic

	B_20mM_IAfraction2 2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction3 2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction4 2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction5 2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction6 2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction7 2023_06_01_KB_NoFAIMS_SO_3B_20mM_IAfraction8			

***** Cyscapping is defined here as the various experimentations of "proteomic sample preparation and processing for unenriched sample analysis" - as described in section (D).

Table S7. Files in Proteomics Identification Database (PRIDE) dataset (PXD052935)

Figure	**Filename**	**Sample ID**	**Experiment**	**Fragpipe Analysis**
Fig. 2	2024-05-31-KB-NoFAIMS-SO-BEA-sample-250mM-1	**BEA sample 1**	Cyscapping*	LFQ, semi-tryptic
Fig. 2	2024-05-31-KB-NoFAIMS-SO-BEA-sample-250mM-2	**BEA sample 2**	Cyscapping*	LFQ, semi-tryptic
Fig. 2	2024-05-31-KB-NoFAIMS-SO-BEA-sample-250mM-3	**BEA sample 3**	Cyscapping*	LFQ, semi-tryptic

Table S8. Preparation of elution buffer solutions for high-pH reversed-phase fractionation of unlabeled, native peptides

Fraction	Triethylamine (TEA) / µL	Acetonitrile (MeCN) / µL	TEA : MeCN (%)	Total Volume (µL)
1	950	50	95 : 5	1000
2	925	75	92.5 : 7.5	1000
3	900	100	90 : 10	1000
4	875	125	87.5 : 12.5	1000
5	850	150	85 : 15	1000
6	825	175	82.5 : 17.5	1000
7	800	200	80 : 20	1000
8	500	500	50 : 50	1000

Table S9. Solvent Systems and Flow Rates used for LC-MS/MS analysis.

Fractions (1-3)				**Fraction 4-7**			
Time	*Duration*	*Flow [nl/min]*	*% B*	*Time*	*Duration*	*Flow [nl/min]*	*% B*
0	00:00	300	1	0	00:00	300	1
5	05:00	220	15	3	03:00	220	15
70	65:00	220	35	63	60:00	220	45
76	06:00	250	95	73	10:00	220	55
80	04:00	250	95	74	01:00	250	95
—			—	80	06:00	250	95

Fraction 8			
Time	*Duration*	*Flow [nl/min]*	*% B*

0	00:00	300	1
3	03:00	220	15
43	40:00	220	40
68	25:00	220	65
70	02:00	250	95
80	10:00	250	95

(Buffer A: water with 3% DMSO and 0.1% formic acid and Buffer B: 80% acetonitrile with 3% DMSO and 0.1% formic acid)

(C) In Silico Studies of Trypsin-CAM Interaction

Protein Structure Preparation:

A Bovine Trypsin complexed with Bis-Phenylamidine Inhibitor (PDB: **1AZ8**) was downloaded from the Protein Data Bank[3]. First, the bis-phenylamidine, heteroatoms, and water molecules were removed from the trypsin protein complex. Next, the protein was prepared by removing water molecules, salts, and crystallographic additives. Then, MGLTools[4] was utilized to add hydrogens and compute Gasteiger-Marsili charges.

Docking of AACAM:

A prepared library of small molecules consisting of the Bis-phenylalanine (control; extracted from the parent protein complex), alanine-alanine-alanine-cysteine-alanine-alanine-alanine (AAACAAA), alanine-alanine-alanine-cysteine-capped-amidine-alanine-alanine-alanine- (AAA-CAM-AAA) and alanine-alanine-alanine-cysteine-capped-ethylamine-alanine-alanine-alanine- (AAA-CEA-AAA) was used for the in silico docking against the prepared/ minimized 1AZ8 protein. The docking studies and calculation of binding energies were performed using AutoDock Vina (version 1.1.2, rigid non-covalent docking)[5]. The center point was a gridbox 20 Å x 20 Å x 20 Å around HIS A 57 CE1 (the catalytic center) of the 1AZ8 protein – with the following dimension in Å: center (X, Y, Z) = (23.966, 13.522, 22.137) with an exhaustiveness of 8.

(D) Biology Methods

Cell lines, culture conditions

HCT-15 cells were cultured in Roswell Park Memorial Institute (Gibco™ PRMI 1640 Medium, 11875119) media supplemented with 10% FBS (Avantor®, Seradigm, Premium Grade Fetal Bovine Serum, Cat. No. 97068-085) and 1% Pen-Strep (Gibco™, Penicillin-Streptomycin 10,000 U/mL, 15140122) at 37℃, and 5% CO2. Cell lines were obtained from ATCC (TIB-152). All protein concentrations were determined using a Bio-Rad DC protein assay kit using reagents from Bio-Rad Life Science (Hercules, CA). Mycoplasma contamination testing in cell cultures was performed using the MycoAlert® kit (Lonza Rockland, Rockland, ME) according to the manufacturer's standard protocol.

Preparation of Cell Lysates for Various Experiments

HCT-15 cell lysates were used for all experiments. HCT-15 cells were grown to 80% confluency in the media described above. The cell media was aspirated and the cell culture plates were placed on ice for all subsequent steps. The cell monolayer was washed gently with 10 mL ice-cold DPBS, with excess DPBS removed. Fresh cold 10mL cold DPBS buffer was added and using a cell scraper, the HCT-15 cells were harvested via centrifugation and resuspended in 400µL cold PBS. The cell pellets were, next, sonicated on an Ultrasonic Sonicator (Misonix S-4000) at Power 2, x10 pulses, one second on, one second off for two rounds on ice. Next, the lysates were incubated on ice for 15 minutes and clarified via centrifugation at 13,000 x g for 5 minutes at 4°C. The various protein concentrations were determined and then normalized to 2mg/mL using a standard DC protein assay (Bio-Rad) and used therefrom.

In-Gel Fluorescence with IA-Rhodamine

Samples of HCT-15 cell lysate (2 mg/ml, 25 µL) were pretreated with either the DMSO vehicle, or varying concentrations (as indicated in Figure 1C, Figure S.1) of the compound, at room temperature: 2-chloroamidine (CAM), or Iodoacetamide (IA) for 1 hour. Next, each sample was incubated with IA-rhodamine (final concentration of 7 µM) at room temperature for 1 h. The reaction was stopped by the addition of a 4x SDS loading buffer (a reducing Laemli buffer system; BioRad). Next, the samples were boiled at 95 °C for 5 min and separated on precast *Any KD* Criterion TGX gels (BioRad #5678124). The various probe-labeled proteins were analyzed by in-gel fluorescence using a ChemiDoc MP Imaging System (BioRad).

Proteomic Sample Preparation for Unenriched Sample Analysis

Freshly prepared HCT-15 lysates (concentration = 2 mg/mL, 200 µL, aliquoted and labeled 1-6) in PBS, as described above, were each incubated with DTT (10 µL of 200 mM stock in water, final concentration = 10 mM) at 65 °C for 15 min. Next, the various compounds (CAM-treated at a final concentration of 200 mM, for samples 1-3, IA-treated samples at a final concentration of 20 mM, for samples 4-6) were added and incubated for 1 hour at 37 °C.

Samples were next subjected to Single-Pot Solid-Phase-enhanced sample preparation (SP3) for protein level cleanup.[6-7] For each 200 µL sample, 20 µL Sera-Mag SpeedBeads Carboxyl Magnetic Beads, hydrophobic (GE Healthcare, 65152105050250) and 20 µL Sera-Mag SpeedBeads Carboxyl Magnetic Beads, hydrophilic (GE Healthcare, 45152105050250) were gently mixed. Washes were repeated three times. The bead slurries were then transferred to the samples, and incubated for 5 min at RT with shaking. Absolute ethanol was added to each sample, and the samples were incubated for 5 min. The beads were further washed three times with 80 % ethanol in water. Beads were then resuspended in 200 µL 0.5 % SDS in PBS containing 2M urea. DTT (10 µL of 200 mM stock) was added to each sample and incubated at 65 °C for 15 min. Next, absolute ethanol was added to each sample, and the samples were incubated for 5 min with shaking (1000 rpm). Beads were then again washed three times with 80 % ethanol in water. Next, beads were resuspended in 200 µL PBS containing 2 M urea followed by the addition of 2 µL trypsin solution (Worthington Biochemical). Digestion was allowed to proceed overnight at 37 °C with shaking.

After digestion, ~4 mL acetonitrile was added to each sample, and the mixtures were incubated for 10 min at RT with shaking. The beads were washed (3×1 mL acetonitrile). After SP3 clean-up, the peptides were then eluted from SP3 beads with 100 µL of 2 % DMSO in water for 30 min at 37 °C. The elution was repeated.

The next day, 90 µL from each digest was combined with 210 µL water and 0.3µL TFA (final concentration ~0.1% TFA and ~180µg peptides). Samples were fractionated into low-bind eppendorf tubes using a high-pH reversed-phase fractionation kit (Pierce, 84868)., using the following buffer systems in Table S.2:

For each treated sample, a total of eight fractions were obtained. Fractions were dried (Speed Vac) and then reconstituted with 15 µL 5% acetonitrile and 1% FA in MB water and analyzed by LC-MS/MS. Samples were fractionated in triplicates for a total of 48 samples each for CAM and IA-treated samples.

The (2-bromoethylamine hydrobromide) BEA samples were processed acoording to reported procedures[8]: Freshly prepared HCT-15 lysates (concentration = 2 mg/mL, 200 µL) were denatured with 8 M urea (550 mM Tris-HCl pH 8.6) and next reduced with 50 mM DTT at 25 °C for 1 h. Next, the samples were s-aminoethylated with 250 mM BEA for 2 h at 25 °C in the dark. Next, the resultants were diluted 5-fold with 50 mM ammonium bicarbonate and subjected to trypsin digestion (protein/enzyme = 100:1). Digestion was allowed to proceed overnight at 37 °C with shaking. The digests were desalted in reversed phase StageTips. The desalted digests were dried (Speed Vac) and then reconstituted in 5% acetonitrile and 1% FA in MB water and analyzed by LC-MS/MS.

LC-MS/MS Data Analysis

Samples were analyzed by liquid chromatography-tandem mass spectrometry using Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Scientific) coupled to an Easy-nLC™ 1200 pump. Data was analyzed using FragPipe V20.0. Raw data collected by LC–MS/MS were searched with MSFragger (version 3.1.1).

Raw data collected by LC-MS/MS were searched using a slightly modified LFQ FragPipe workflow - with semi-tryptic searches: MSFragger (version 3.5), Philosopher (version 4.2.2), and IonQuant (version 1.8.0) enabled (Kong et al. 2017; Yu, Haynes, and Nesvizhskii 2021; Yu et al. 2020; Leprevost et al. 2020). Peptide length was set to 7 - 50 and peptide mass range was set to 500 - 5000. Cysteine residues were searched with variable modifications at cysteine residues for cysteine capping (+56.0374), IA capping (+57.0215) or BEA capping (+43.0422). Mean LFQ intensities were calculated.

The MS search results and Fasta files have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Perez-Riverol et al. 2019) with the dataset identifiers PXD047844 and (PXD052935). Datasets are in **Table S1-S5** and file details in **Table S6, S7.**

Mass Offset Search Analysis for CAM Selectivity

We applied established mass offset search-based workflow (released in FragPipe 22 under the name 'PAL') to our datasets[1-2]. Using this workflow, the CAM modification masses were specified as mass offsets (this delta masses restricted on any nucleophilic amino acid: CSTMHK) with fixed modifications allowed on all amino acids as well as all c-term peptides, n-term peptides, c-term proteins and n-term proteins. The labelled amino acids for the mod and their score best position was determined and represented as bar plots.

Proteomics acquisition

The samples were analyzed by liquid chromatography-tandem mass spectrometry using a Thermo ScientificTM Orbitrap EclipseTM TribridTM mass spectrometer or coupled with a High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) Interface. Peptides were fractionated online using a 16 cm long, 100 µM inner diameter (ID) fused silica capillary packed in-house with bulk C18 reversed phase resin (particle size, 1.9 µm; pore size, 100 Å; Dr. Maisch GmbH). The 70-minute water-acetonitrile gradient was delivered using a Thermo ScientificTM EASY-nLCTM 1200 system at different flow rates shown in the table below:

Data processing

Custom python scripts were implemented to compile labeled peptide datasets. Unique proteins, unique cysteines, and unique peptides were quantified for each dataset. Unique proteins were established based on UniProt protein IDs. Unique cysteines were classified by an identifier consisting of a UniProt protein ID and the amino acid number of the modified cysteine (ProteinID_C#); residue numbers were found by aligning the peptide sequence to the corresponding UniProt protein sequence. Unique peptides were found based on sequences containing a modified cysteine residue.

C-Terminus, Non-proteotypic, and CysDB Annotations:

CAM-capped c-terminus cysteines were determined based on the location of the modified residue within each detected peptide sequence. Peptides that had a cam-capped cysteine as the last residue in the sequence were marked as c-terminus-labeled cysteines. Additionally, we performed various in silico trypsin digests, with methodology previously reported in (3). For our first in silico digest, we considered only lysine and arginine residues as potential cleavage sites, we did not allow any miscleavage events and considered peptides ≥ 45 amino acids in length as long non-proteotypic. Then, we performed a second in silico digest, where we considered lysine, arginine, and cysteine as cleavage sites. For Lys/Arg/Cys (KRC) trypsin digest, we allowed up to two miscleavage events and considered peptides of lengths between 6 to 45 amino acids. The peptides found from the KR and KRC trypsin digests were then compared to identify unique KRC non-proteotypic tryptic peptides. Finally, cysteines annotated as CAM-capped, c-terminus CAM-capped or CAM-capped non-proteotypic were then compared with unique cysteine identifiers reported in CysDB (2306-release)[9]

Charge State and Peptide Length Analysis:

Peptides were grouped based on their listed charge state from FragPipe outputs. If a peptide had multiple charge states it would get binned into each of those charge state groups. The median peptide intensity was plotted per charge state. For peptide lengths, peptides were binned into lengths between 7 and 49 amino acids long. Total peptides in each category were then normalized by the percentage of peptides in one length group out of the total number of peptides. Data was plotted and statistical analysis was performed in GraphPad Prism (version 10).

Function and Location Analysis:

For all the peptides labeled by CAM, the UniprotKB annotations for function and location were assigned for each protein assigned to the labeled peptide. The peptides were grouped within these annotations and if a protein had multiple annotations, the peptide would be grouped in all of the relevant annotations. The frequency of each function/location was calculated by dividing the number of peptides belonging to those groups by the number of total peptides. This analysis was also performed for the CAM c-terminal cysteine peptides. The fold change difference was calculated by comparing the frequency for all CAM-labeled peptides to the CAM c-terminal cysteine peptides to find a fold change enrichment.

(E) Chemistry Methods

Synthetic Procedures

For chemical synthesis, all solution-phase reactions were performed in dried glassware under an atmosphere of dry N_2 except where water was used as a solvent. Silica gel P60 (SiliCycle) was used for column chromatography. Analytical thin layer chromatography (TLC) was performed on 0.25 mm silica gel 60- F254, which was visualized by fluorescence quenching under UV light or by staining with iodine. Solvents and reagents were purchased from Sigma-Aldrich (St. Louis, MO), Alfa Aesar (Ward Hill, MA), EMD Millipore (Billerica, MA), Fisher Scientific (Hampton, NH),

Oakwood Chemical (West Columbia, SC), Combi-blocks (San Diego, CA) and Cayman Chemical (Ann Arbor, MI) and used without further purification. Low-resolution LC-MS analysis of the various reaction mixtures was analyzed on an Agilent Technologies InfinityLab LC/MSD single quadrupole LC/MS (ESI source). High-resolution mass spectrometry was analyzed on a Waters LCT Premier with ACQUITY LC and autosampler (ESI source).

NMR analysis

^{1}H NMR and ^{13}C NMR spectra for characterization of new compounds and monitoring reactions were collected in $CDCl_3$, CD_3OD, CD_6CO, or DMSO-d_6 (Cambridge Isotope Laboratories, Cambridge, MA) on a Bruker AV 400 MHz spectrometer or Bruker AV 500 MHz in the Department of Chemistry & Biochemistry at The University of California, Los Angeles. All chemical shifts are reported in the standard notation of parts per million using the peak of residual proton signals of the deuterated solvent as an internal reference. Coupling constant units are in Hertz (Hz). Splitting patterns are indicated as follows: br, broad; s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublets; dt, doublet of triplets.

Synthesis of CAM

Using slightly modified procedures[10], 2-chloroethanimidoate·HCl (500 mg, 4. 11 mmol) was added to a stirred solution of sodium ethoxide (431 mg, 6.33 mmol) in ethanol under a nitrogen atmosphere and stirred for 3 h. Next, ammonium chloride (249 mg, 4.75 mmol) was added portion-wise for 10 min to the reaction mixture and stirred for 24 h under the same inert conditions. Next, the reaction mixture was acidified with HCl (4 mL, 4M in dioxane). Then concentrated in vacuo to remove all the volatiles. Next, diethyl ether was added to the crude, and the precipitates were filtered and dried under a high vacuum to yield the titled compound (300 mg, 74%).

^{1}H NMR (400 MHz, DMSO-d_6): δ 9.38 (d, J = 1.3 Hz, 4H), 4.42 (s, 2H). **^{13}C NMR (101 MHz, DMSO-d6)** δ 166.28, 39.09.

HRMS $[M+H]^+$ (ESI-TOF) = $C_2H_6ClN_2^+$: calculated for 93.0219; Found 92.9585

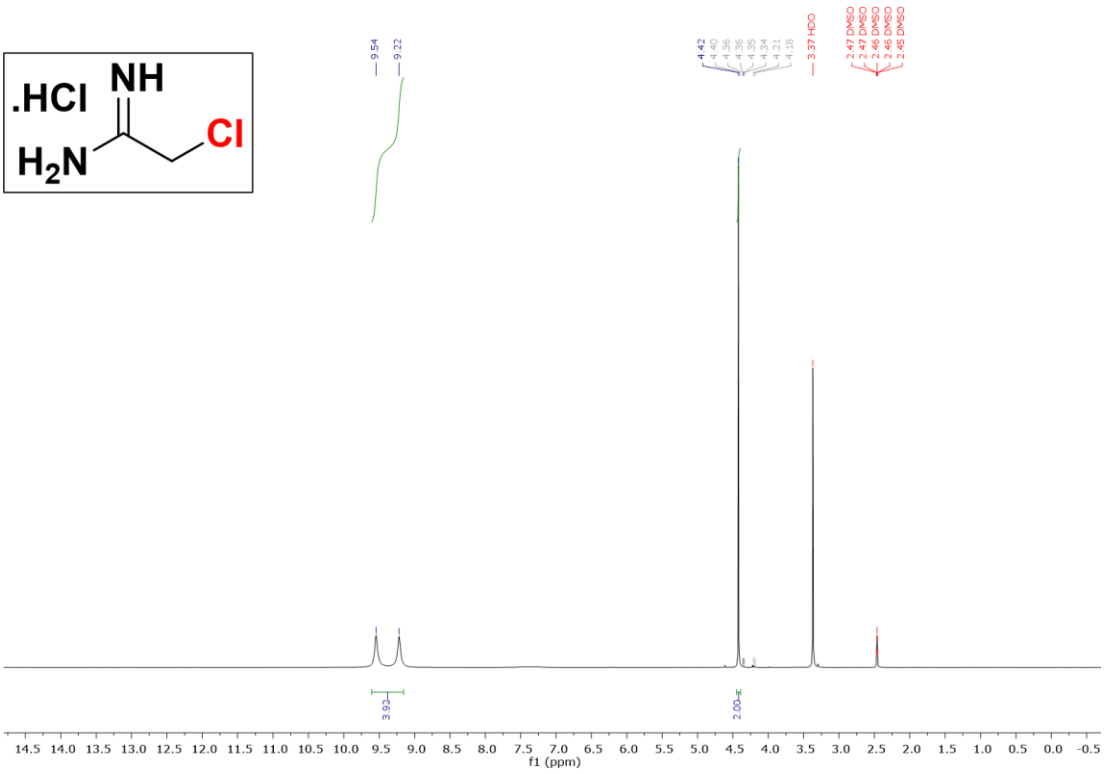

Figure S.2. [1]H NMR of 2-Chloroacetamidine Hydrochloride, **CAM**, in DMSO-d_6

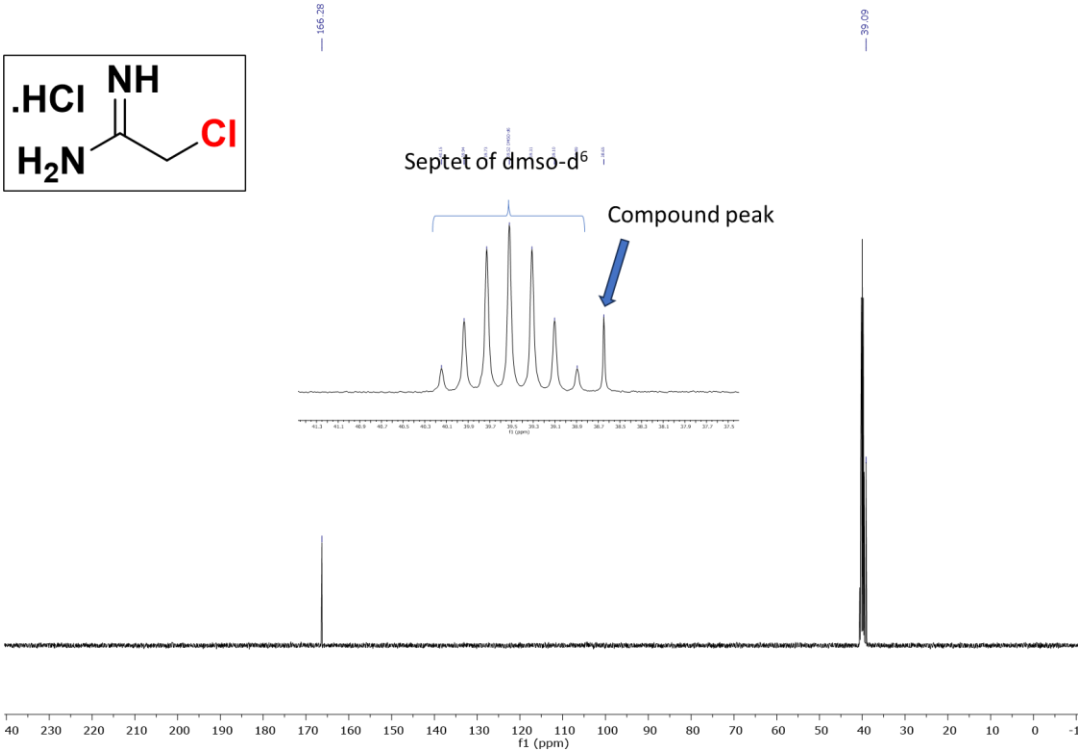

Figure S3. [13]C NMR of 2-Chloroacetamidine Hydrochloride, **CAM** in DMSO-d_6

Supplemental References

1. Takechi, S. N., C.; Burton, N.; Villanueva, M.; Boatner, L.; Yu, F.; Kim, P.; Biletch, E.; Turmon, A.; Nesvizhskii, A.; Backus, K., Silyl Ether Enables High Coverage Chemoproteomic Interaction Site Mapping. *ChemRxiv This content is a preprint and has not been peer-reviewed.* **2024,** *10.26434/chemrxiv-2024-21r7b*.
2. Yu, F.; Teo, G. C.; Kong, A. T.; Haynes, S. E.; Avtonomov, D. M.; Geiszler, D. J.; Nesvizhskii, A. I., Identification of modified peptides using localization-aware open search. *Nat Commun* **2020,** *11* (1), 4065.
3. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L., The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* **2007,** *35* (Database issue), D301-3.
4. Sanner, M. F., Python: a programming language for software integration and development. *J Mol Graph Model* **1999,** *17* (1), 57-61.
5. Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S., AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* **2021,** *61* (8), 3891-3898.
6. Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J., Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* **2019,** *14* (1), 68-85.
7. Yan, T.; Desai, H. S.; Boatner, L. M.; Yen, S. L.; Cao, J.; Palafox, M. F.; Jami-Alahmadi, Y.; Backus, K. M., SP3-FAIMS Chemoproteomics for High-Coverage Profiling of the Human Cysteinome*. *Chembiochem* **2021,** *22* (10), 1841-1851.
8. Tomioka, R.; Tomioka, A.; Ogata, K.; Chan, H.-J.; Chen, L.-Y.; Guzman, U. H.; Xuan, Y.; Olsen, J. V.; Chen, Y.-J.; Ishihama, Y., Extending the Coverage of Lys-C/Trypsin-Based Bottom-up Proteomics by Cysteine S-Aminoethylation. *Journal of the American Society for Mass Spectrometry* **2024,** *35* (2), 386-396.
9. Boatner, L. M.; Palafox, M. F.; Schweppe, D. K.; Backus, K. M., CysDB: a human cysteine database based on experimental quantitative chemoproteomics. *Cell Chem Biol* **2023,** *30* (6), 683-698.e3.
10. [US]; MURRAY ROBERT S [US]; RENOU CHRISTELLE C [US]; REYNOLDS DOMINIC [US], M. P. I. U. E. A. M. U. G. S. U. H. S. J. U. H. M. E. U. L. J. F. Beta carbolines and uses thereof.