

An unbiased proteomic platform for ATE1-based arginylation profiling

Received: 17 June 2024

Accepted: 10 July 2025

Published online: 25 August 2025



Zongtao Lin¹✉, Yixuan Xie¹, Joanna Gongora¹, Xingyu Liu¹, Emily Zahn¹, Bibhuti Bhusana Palai¹, Daniel H. Ramirez¹, Richard M. Searfoss¹, Francisca N. Vitorino¹, Rashmi Karki¹, Geoffrey P. Dann², Chenfeng Zhao³, Xian Han¹, Brittany MacTaggart⁴, Xin Lan⁵, Dechen Fu⁵, Lina Greenberg¹, Yi Zhang⁵, Kory J. Lavine⁶, Michael J. Greenberg¹, Dongwen Lv^{7,8}, Anna Kashina⁴ & Benjamin A. Garcia¹✉

Protein arginylation is an essential post-translational modification catalyzed by arginyl-tRNA-protein transferase 1 (ATE1) in mammalian systems. Arginylation features a post-translational conjugation of an arginyl to a protein, making it extremely challenging to differentiate from translational arginine residues with the same mass. Here we present a general ATE1-based arginylation profiling platform for the unbiased discovery of arginylation substrates and their precise modification sites. This method integrates isotopic arginine labeling into an ATE1 assay utilizing biological lysates (ex vivo) rather than live cells, thus eliminating ribosomal bias and enabling bona fide arginylation identification. The method has been successfully applied to peptide, protein, cell, patient and mouse samples, with 235 unique arginylation sites revealed from human proteomes using 20 µg of input. Representative sites were validated and followed up for their biological functions. This global platform, applicable to various sample types, paves the way for functional studies of this difficult-to-characterize protein modification.

As a critical post-translational modification, arginylation is catalyzed by arginyltransferase (ATE1), which is the only known enzyme installing arginylation in mammalian systems¹. The absence of arginylation in ATE1 knockout (KO) models resulted in embryonic lethality due to heart defects¹. Specifically, ATE1 KO caused thinned myocardium with immature septa, and nonseparation of the aorta and pulmonary artery, resulting in defects in cardiac contractility, myofibril dysfunction and eventually embryonic death¹, demonstrating the essential nature of arginylation¹. Tissue-specific knockdown or deletion (KO) of arginylation resulted in impaired myosin phosphorylation and thrombus

formation², elevated myocardial fibrosis and progressive heart failure³, cardiomyocyte hypertrophy⁴ and many other symptoms^{5,6}. At the molecular level, the arginylation field has long tried to understand the biological roles of arginylation in cardiovascular-associated proteins including β-actin (Asp3 arginylation) regulating cytoskeleton and cell motility^{7,8}, calreticulin (CALR) (Asp18 arginylation) regulating stress granules in a Ca²⁺-dependent manner^{9,10} and RGS4/5/16 (tri-oxidized Cys2 arginylation) acting as nitric oxide and oxygen sensors^{11,12}. Many of these studies focused on the implications of shortened half-lives of proteins involved in cardiovascular biology^{1,7–12}.

¹Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, MO, USA. ²Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA. ³McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO, USA.

⁴School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁵Department of Biochemistry, Case Western Reserve University, Cleveland, OH, USA. ⁶Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. ⁷Department of Biochemistry and Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. ⁸Present address: Department of Cellular Biology and Anatomy, Medical College of Georgia, Augusta University, Augusta, GA, USA. ✉e-mail: zongtao@wustl.edu; bagarcia@wustl.edu

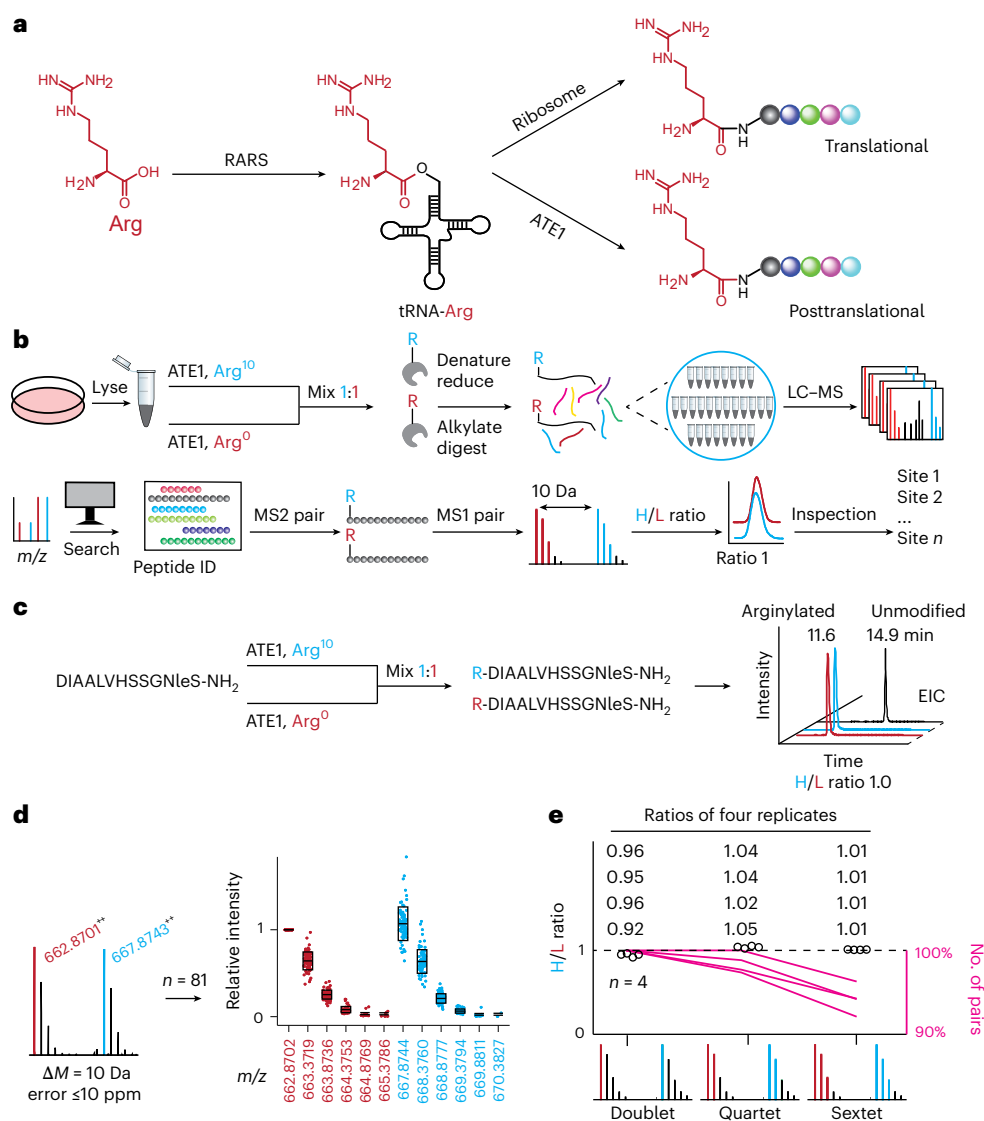


Fig. 1 | Isotopic arginine labeling and detection strategy for ATE1 substrates and arginylation sites. a, Scheme of arginyl installation onto proteins by ribosomal synthesis (translational) and ATE1 (post-translational). **b**, Arginylation profiling platform for arginylation site and substrate discovery from biological samples. Lysate is labeled by isotopic arginine molecules (Arg¹⁰ and Arg⁰), respectively, using ATE1 assay, mixed and digested. Peptides are fractionated and analyzed by mass spectrometry in data-dependent acquisition mode. Proteomics data are searched to produce peptide identifications (IDs), among which peptide pairs modified by H and L Arg are further evaluated for MS1 isotopic features. **c**, Isotopic arginylation of a peptide ATE1 substrate. EICs were extracted using monoisotopic peaks based on calculated *m/z* values. EIC in black

indicates chromatography of the unmodified peptide. **d**, Isotopic feature in MS1 spectra and their summary (upper, center and lower: 75, 50 and 25 percentiles, respectively). The first monoisotopic peaks in technical MS1 scans (*n* = 81) from a representative LC-MS run is set to 1 for normalization. Relative intensities of other isotopic peaks (*n* = 81, 80, 61, 28, 8, 81, 81, 79, 53, 25 and 3 technical scans) are displayed. A 10-ppm error is set for all MS1 isotopic peaks. **e**, Ratio summary of MS1 pairs in four replicates (*n* = 4) using doublet, quartet and sextet peaks. Detailed ratios from the replicates are provided. The numbers of pairs in quartet and sextet are normalized to the numbers of pairs in doublet from respective replicate. Nle, norleucine.

Meanwhile, many studies suggested that arginylation is also important for certain proteins to function properly (noncanonical roles) beyond its canonical role in the N-degron pathway^{13,14}. It is reported that arginylation of β -amyloid guides proper α -helical shape preventing misfolding and aggregation¹⁵. Arginylation of α -synuclein (α -syn) facilitates brain health by preventing neurodegeneration^{16,17}. Arginylation of CALR promotes its association with stress granules^{18,19}. Decreased arginylation of nuclear proteins results in smaller nucleus size and architecture²⁰. In addition, arginylation has been detected on histone proteins^{20,21}, potentially facilitating their interaction with DNA through the positively charged Arg residue. Other prominent arginylated examples are chaperone HSPA5 (ref. 22), BRCA1 (ref. 23), PDI²⁴ and CDC6 (ref. 25). It is becoming clear that protein arginylation

has many biological functions yet to be determined. However, only a handful of proteins have been validated for their arginylation sites, and the discovery of arginylation substrates and sites is becoming a bottleneck for arginylation biology.

Identification of arginylation has been inherently challenging because both ATE1 and ribosomes use the same arginine (Arg or R) source, arginyl-transfer RNA (tRNA), to add an arginyl group to a protein, yielding post-translational and translational addition, respectively, of arginyl with the same mass shift (+156 Da)¹³ (Fig. 1a). This makes unbiased differentiation of post-translational arginylation and translational arginine residues extremely difficult when both ATE1 and the ribosome are active (that is, when cells are alive). In addition, arginylated proteins may go through the Arg-N-degron

pathway for rapid ubiquitin-mediated degradation, decreasing the endogenous arginylation level for detection^{13,25}. Furthermore, arginylation mostly happens on the protein N-termini after proteolytic cleavage, which is poorly understood for individual proteins in the whole proteome¹³. Further confusion comes from trypsin mis-cleavage resulting in peptides starting with an R residue in bottom-up proteomics. Disregarding those challenges, efforts have been made aiming to systematically identify arginylation.

One early approach to screening arginylation substrates involved adding subdivisions of a complete complementary DNA library to a transcription–translation–degradation system, which assumes that arginylated substrate is prone to the N-degron pathway²⁶. A few arginylation substrates (for example, BiP and PDI) have been identified from the whole proteome through the combination of ATE1 assay *ex vivo* and [³H]arginine autoradiography²⁴. The autoradiography was also used for further identification of individual proteins (for example, CALR¹⁰ and α -syn¹⁶) as ATE1 substrates. Notably, these methods heavily relied on mass spectrometry to reveal the arginylation sites. In contrast, proteomic profiling was attempted to identify the endogenous (*in vivo*) arginylation sites using anti-arginylation antibody enrichment. Such an approach enabled the first proteomic analysis to identify 43 plausible arginylated proteins²¹. Similarly, two later studies identified 19 and 15 proteins potentially arginylated on their side-chain Asp and Glu residues^{27,28}, a different ATE1 catalysis mechanism from canonical N-terminal arginylation. However, most of those sites have little overlap with existing confirmed proteins and/or sites from *ex vivo* experiments and were largely postulated from search algorithms without experimental validation. Therefore, there is still a huge unmet need for an unbiased arginylation profiling method.

Inspired by unbiased labeling using isotopic Arg²⁴ and proteomics²¹, here we present a technology for the unbiased discovery of ATE1 substrates and arginylation sites from whole proteomes (Fig. 1b). This work could serve as the technological foundation for studying the functions of arginylation, will open new biochemical frontiers and thus have a long-lasting impact on the arginylation field.

Results

Isotopic arginylation of substrate peptide by ATE1 *in vitro*

To test the proof-of-concept of the arginylation profiling strategy based on ATE1 enzymatic activity (Fig. 1b), we started by isotopically labeling a peptide with an unmodified Asp (D) residue at the N-terminus (Fig. 1c and Supplementary Fig. 1). The ATE1 assay was slightly modified from the well-established conditions²⁹, where arginine-tRNA ligase (RARS) and ATE1 enzymes were kept at the micromolar range. The arginylation activity of the assay was dependent on essential components such as arginine, ATP, tRNA, RARS, ATE1 and substrate (Supplementary Fig. 2). Replacement of N-terminal D residue to acetylated D or Val produced no detectable arginylation (Supplementary Figs. 3 and 4). The arginylation was time dependent (0–60 min), and most peptides were arginylated within 30 min or longer (Supplementary Fig. 5). Higher concentrations (20–300 μ M) of substrate produced more arginylation product but correlated to lower yields (Supplementary Fig. 5). The incubation time and substrate concentration were set at 30 min and 100 μ M, respectively, while arginine (2 mM) and ATP (2 mM) remained excessive.

After the ATE1 assays using isotopic Arg¹⁰ and Arg⁰, respectively, the reactions were mixed (1:1) and desalted for analysis (Fig. 1c). The extracted ion chromatograms (EICs) of the products were manually examined for hypothetical isotopic ratio (1.0) based on coelution peak intensities (Fig. 1c), demonstrating that the isotopic labeling strategy could introduce Arg¹⁰ to arginylation sites. The data were also searched to confirm the products (Supplementary Fig. 1). The MS1 spectra were extracted to validate the Arg¹⁰–Arg⁰ pair features in doublets ($1 + 1' = 2$ isotopic peaks, error ≤ 10 ppm) and summarized into a box plot (Fig. 1d). The coeluting isotopic peaks indicated unbiased identification of arginylation. The average value of heavy/light (H/L) intensity ratios

calculated from individual MS1 spectra was considered the H/L ratio in a liquid chromatography with mass spectrometry (LC–MS) run. The isotopic labeling of the standard peptide was replicated four times with consistent ratios (Fig. 1e). An increasingly stringent MS1 pair detection (error ≤ 10 ppm) using a quartet ($2 + 2' = 4$ isotopic peaks) and sextet ($3 + 3' = 6$ isotopic peaks) yielded H/L ratios closer to hypothetical 1.0 than the doublet but gave fewer MS1 pairs. To ensure all MS1 pairs are included, we decided to use the MS1 doublet as the threshold filter and include a duodecet ($6 + 6' = 12$ peaks) for pair overview (Fig. 1d). When isotopic peptides were mixed at other ratios (R¹⁰:R⁰-peptide, 1:2 and 7:10), experimental H/L ratios around 0.5 and 0.7 were observed as expected (Supplementary Figs. 6 and 7), another approach to confirm the products. For the convenience of experiments and data analysis, an isotopic 1:1 mix strategy is used in the rest of this study.

Development of in-house software for dataset analysis

We then developed a computational tool ‘ArginylomePlot’ (Fig. 2a) to facilitate handling datasets from complex samples. The software should be able to (1) filter out MS2 pairs with the same sequence modified by Arg¹⁰ and Arg⁰ (change in mass (Δ m) of 10.008269 Da) disregarding other modifications (for example, carbamidomethylation), (2) extract their MS1 scans from raw data, (3) summarize H/L ratios of each arginylation peptide and (4) export paired MS1 and MS2 scans for data visualization. The software is publicly available to download. Briefly, mzXML data converted from raw data were used as input, and search results from proteomics software (for example, Byonic) were used as peptide input. Once a coelution MS2 pair was identified, their MS1 information was used to extract pairs from mzXML for ratio calculation. To present high-confidence proteomics data, peptide score (H or L ≥ 300) was applied as a filter before MS2 spectrum export. The instructions and an exemplary dataset are provided in Supplementary Data 1.

Isotopic arginylation of HEK293T peptides by ATE1 *in vitro*

Assisted by the software, we tested the arginylation of a whole-proteome peptide mixture from human embryonic kidney 293T (HEK293T) cells with ATE1 knocked out to avoid potential interference from endogenous arginylation (Fig. 2b). After data analysis, a total of 772 arginylation pairs belonging to 167 unique arginylation sites were detected at the MS1 level (Fig. 2c, Supplementary Fig. 8 and Supplementary Data 2). The data showed exclusive N-terminal arginylation (Supplementary Data 2), even though we searched for both N-terminal and side-chain arginylation at the same time. While charges 2 (83.7%) and 3 (13.8%) account for 97.5% of all identified peptides, most arginylated peptides (98.3%) are detected at charges 2 (47.3%) and 3 (51.0%) with a small portion in charge 4, indicating a charge shift for arginylated peptides (Supplementary Fig. 9). The distribution profile of H/L ratios (threshold $0.1 \leq \text{ratio} \leq 10$) from all MS1 pairs is centered at hypothetical 1.0 (Fig. 2d) per the experimental design, demonstrating the success of the workflow. The doublet ratios were compared with the quartet and sextet profiles (Supplementary Fig. 10). When a peptide pair was observed with the same charge state and a Δ m of 10.008269 Da, the averaged retention time of relevant peptide IDs containing both Arg¹⁰ and Arg⁰ was used to create a ± 1.25 -min window for paired MS1 scan extraction. MS1 scans containing doublets (10-ppm error) were exported. The average value of H/L intensity ratios calculated from individual MS1 scans was considered the H/L ratio of a peptide pair. The numbers of MS1 scans with the doublet feature used for generating respective H/L ratios are provided (Fig. 2d and Supplementary Fig. 11). When looking at the arginylated N-terminal residues of all unique peptides, most arginylation happened on D and E (including Q \rightarrow E) residues (96.36%) (Fig. 2e) while the data suggested possible arginylation on other N-terminal residues (for example, A, G, L and V), consistent with reported specificities and activities of ATE1 (refs. 30,31). We then generated the sequence logo of unique peptides using WebLogo v.3.0 (ref. 32) (Fig. 2f) and the arginylation motif using pLogo³³

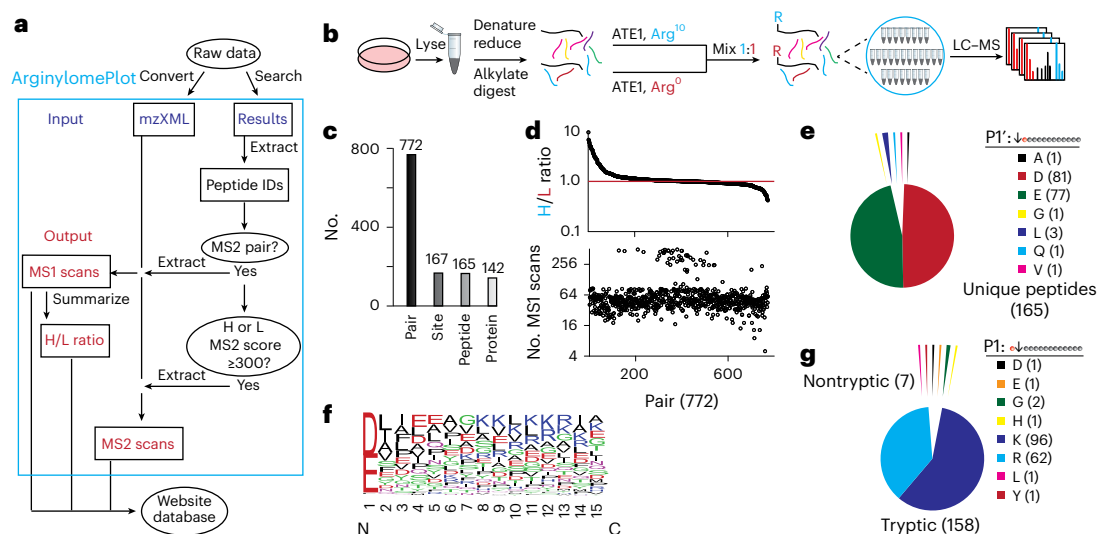


Fig. 2 | Arginylation analysis of whole-proteome peptides using in-house software. **a**, Software flowchart of customized ArginylomePlot. Input data are mzXML and peptide identification files. Output data are MS1, MS2 and the summary of unbiased arginylation sites. **b**, Experimental workflow for arginylation of tryptic peptides from HEK293T (ATE1 KO) cells. **c**, Numbers of identified MS1 pairs, arginylation sites, unique peptides and unique proteins. **d**, H/L ratio distribution of all MS1 pairs and their corresponding numbers of MS1 scans (doublets, error ≤ 10 ppm). The retention times of peptide IDs belonging to a pair were averaged. The software will extract matching MS1 scans within

1.25 min (± 1.25 min) of the averaged retention times. **e**, Analysis of the N-terminal residues of all unique arginylated peptides. Q is arginylated after deamidation (Q \rightarrow E). **f**, Sequence logo calculated from unmodified forms of all unique arginylated peptides. The frequency plot is generated by WebLogo. The peptide sequences were aligned and extended to include 14 positions downstream of the arginylation sites as the P1' position. **g**, Comparison of tryptic and nontryptic N-terminal of all arginylated peptides. A nontryptic N-terminal may indicate endogenously exposed N-termini of proteins.

(Supplementary Fig. 12). The result suggested that arginylation favors N-terminal D and E residues substantially, while subsequent nonpolar residues (I, L, F, V and A) and acidic residues (D and E) may facilitate arginylation (Fig. 2f and Supplementary Fig. 12). While most of the unique peptides (95.76%) are tryptic after K or R cleavage, a few nontryptic peptides were arginylated (Fig. 2g). For example, the arginylation of SSBP E17 (Supplementary Fig. 13) may have resulted from endogenous protein N-terminal cleavage³⁴. This observation demonstrates the potential application of our peptide workflow for the discovery of the cleavage and arginylation of the endogenous protein N-terminus.

Isotopic arginylation of proteins by ATE1 in vitro

We then moved on to protein arginylation using CALR with reported N-terminal E18 arginylation after endogenous cleavage of a signaling peptide (amino acids 1–17)¹⁸. Purified 18E-CALR in HEK293T ATE1 KO cells was isotopically labeled and yielded the H/L ratio of 1.0 in the arginylated E18 peptide (R^{10}/R^0 -E₁₈PAVYFK) (Fig. 3a and Supplementary Fig. 14). The in vitro arginylation efficiency of CALR is preliminarily estimated to be 64.7% for the E18 site based on MS2 spectrum counts of modified (R^{10}/R^0 -E₁₈PAVYFK, counts 33) and unmodified (EPAVYFK, counts 18) peptides. When counting the unmodified peptide (EPAVYFK) in samples with and without ATE1, the arginylation efficiency was estimated to be 80.2% ($n = 3$, Supplementary Fig. 15). Replication of CALR arginylation ($n = 5$) showed consistent ratios near 1.0 (Fig. 3a and Supplementary Fig. 16). In addition, arginylation results ($n = 5$) of a commercial 18E-CALR (Abcam, cat. no. ab276554) showed consistent E18 labeling (Supplementary Fig. 17). In comparison, the 18R-CALR protein was purified and tested, and the results ($n = 5$) showed almost no R^{10} arginylation due to the absence of an open N-terminal E18 residue (Supplementary Fig. 18). The results indicated that the isotopic arginylation strategy could be used to discover modification sites in proteins.

Arginylation profiling in human proteomes

Using this established platform, we profiled 16 human samples by ATE1 assay ex vivo, a key to installing isotopic Arg post-translationally under

ribosome-free conditions to bypass translational Arg incorporation and bias. A total of 1,876 isotopically labeled pairs belonging to 235 unique arginylation sites on 165 proteins were identified (Fig. 3b, Table 1, Supplementary Data 3 and Supplementary Figs. 19 and 20). As a negative control, the HEK293T sample without the addition of ATE1 did not show any arginylation site (RAW data are available). Only seven arginylation sites (ACTC C259 RO2, BII N9 R_{deami}, HBA N79 R_{deami}, HMSD Q43 R_{deami}, PSD12 E348 R, S10AE C74 RO3 and SCN1A C1588 RO2) were assigned as side-chain arginylation accounting for 3% of all unique sites, while the remaining 97% of sites were identified as N-terminal arginylation. Similar to results from HEK293T peptides, most of the peptide pairs are detected at charge states 2 (45.0%) and 3 (49.9%), with small portions in charges 4 and 5 (Supplementary Fig. 21). As cross-validation, CALR E18 and PDIA D18 are among the top high-frequency sites with 185 and 135 detections (MS1 pairs), respectively, consistent with the literature on their confirmed E18 (ref. 10 and D18 (ref. 24) arginylation. The number of unique sites from each sample is listed in Fig. 3c that shows an average of 29 sites per sample. The H/L ratios (threshold used $0.1 \leq \text{ratio} \leq 10$) of MS1 pairs are symmetrically distributed and perfectly centered at the hypothetical ratio of 1.0 (Fig. 3d). Endogenous Arg⁰ at the micromolar level may interfere with the Arg¹⁰ labeling; however, the final concentration of Arg¹⁰ is at 2 mM thus Arg⁰ interference on final H/L ratios might be minimal. Since isotopic labeling was carried out separately on complex whole proteomes, deviations of H/L ratios (range 0.3–8.5) from the hypothetical ratio of 1.0 were expected, which might be introduced from the multi-step preparation procedures including ATE1 arginylation on substrates in the whole proteome, H/L mixing, digestion and peptide fractionation (Fig. 1b). The doublet ratios were compared with the quartet and sextet profiles (Supplementary Fig. 22). The numbers of MS1 scans with the doublet, quartet and sextet features for generating respective H/L ratios are provided, quartet and sextet pairs showed lower numbers of qualified MS1 scans than doublets based on their trendlines (Supplementary Fig. 23). The sequence logo (Fig. 3e and Supplementary Fig. 24) and arginylation motif (Supplementary Fig. 25) from all unique

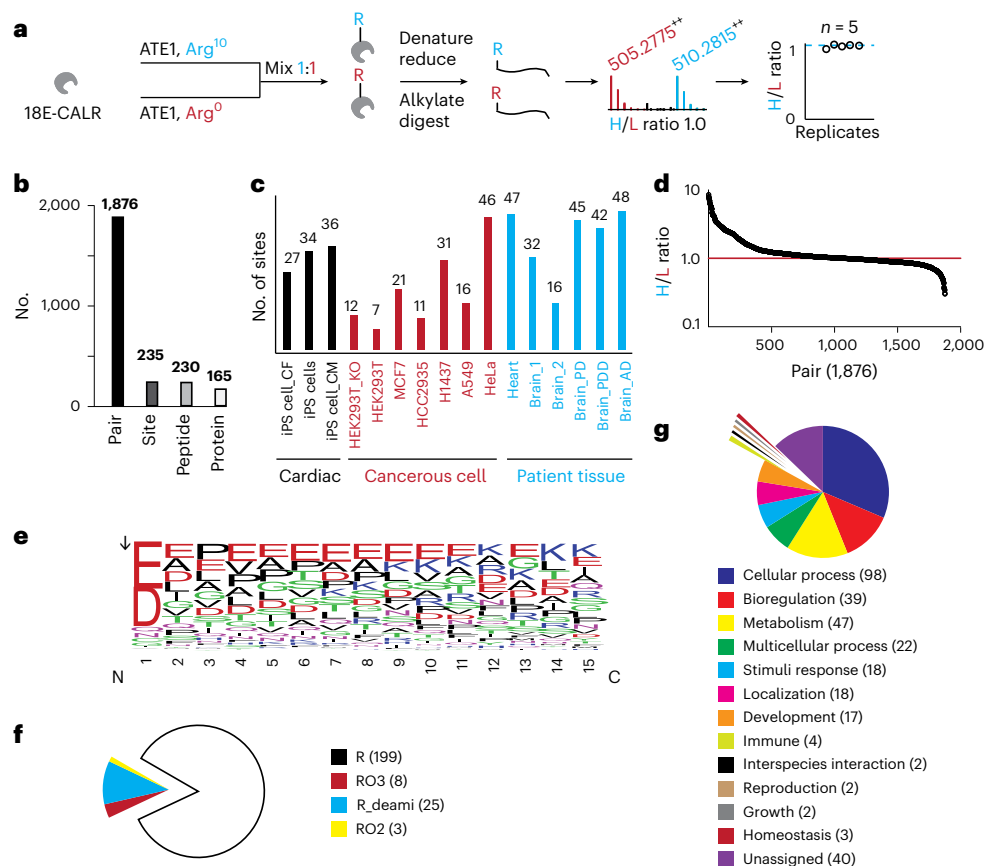


Fig. 3 | Summary of arginylation sites and ATE1 substrates in human proteins and proteomes. **a**, Experimental workflow for CALR arginylation. **b**, Numbers of identified MS1 pairs, arginylation sites, unique peptides and unique proteins in human cells and patient tissues. **c**, Overview of arginylation sites in human cells and patient tissues. **d**, Ratio plot of all MS1 pairs detected in all sample fractions. **e**, Sequence logo calculated from unmodified forms of all unique arginylated

peptides. Frequency plots are generated by WebLogo. The arrow indicates the cleavage site before arginylation. **f**, Arginylation type comparison of all unique sites. **g**, Biological function analysis of ATE1-protein substrates using PANTHER. iPS, induced pluripotent stem. CF, cardiac fibroblasts. CM, cardiomyocytes. R, arginylation. R_deami, N/Q deamidation after deamidation. PD, Parkinson's disease. PDD, Parkinson's disease with dementia. AD, Alzheimer's disease.

peptides suggested that arginylation favors N-terminal D and E residues substantially, while subsequent nonpolar residues (A, L, V and P) and acidic residues (D and E) may facilitate N-terminal arginylation. All four types of arginylation based on different modification masses (Supplementary Fig. 26) have been detected including arginylation (R, 84.7%), Cys tri-oxidation arginylation (RO3, 3.4%), N/Q deamidation arginylation (10.6%) and Cys di-oxidation arginylation (RO2, 1.3%), (Fig. 3f). Comparison of arginylation sites between induced pluripotent stem cells, cancerous cells and patient samples showed only 32 shared sites, while most of the remaining sites are unique to their respective sample groups (Supplementary Fig. 27). Comparisons between individual samples are listed in Supplementary Fig. 28 (cardiac), Supplementary Fig. 29 (brains) and Supplementary Fig. 30 (cancerous), indicating unique and shared arginylation sites from individual samples. Replication ($n = 3$) analysis of four samples showed that most arginylation sites were detected in replicates (Supplementary Fig. 31), indicating the repeatability of the profiling method. Among 165 proteins identified, 162 protein targets are involved in many key biological processes (for example, cellular process, biological regulation and cellular metabolism) according to the PANTHER classification (Fig. 3g).

Validation of arginylation sites using synthetic peptides

We validated representative sites using synthetic peptides (for example, REPAVYFK for CALR E18 peptide) whose MS behaviors (MS1 spectra in Fig. 4a and MS2 spectra in Supplementary Fig. 32) matched those in our MS data. The chosen sites are a mixture of known (CALR and PDIA)

and unknown, and they represent high-frequency (CALR, PDIA, SSBP and RM12), mid-frequency (ERO1A and CH60) and low-frequency (GATB) sites in our data (Table 1). Such a validation added extra confidence to the Arg¹⁰ arginylation that is an unnatural process and thus can be considered an internal validation of our workflow. For example, an H/L ratio of 1.0 for RM12 C45 arginylation (Fig. 4b) eliminated the possibility of tryptic missed cleavage between Arg44 and tri-oxidized Cys45; the standard peptide confirmed this type of arginylation with the copresence of Cys tri-oxidation¹ (Supplementary Fig. 32). It is worth noting that the low frequency of a site does not necessarily mean low confidence since most of the reported sites are based on many isotopic MS1 scans (Supplementary Fig. 23 and Supplementary Data 3) and paired MS2 spectra (Supplementary Data 4). To confirm that these sites possess arginylation at the N-termini but not on the side chain of N-terminal residues, we purchased peptides containing arginylation on the side chain (for example, E(R)PAVYFK for CALR E18 peptide). Analyzed by LC-MS, side-chain arginylation peptides may behave similarly (for example, CALR E18 peptide) or differently (for example, SSBP E17 peptide) with N-terminal arginylation peptides (Supplementary Fig. 33). Arginine residue on the side chain produced a signature ion at m/z 175.1190 when the peptide did not end with a C-terminal R residue (for example, E(R)PAVYFK). When the C-terminal R residue was present on side-chain arginylation peptides (for example, E(R)EQPPETAAQR), both R residues could produce the same ion at m/z 175.1190 with a higher intensity than that from N-terminal arginylation peptides (for example, REEQPPETAAQR) (Supplementary Fig. 33). By adding the side-chain

Table 1 | Number of detections in isotopic pairs and MS1 scans for each arginylation site from human proteomes

Site	n1 ^c	n2 ^c	Site	n1	n2	Site	n1	n2	Site	n1	n2	Site	n1	n2
CALR E18 R ^a	185	8434	LMAN2 D45 R	6	147	CNBP Q44 R_deami	3	84	NDKM K91 R	1	192	VIME D451 R	1	26
MBP D168 R	139	5153	MTAP2 Q431 R_deami	6	122	MYH7 E1902 R	3	71	ACTC E109 R	1	121	HOOKE1 D659 R	1	25
PDIA1 D18 R ^a	135	6071	SRCA E21 R	6	108	CALU D62 R	3	70	AKAP2 D638 R	1	84	MYH7 E1489 R	1	25
SSBP E17 R ^a	127	4744	CBS E4 R	6	101	KTN1 D238 R	3	61	ATAD5 V1449 R	1	70	TBA4A D69 R	1	25
RM12 C45 RO3 ^a	115	3085	MYH7 E1461 R	6	96	HACD3 P100 R	3	58	SRRT D868 R	1	70	ACTA E318 R	1	24
NFL D468 R	66	3260	ACTB E241 R	6	79	PLEC E4050 R	3	56	SFPQ E696 R	1	68	HNRPK Q358 R_deami	1	24
SAP D405 R	52	1485	H2B1K E94 R	6	35	MYH7 D1229 R	3	54	VATD E242 R	1	63	MYH7 E1376 R	1	24
A1AT E25 R	48	1648	NFM E497 R	5	321	TBA1B D327 R	3	52	CC148 I574 R	1	54	MYH6 E73 R	1	23
CATD E64 R	38	986	MBP T200 R	5	264	NFL Q75 R_deami	3	46	NFM E884 R	1	51	NOP16 E77 R	1	23
CATD G65 R	36	1008	CH60 Q42 R_deami	5	207	CALD1 E161 R	3	42	HMGA1 E47 R	1	49	NPC2 E20 R	1	23
ERO1A E24 R ^a	36	984	LN28A E18 R	5	200	NT5D3 C44 RO3	3	35	RS7 E188 R	1	49	ACLY E830 R	1	21
F162A C41 RO3	26	516	LETM1 D116 R	5	188	PTBP1 D7 R	2	124	ACTA E109 R	1	47	ATS15 D749 R	1	21
TBB5 E3 R	25	1342	TNNI3 E165 R	5	169	FCN1 D120 R	2	111	MYOZ2 D255 R	1	44	ATP5J N33 R_deami	1	20
PDIA4 E25 R	25	756	VIME E425 R	5	166	HS90B E539 R	2	102	NCAM1 E833 R	1	43	KS6A1 D37 R	1	20
GFAP D417 R	22	722	NFL E453 R	5	145	LDB3 E257 R	2	79	BI1 N9 R_deami	1	42	ATIF1 E42 R	1	19
G3P D296R	22	482	MYL3 E131 R	5	119	NPM D178R	2	77	CH60 D209 R	1	41	CH60 D100 R	1	18
ODO2 D68 R	21	735	CH60 D49 R	5	111	ROA2 D161 R	2	76	RON D1232 R	1	41	HNRPL D284 R	1	18
MBP N226 R_deami ^b	21	364	TXND5 E40 R	5	94	VIME D90 R	2	76	CH60 C447 R	1	40	MYH7 E927 R	1	18
MBP E217 R	20	582	SPTN1 E1124 R	5	61	S10AE C74 RO3	2	75	DJC10 D33 R	1	40	SNX17 E463 R	1	18
MBP G277 R	18	488	RL23A E8 R	5	59	XRCC5 E583 R	2	66	RM12 C45 R	1	40	JTB E31 R	1	17
H4 D25 R	18	466	PGRC1 D49 R	4	201	NPM D180 R	2	63	RT18A E37 R	1	40	MITOK C37 RO3	1	17
MBP D216 R	17	550	NFL E469 R	4	193	TCAF1 V114 R	2	59	MYH9 D1846 R	1	39	RS27 D6 R	1	17
NFM E454 R	17	466	CASQ2 E20 R	4	151	MAP1A E663 R	2	58	DREB D636 R	1	38	HBA D48 R	1	16
PGCB D23 R	16	772	TBA1B D69 R	4	146	HSP7C G372 R	2	55	NEUG C9 RO3	1	37	PCDH1 D944 R	1	16
GFAP E391 R	15	600	NFM E599 R	4	129	^a GATB E42 R	2	53	DZ11L Q132 R	1	35	AT5G1 D62 R	1	15
MYH9 D1136 R	15	495	ACTC E318 R	4	119	NPM D179 R	2	53	RLA1 D19 R	1	34	F205A E1023 R	1	15
CH60 E107 R ^a	15	452	NFH E469 R	4	88	EFHD2 Q67 R_deami	2	50	PTBP1 G306 R	1	33	ILF2 D347 R	1	15
MLRV E22 R	14	266	NFL E507 R	4	88	TBB4A D67 R	2	50	TYB4 D6 R	1	32	MAP1B Q907 R_deami	1	14
SAP G407 R	13	172	NHRF1 E121 R	4	77	HBA N79 R_deami	2	48	CH60 G43 R	1	31	MTDC E36 R	1	14
NFM E470 R	12	354	RD23B E136 R	4	65	SCMC1 Q16 R_deami	2	45	HNRPK D279 R	1	31	MYH7 D1008 R	1	14
TBA1B D218 R	12	132	SCG1 D88 R	4	51	TFF1 E25 R	2	42	ROA1 E93 R	1	31	GLYM Q27 R_deami	1	13
API80 N767 R_deami	12	77	EF1D E27 R	4	21	CWC15 Q45 R_deami	2	40	SPTB2 E1469 R	1	31	SF3B1 D93 R	1	13
NFM E496 R	10	770	HNRPK N306 R	3	146	AKA12 E952 R	2	37	SRRT D866 R	1	31	MBP D279 R	1	12
CLUS D23 R	10	179	MBP N226 R	3	142	MYH7 E73 R	2	35	FUBP1 N369 R_deami	1	30	NECP1 C162 RO3	1	11
NEUM E89 R	10	152	APOA1 D25 R	3	140	RL1D1 D319 R	2	32	COX17 D6 R	1	29	RM12 C45 RO2	1	11
ATIF1 E65 R	10	148	H15 E8 R	3	119	BACH C36 RO3	2	31	NUCL E628 R	1	29	ATIF1 D28 R	1	10
DREB D634 R	9	460	CD99 D23 R	3	115	PGCB E412 R	2	29	TBA1B D251 R	1	29	MAP1A E1580 R	1	10
MYH7 E1821 R	9	232	C99L2 D26 R	3	114	UTP18 E374 R	2	26	E2F7 D10 R	1	28	MYPT2 D816 R	1	9
ENPL D22 R	8	555	LDB3 N197 R_deami	3	112	AAK1 Q374 R_deami	2	23	IF4H D196 R	1	28	TNNT2 D108 R	1	9
NCAN E23 R	8	177	HNRPU E259 R	3	110	MLRV N154 R_deami	2	23	AQP4 E279 R	1	27	KAD4 G82 R	1	8
BCLF1 E463 R	8	160	HNRPC D123 R	3	109	TMM11 D163 R	2	20	ATP5H D129 R	1	27	NFL E509 R	1	7
RLA1 N62 R_deami	8	159	RLA1 D18 R	3	109	ACTC C259 RO2	2	17	CALX E275 R	1	27	PDLI5 E297 R	1	7
TAU N644 R_deami	8	136	PELP1 E1008 R	3	104	CDY1 D285 R	2	17	PSMD6 L84 R	1	27	DPEP3 D150 R	1	5

Table 1 (continued) | Number of detections in isotopic pairs and MS1 scans for each arginylation site from human proteomes

Site	n1 ^c	n2 ^c	Site	n1	n2	Site	n1	n2	Site	n1	n2	Site	n1	n2
ACTC D53 R	7	270	CAVN2 Q161 R_deami	3	95	NFM E763 R	2	17	AKA12 E418 R	1	26	HMSD Q43 R_deami	1	5
ACTB E107 R	7	179	G3P D315 R	3	92	ANK2 E1760 R	2	15	COX41 D43 R	1	26	PTGDS E25 R	1	3
ACTB D51 R	6	257	MESD E35 R	3	92	PACN1 N330 R_deami	2	8	MPPB Q46 R_deami	1	26	PSD12 E348 R	1	1
MCFD2 E27 R	6	231	NFM E455 R	3	87	PGCB E26 R	2	7	PDLI5 E215 R	1	26	SCN1A C1588 RO2	1	1

^aValidated by standard peptides. ^bR_deami, N/Q arginylation after deamidation. ^cn1, the total number of fractions and n2, the total number of paired MS1 scans.

arginylation peptides after ATE1 assay, the N-terminal arginylation was further confirmed for the standard peptide (DIAALVHSSGNleS-NH₂) (Supplementary Fig. 34), CALR pure protein (Supplementary Fig. 35) and CALR in HEK293T cells (Supplementary Fig. 36).

Site validation by in-bacteria arginylation assay

We selected two new sites for validating their arginylation in an ATE1–protein coexpression system in *Escherichia coli*³⁵. The two proteins are (1) a chaperone ERO1A in the endoplasmic reticulum, being similar to CALR and PDIA; and (2) SSBP being distinct from endoplasmic reticulum proteins in molecular size (17 kDa), localization (mitochondria) and function (single-stranded DNA binding). Briefly, a protein or peptide containing an arginylation site is fused to ubiquitin and coexpressed with human ATE1 and Ulp1 protease in *E. coli*. The results showed that both ERO1A E24 and SSBP E17 were arginylated (Fig. 4c and Supplementary Figs. 37 and 38). As a bonus, we detected leucylation on top of arginylation (Supplementary Figs. 37 and 38), a unique process to modify N-terminal Arg by leucyl–phenylalanyl-tRNA-protein transferase³⁶. We also tested the in-bacteria arginylation of ERO1A and SSBP peptides, whose respective N-terminal E24 and E17 were also arginylated (Supplementary Fig. 39). The arginylation and leucylation were observed by top-down proteomics (Supplementary Figs. 40 and 41). The results concluded that ERO1A E24 and SSBP E17 are arginylation sites.

Validation of representative arginylation sites in cells

To confirm arginylation of proteins in cells, we overexpressed the wild-type (WT) ERO1A and SSBP in HEK293T cells using CALR as a positive control. Due to the low abundance of endogenous arginylation, arginylated peptides were not detectable in peptide mixtures obtained from whole proteomes if overexpressed proteins were not purified by immunoprecipitation. For example, the MS2 spectrum of arginylated peptide for the SSBP E17 site was only detected once from a pull-down sample (Supplementary Fig. 42). Nevertheless, the arginylated peptides corresponding to all three sites (CALR E18, ERO1A E24 and SSBP E17) were identified after overexpression and immunoprecipitation (Fig. 4d and Supplementary Fig. 42). We also performed the same experiment in HEK293T (ATE1 KO) cells (Fig. 4d) or on E-to-V mutants of ERO1A and SSBP (Supplementary Fig. 43), which did not show any arginylation indicating arginylation dependency on ATE1 and arginylation selectivity on E residue. To test dose dependency on ATE1, we cooverexpressed ATE1 and protein substrates in HEK293T (ATE1 KO) cells. The data show the arginylation levels of the tested sites are dose dependent on the ATE1 expression (Fig. 4e). Such dose dependency was also observed when cooverexpressing CALR and ATE1 in WT HEK293T cells (Supplementary Fig. 44). Both ERO1A (ref. 37) and SSBP³⁸ contain cleavage peptides before the arginylation sites (E24 and E17, respectively), similar to that of CALR. Taking advantage of this feature, we further prepared 24R-ERO1A and 17R-SSBP plasmids by inserting an R residue before E into the full-length plasmids as positive transfection controls, the spectra of their arginylated peptides (Supplementary Fig. 45) further confirmed the arginylation of their WT counterparts.

We have additionally validated the endogenous arginylation in HEK293T cells of two more sites: A1AT E25 (Supplementary Fig. 46) and Tau N644 after deamidation (Supplementary Fig. 47). Similar to CALR, ERO1A and SSBP, A1AT has a signal peptide³⁹ before E25, thus it is not surprising to observe its E25 arginylation. Tau N644 arginylation after deamidation has not been observed previously. We confirmed the arginylation of this site using overexpression of WT Tau plasmid in HEK293T cells (Supplementary Fig. 47). We have also confirmed the arginylation of the Tau E3 site (Supplementary Fig. 47), which was proposed previously as an ATE1 site, but not validated for arginylation⁴⁰. A previous study showed that purified Tau did not show detectable arginylation by ATE1 in vitro¹⁶, possibly due to the lack of exposed arginylation sites (for example, E3 and N644 deamidation). Different from Tau E3 as a calpains cleavage product⁴⁰, the mechanism of generating an open N644 followed by deamidation for arginylation remains to be elucidated.

Biological functions associated with ERO1A arginylation

To investigate whether arginylated ERO1A resides in endoplasmic reticulum, we did imaging on overexpressed ERO1A and 24R-ERO1A in HEK293T ATE1 KO cells. Both proteins colocalized with endoplasmic reticulum marker and PDI (Fig. 5a and Supplementary Fig. 48), suggesting that 24R-ERO1A was translocated into endoplasmic reticulum. As a positive control, we showed that CALR-halo and its arginylation form colocalized with a GFP-tagged endoplasmic reticulum marker (Supplementary Fig. 49). Similar to CALR and PDIA, ERO1A is translocated into endoplasmic reticulum after signal peptide cleavage. A previous study reported the formation of the ERO1A–PDI complex in regulating cellular function⁴¹, our pull-down data showed that 24R-ERO1A interacted with PDI to a comparable extent with and without PDI cooverexpression (Supplementary Figs. 50 and 51). We also showed that arginylation of ERO1A maintains its enzymatic activity on oxidative protein folding compared with WT ERO1A as indicated by the disappearance of the reduced immunoglobulin J chain (JcM) band (Supplementary Fig. 52).

To investigate the arginylation effect on ERO1A stability and degradation using HEK293T ATE1 KO cells, we compared the degradation patterns of ubiquitin-ERO1A fusion proteins with and without arginylation (Supplementary Fig. 52) according to similar experiments performed on CALR⁴². Briefly, ubiquitin is cleaved by de-ubiquitinase after protein synthesis to expose E24 and R24 residues in ERO1A (no arginylation on E24) and 24R-ERO1A (~100% arginylation on E24), respectively. Proteasome inhibitor MG132 treatments after transient transfection increased the levels of ERO1A and 24R-ERO1A, suggesting that both are involved in proteasome-mediated degradation. Similar dose dependence on MG132 was observed in cotreatments when the ribosomal translation of both species was inhibited by cycloheximide (CHX). In addition, 24R-ERO1A showed slightly higher levels after MG132-CHX cotreatment, indicating better stability than ERO1A against proteasomal degradation (Supplementary Fig. 52). To take a closer look at their stabilities, we did CHX chase experiments⁴² where cells were treated with CHX after a 2-day transient transfection. The results showed that 24R-ERO1A was less prone to degradation than ERO1A with

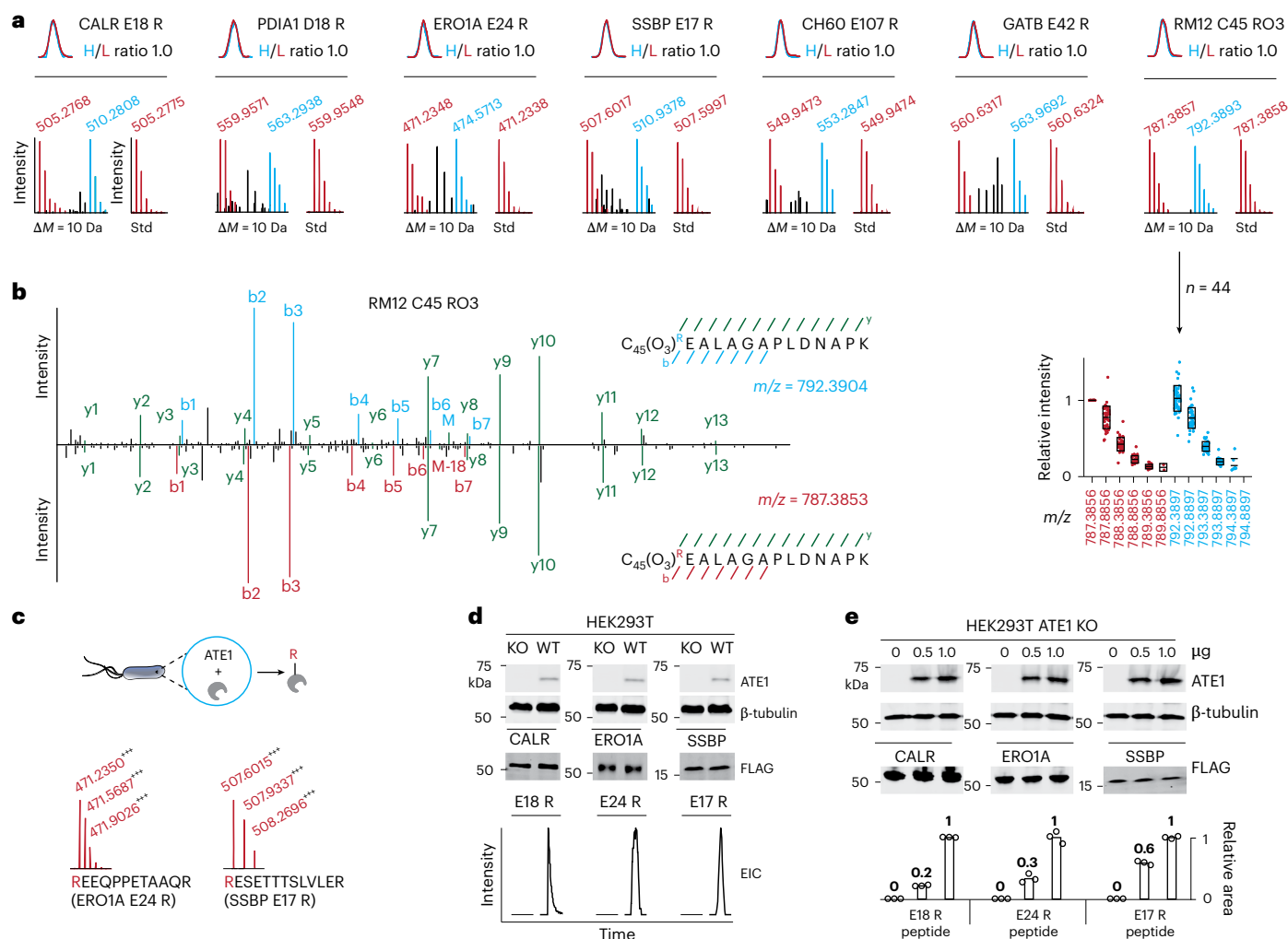


Fig. 4 | Validation of representative arginylation sites. a, Synthetic peptide validation of seven arginylation sites. **b**, Representative MS2 spectra indicating RM12 arginylation on Cys45 with tri-oxidation (C45 RO3) and its MS1 summary (upper, center and lower: 75, 50 and 25 percentiles, respectively). Its first monoisotopic peak in MS1 scans ($n = 44$ technical scans) from a representative run is set at 1 for normalization. Relative intensities of other isotopic peaks ($n = 43, 38, 30, 18, 2, 4, 44, 40, 30, 9$ and 0 technical scans) are displayed. **c**, Open E24 in ERO1A and E17 in SSBP are arginylated in an in-bacteria arginylation system. Protein and ATE1 are coexpressed in *E. coli* for in-bacteria arginylation. The protein was purified and digested for proteomics analysis. **d**, Arginylation dependency of CALR, ERO1A and SSBP sites on endogenous ATE1 ($n = 1$ biological replicate). Anti-ATE1 was used to detect the expression of

endogenous ATE1. β -tubulin was used as a loading control. Anti-Flag antibody was used to detect expressions of CALR, ERO1A and SSBP proteins. EICs of arginylated peptide in each protein in WT and KO cells after pull-down and proteomics are provided. **e**, Relative arginylation levels of CALR, ERO1A and SSBP sites after coexpression of ATE1. Protein was purified by antibody pull-down experiment followed by proteomics analysis. Peak areas of arginylated peptides were normalized to the sample with the highest signal and relative ratios are displayed. Different amounts (0, 0.5 and 1 μ g) of ATE1 plasmids were used for transfection. β -tubulin was used as a loading control. Anti-Flag antibody was used to detect expressions of CALR, ERO1A and SSBP proteins. KO, ATE1 KO; Std, standard peptide.

clear degradation after CHX treatment (Fig. 5b and Supplementary Fig. 53). Such behavioral differences have been previously observed in cytosolic CALR⁴², suggesting that arginylation of ERO1A may also have a stabilization effect against proteasomal degradation.

Biological functions associated with SSBP arginylation

SSBP matures by translocation into mitochondria after cleavage of its transit peptide (amino acids 1–16)³⁸. We did imaging on SSBP and 17R-SSBP to investigate whether arginylated SSBP resides in the mitochondria. Both species colocalized with mitochondria protein COX4 (Fig. 5c and Supplementary Fig. 54), suggesting that 17R-SSBP was translocated into mitochondria after cleavage. To assess the cytosolic degradation of SSBP with and without arginylation in the N-degron pathway, we evaluated its expression levels using a ubiquitin cleavage plasmid system. As a result, we did not observe an obvious

response toward MG132 treatment (proteasome inhibitor) for both SSBP species (Supplementary Fig. 55). CHX (protein synthesis inhibitor) treatment reduced the expression levels of both SSBP and 17R-SSBP compared with cells without CHX treatment, indicating that ribosomal synthesis was contributing to the cellular expression levels of both species. The 17R-SSBP expression levels were higher than WT SSBP with and without CHX treatment, indicating a slower turnover of 17R-SSBP than SSBP. The results indicate that the major role of SSBP arginylation might not be degradation (Supplementary Fig. 55).

We then turned our attention to mitochondrial SSBP. SSBP KO in mouse models is embryonic lethal whereas conditional KO in the heart results in cardiomyopathy and reduced life span (~18 weeks) accompanied by heart respiratory chain deficiency⁴³. It is worth noting that ATE1 KO is also embryonic lethal with heart defects¹, being similar to SSBP KO. Thus, arginylation of SSBP may regulate cells through

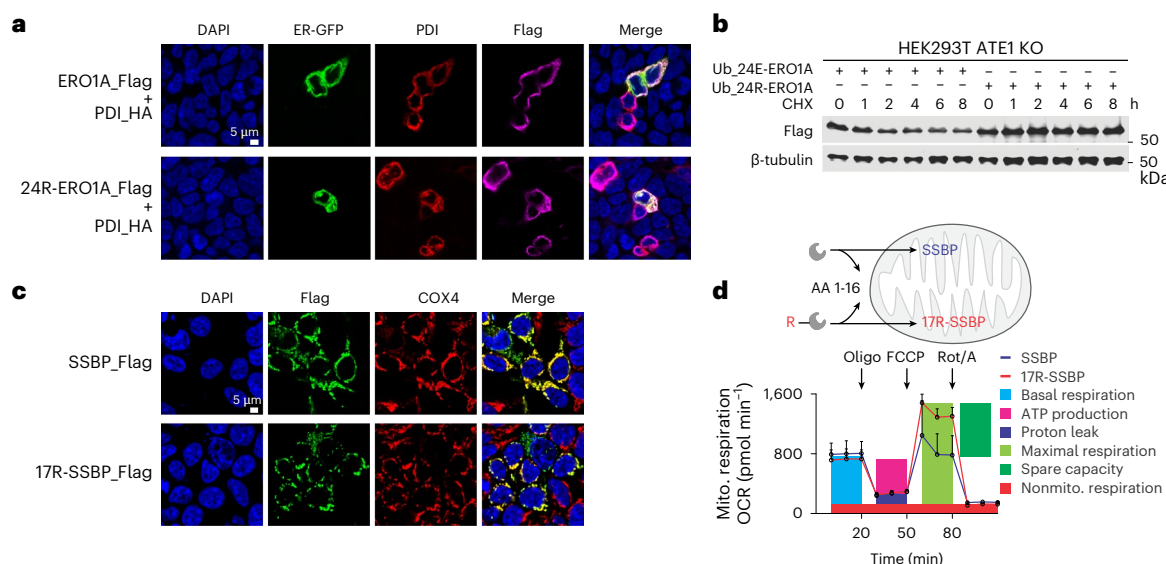


Fig. 5 | Biological functions of representative arginylation sites. a, Imaging of ERO1A and 24R-ERO1A compared with endoplasmic reticulum marker and cooverexpressed PDI. **b**, Expression levels of cytosolic ERO1A species in the CHX chase experiment. **c**, Imaging of SSBP and 17R-SSBP compared with mitochondria protein COX4. **d**, Bioenergetic profiles of HEK293T cells after SSBP and 17R-SSBP transfection measured by Seahorse XF24. Data are presented as mean \pm s.d. ($n = 5$ biological replicates). Mitochondria (mito.) respiration was analyzed

with basal respiration, ATP production, proton leak, maximal respiration, spare capacity and nonmitochondria (Nonmito.) respiration. The average values of three Seahorse measurements were used for comparison. DAPI, 4,6-diamidino-2-phenylindole; ER, endoplasmic reticulum; FCCP, carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone; Rot/A, rotenone/antimycin A; OCR, oxygen consumption rate.

mitochondria functions. Considering the importance of mitochondria in energy generation and heart functions, we investigated the functional differences between SSBP and 17R-SSBP on cellular bioenergetics (Fig. 5d) based on a previous study on SSBP mutants⁴⁴. The Seahorse XF24 Cell Mito Stress Test was performed to measure the arginylation effects of SSBP on mitochondrial respiration before and after sequential addition of oligomycin (oligo), carbonyl cyanide-4-(trifluoromethoxy) phenylhydrazone and rotenone/antimycin A. Analysis of cellular oxygen consumption rate showed that HEK293T cells overexpressing 17R-SSBP displayed comparable basal respiration, ATP production and nonmitochondria respiration than WT SSBP (Fig. 5d). 17R-SSBP resulted in significantly higher maximal respiration ($P = 0.0104$) and spare capacity ($P = 0.0076$) than cells overexpressing WT SSBP (Fig. 5d, Supplementary Fig. S6 and Supplementary Data 5). The relative ratios of maximal respiration and spare capacity in 17R-SSBP are 1.6 and 8.5 compared with WT SSBP (Supplementary Fig. S6), indicating improved mitochondria respiration potential after SSBP arginylation. 17V-SSBP did not display these effects on mitochondria when compared with WT SSBP (Supplementary Data 6).

Arginylation profiling in mouse tissues

We next applied our profiling platform to mouse tissues. We acknowledge that it is not ideal to use human ATE1 to arginylate mouse proteomes, although such a cross-species approach has been previously used¹⁶. Results from three tissues (lung, heart and brain) showed that the platform is applicable to arginylation discovery from mouse proteome (Supplementary Data 7). A total of 14 sites from 69 MS1 pairs were identified (Supplementary Fig. S7). CALR D18 and A1AT1 E25 are the two shared arginylation sites across three tissues. We noticed that the mouse A1AT1 E25 site was identified by a different peptide sequence (EDVQETDTSQK) (Supplementary Fig. S8) from the human A1AT E25 site (EDPQGDAQK) (Supplementary Fig. S46). In addition, human CALR E18 and mouse CALR D18 sites were also identified from species-dependent sequences (EPAYVFK and DPAIYFK, respectively) (Supplementary Fig. S59), indicating that mouse and human sites can be used to cross-validate when overlapped.

Establishing a public database for arginylation sites

The fact that our high-frequency hits (CALR, PDIA and TAU and others)^{25,40} are among the most studied arginylation proteins instilled confidence to follow the biology of new sites. This represents a major advantage of our method compared to existing methods^{21,28}. We understand that we alone cannot follow up on all targets or sites, a public database (Extended Data Fig. 1) was thus formulated to allow everyone to examine our data and follow the biology. The isotopic MS1 and MS2 scans from standard peptides, peptide mixtures, proteins, human proteomes and mouse proteomes are included on the website.

Discussion

ATE1 is the only enzyme known to catalyze arginylation in mammalian systems, it has two isoforms in humans (ATE1-1 and ATE1-2) and four isoforms in mice (ATE1-1, ATE1-2, ATE1-3, ATE1-4). (Supplementary Fig. S60). The expression levels and enzymatic functions vary among different isoforms^{24,45}. While our work mostly focuses on establishing a general proteomic platform using human ATE1-1, using other isoforms is a necessary next step to uncover the ATE1 substrates comprehensively. We preliminarily compared human ATE1-1 and ATE1-2 for arginylation in two samples. The isoforms showed unique and shared sites, more arginylation sites were revealed by ATE1-1 than ATE1-2 for both samples (Supplementary Figs. S61 and S62), indicating that ATE1-1 may have a slightly broader substrate scope. Previous studies showed that the expression of ATE1 isoforms is tissue specific and cell dependent^{24,45}, thus the expression levels might be an important consideration when choosing ATE1 isoforms for arginylation.

Arginylation was initially believed to occur only on exposed N-terminal Asp, Glu and Cys residues in the N-degron pathway⁴⁶. Later studies identified proteins (for example, α -synuclein¹⁶) potentially arginylated on their side chain²⁷, resulting from a minor catalytic mechanism of ATE1. Our search algorithm considered both N-terminal and side-chain arginylation situations (Supplementary Fig. S26). The resulting sites from peptide mixture (Fig. 2) and proteomes (Fig. 3) are in agreement with the fact that ATE1 predominantly arginylates N-terminal C (oxidized), D and E residues, while other residues and

sidechains are less preferably arginylated. Since the ATE1 assay was performed on proteomes, we believe most sites were arginylated at protein N-termini (Fig. 3 and Supplementary Data 3). A preliminary analysis of the P1 residues in arginylated peptides showed potential protease cleavage after every amino acid (Supplementary Fig. 63 and Supplementary Data 3). Cleavages at K and R sites accounting for 18.6% and 18.2%, respectively, might be produced from endopeptidases such as trypsin-like proteases; in contrast, how cleavages at A (13.3%) happened remains unanswered. While some sites (for example, CALR E18, ERO1A E24 and SSBP E18) were known to be protein N-terminal, many sites (for example, TAU N644) are left unexplained as to how the parent proteins were cleaved. Previous studies have offered some insights into protease cleavage (for example, calpains for TAU E3 (ref. 40) and caspase 3 for CDC6 D101 (ref. 25); however, the cleavage products of more than 600 proteases remain elusive. Great efforts are needed to understand the interplay between proteolytic cleavage and arginylation.

Proteomic discovery of endogenous arginylation is challenging^{21,28}. This is largely due to (1) the mass difference (+156 Da) introduced by arginylation being the same as the normal arginyl residue, making label-free proteomics biased to predefined arginylation sites by nature; and (2) efficient enrichment methods are lacking. The fact that arginyl-tRNA is the same source of arginylation and translation makes unbiased endogenous arginylation discovery almost impossible since one cannot exclude ribosomal incorporation when cells are alive¹³. Our initial effort using Arg¹⁰ and Arg⁰ for pulse labeling of arginylation in HEK293T cells yielded a putative list of sites (Supplementary Fig. 63). Further efforts using CHX to stop protein synthesis, MG132 to accumulate arginylation and peptide fractionation to boost peptide IDs did not seem to increase our confidence. Arginylation identification from a proteome is not ideal. Our efforts on in vivo strategy also suggested that endogenous arginylation should have minimal contribution to the light (Arg⁰) labeling in our ex vivo approach. Previous studies tried to use anti-arginylation antibodies to enrich endogenous arginylation^{21,28}; however, these efforts by design were biased toward nondegradative ATE1 targets and high-abundance proteins. Sites from those studies showed nearly no overlap with those previously characterized in ref. 13. We thus turned our attention to an ex vivo approach using lysates, which are less physiologically relevant than cells or tissues. Arginylated proteins ex vivo were not localized correctly in cell compartments, thus these may be false-positive hits. However, this method, due to its unbiased nature, represents a superior strategy for arginylation discovery.

This method relies on deep peptide fractionation because intact peptides that are not of interest (accounting for >99.99%) are interfering signals in proteomics. The reduced complexity of peptide fraction after fractionation is key to identifying arginylated pairs from proteomes. The additional charge in ionization (Supplementary Fig. 9) and polarity in chromatography of arginylated peptides (Fig. 1c) may help validate peptide identity using LC-centric prediction tools such as DeepLC (Supplementary Data 8)⁴⁷ and Chronologer⁴⁸. The sites revealed by our method could be dependent on the expression levels of the proteins. We foresaw that enrichment is essential to achieve improved sensitivity in arginylation detection. Since ATE1 mainly arginylates N-termini, some technologies targeting N-terminal peptides might be helpful for arginylation enrichment. Examples are terminal amine isotopic labeling of substrates technology⁴⁹ and the integrated N-terminal peptide enrichment method⁵⁰, which are currently under our investigation. The anti-arginylation antibody enrichment of in vivo arginylated proteins is also under consideration. However, antibody enrichment may require scaling up the assay to milligram scale from the current 20-μg scale. Looking ahead, combining this strategy with targeted enrichment techniques and in vivo validation should substantially advance our understanding of arginylation's functional roles in biological contexts.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-025-01996-z>.

References

1. Kwon, Y. T. et al. An essential role of N-terminal arginylation in cardiovascular development. *Science* **297**, 96–99 (2002).
2. Lian, L. et al. Loss of ATE1-mediated arginylation leads to impaired platelet myosin phosphorylation, clot retraction, and in vivo thrombosis formation. *Haematologica* **99**, 554–560 (2014).
3. Kurosaka, S. et al. Arginylation regulates myofibrils to maintain heart function and prevent dilated cardiomyopathy. *J. Mol. Cell Cardiol.* **53**, 333–341 (2012).
4. Singh, K. et al. Arginyltransferase knockdown attenuates cardiac hypertrophy and fibrosis through TAK1-JNK1/2 pathway. *Sci. Rep.* **10**, 598 (2020).
5. Kurosaka, S. et al. Arginylation-dependent neural crest cell migration is essential for mouse development. *PLoS Genet.* **6**, e1000878 (2010).
6. Leite, F. S. et al. Reduced passive force in skeletal muscles lacking protein arginylation. *Am. J. Physiol. Cell Physiol.* **310**, C127–135 (2016).
7. Karakozova, M. et al. Arginylation of beta-actin regulates actin cytoskeleton and cell motility. *Science* **313**, 192–196 (2006).
8. Zhang, F., Saha, S., Shabalina, S. A. & Kashina, A. Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science* **329**, 1534–1537 (2010).
9. Carpio, M. A., Lopez Sambrooks, C., Durand, E. S. & Hallak, M. E. The arginylation-dependent association of calreticulin with stress granules is regulated by calcium. *Biochem. J.* **429**, 63–72 (2010).
10. Decca, M. B. et al. Post-translational arginylation of calreticulin: a new isospecies of calreticulin component of stress granules. *J. Biol. Chem.* **282**, 8237–8245 (2007).
11. Hu, R. G. et al. The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature* **437**, 981–986 (2005).
12. Lee, M. J. et al. RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. *Proc. Natl Acad. Sci. USA* **102**, 15030–15035 (2005).
13. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl Acad. Sci. USA* **116**, 358–366 (2019).
14. Kim, B. H. et al. Crystal structure of the Ate1 arginyl-tRNA-protein transferase and arginylation of N-degron substrates. *Proc. Natl Acad. Sci. USA* **119**, e2209597119 (2022).
15. Bongiovanni, G., Fidelio, G. D., Barra, H. S. & Hallak, M. E. The post-translational incorporation of arginine into a beta-amyloid peptide increases the probability of alpha-helix formation. *Neuroreport* **7**, 326–328 (1995).
16. Wang, J. et al. Protein arginylation targets alpha synuclein, facilitates normal brain health, and prevents neurodegeneration. *Sci. Rep.* **7**, 11323 (2017).
17. Zhao, J. et al. alpha-Synuclein arginylation in the human brain. *Transl. Neurodegener.* **11**, 20 (2022).
18. Lopez Sambrooks, C., Carpio, M. A. & Hallak, M. E. Arginylated calreticulin at plasma membrane increases susceptibility of cells to apoptosis. *J. Biol. Chem.* **287**, 22043–22054 (2012).
19. Carpio, M. A. et al. Calreticulin-dimerization induced by post-translational arginylation is critical for stress granules scaffolding. *Int. J. Biochem. Cell Biol.* **45**, 1223–1235 (2013).
20. Saha, S. et al. Arginylation and methylation double up to regulate nuclear proteins and nuclear architecture in vivo. *Chem. Biol.* **18**, 1369–1378 (2011).

21. Wong, C. C. et al. Global analysis of posttranslational protein arginylation. *PLoS Biol.* **5**, e258 (2007).
22. Cha-Molstad, H. et al. Amino-terminal arginylation targets endoplasmic reticulum chaperone BiP for autophagy through p62 binding. *Nat. Cell Biol.* **17**, 917–929 (2015).
23. Xu, Z., Payoe, R. & Fahlman, R. P. The C-terminal proteolytic fragment of the breast cancer susceptibility type 1 protein (BRCA1) is degraded by the N-end rule pathway. *J. Biol. Chem.* **287**, 7495–7502 (2012).
24. Hu, R. G. et al. Arginyltransferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. *J. Biol. Chem.* **281**, 32559–32573 (2006).
25. Yoo, Y. D. et al. N-terminal arginylation generates a bimodal degron that modulates autophagic proteolysis. *Proc. Natl Acad. Sci. USA* **115**, E2716–E2724 (2018).
26. Davydov, I. V. & Varshavsky, A. RGS4 is arginylated and degraded by the N-end rule pathway in vitro. *J. Biol. Chem.* **275**, 22931–22941 (2000).
27. Wang, J. et al. Arginyltransferase ATE1 catalyzes midchain arginylation of proteins at side chain carboxylates in vivo. *Chem. Biol.* **21**, 331–337 (2014).
28. MacTaggart, B. et al. Global analysis of post-translational side-chain arginylation using pan-arginylation antibodies. *Mol. Cell Proteom.* **22**, 100664 (2023).
29. Wang, J., Yates, J. R. 3rd & Kashina, A. Biochemical analysis of protein arginylation. *Methods Enzymol.* **626**, 89–113 (2019).
30. Wadas, B., Piatkov, K. I., Brower, C. S. & Varshavsky, A. Analyzing N-terminal arginylation through the use of peptide arrays and degradation assays. *J. Biol. Chem.* **291**, 20976–20992 (2016).
31. Wang, J. et al. Target site specificity and in vivo complexity of the mammalian arginylome. *Sci. Rep.* **8**, 16177 (2018).
32. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
33. O'Shea, J. P. et al. pLogo: a probabilistic approach to visualizing sequence motifs. *Nat. Methods* **10**, 1211–1212 (2013).
34. Tan, K. et al. Mitochondrial SSBP1 protects cells from proteotoxic stresses by potentiating stress-induced HSF1 transcriptional activity. *Nat. Commun.* **6**, 6580 (2015).
35. Lan, X. et al. Oligomerization and a distinct tRNA-binding loop are important regulators of human arginyl-transferase function. *Nat. Commun.* **15**, 6350 (2024).
36. Watanabe, K. et al. Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. *Nature* **449**, 867–871 (2007).
37. Pagani, M., Pilati, S., Bertoli, G., Valsasina, B. & Sitia, R. The C-terminal domain of yeast Ero1p mediates membrane localization and is essential for function. *FEBS Lett.* **508**, 117–120 (2001).
38. Xu, G., Shin, S. B. & Jaffrey, S. R. Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc. Natl Acad. Sci. USA* **106**, 19310–19315 (2009).
39. Long, G. L., Chandra, T., Woo, S. L., Davie, E. W. & Kurachi, K. Complete sequence of the cDNA for human alpha 1-antitrypsin and the gene for the S variant. *Biochemistry* **23**, 4828–4837 (1984).
40. Brower, C. S., Piatkov, K. I. & Varshavsky, A. Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway. *Mol. Cell* **50**, 161–171 (2013).
41. Zhang, Y. et al. Targeting the functional interplay between endoplasmic reticulum oxidoreductin-1alpha and protein disulfide isomerase suppresses the progression of cervical cancer. *EBioMed.* **41**, 408–419 (2019).
42. Goitea, V. E. & Hallak, M. E. Calreticulin and arginylated calreticulin have different susceptibilities to proteasomal degradation. *J. Biol. Chem.* **290**, 16403–16414 (2015).
43. Jiang, M. et al. The mitochondrial single-stranded DNA binding protein is essential for initiation of mtDNA replication. *Sci. Adv.* **7**, eabf8631 (2021).
44. Del Dotto, V. et al. SSBP1 mutations cause mtDNA depletion underlying a complex optic atrophy disorder. *J. Clin. Invest.* **130**, 108–125 (2020).
45. Rai, R. & Kashina, A. Identification of mammalian arginyltransferases that modify a specific subset of protein substrates. *Proc. Natl Acad. Sci. USA* **102**, 10123–10128 (2005).
46. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186 (1986).
47. Bouwmeester, R., Gabriels, R., Hulstaert, N., Martens, L. & Degroove, S. DeepLC can predict retention times for peptides that carry as-yet unseen modifications. *Nat. Methods* **18**, 1363–1369 (2021).
48. Wilburn, D. B. et al. Deep learning from harmonized peptide libraries enables retention time prediction of diverse post translational modifications. Preprint at *bioRxiv* (2023).
49. Kleifeld, O. et al. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28**, 281–288 (2010).
50. Ju, S. et al. iNrich, rapid and robust method to enrich N-terminal proteome in a highly multiplexed platform. *Anal. Chem.* **92**, 6462–6469 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Methods

Expression and purification of RARS1 and ATE1 proteins

BL21 (DE3*) (Invitrogen) and BL21-CodonPlus (Agilent Technologies, catalog no. 230245) were used for expressing RARS1 and ATE1, respectively, according to a previous study²⁹. Bacteria expressing protein were grown in large cultures (1 l per flask) at 37 °C in rich Luria-Bertani medium to an optical density of 0.7–0.9 at 600 nm. The flasks were cooled rapidly on ice before induction with isopropyl β -D-1-thiogalactopyranoside at a final concentration of 1 mM. Cultures were grown at 16 °C for an additional 16–20 h and cells were collected by centrifugation (5,000g, 30 min, 4 °C), flash frozen in liquid nitrogen and stored at –80 °C. Cells were lysed in lysis buffer (50 mM Tris, 500 mM NaCl, 5 mM β -mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride, 10% glycerol, pH 7.8) at 4 °C by probe sonication. The His-tagged proteins were purified on the HisTrap HP column (Cytiva, catalog no. 17-5248-01) in the lysis buffer containing 300 mM imidazole by fast protein LC (Cytiva). Fractions containing the protein were pooled and concentrated by centrifugation to a final volume of ~1 ml. Protein was loaded to the size exclusion column, eluted by elution buffer (50 mM Tris, 100 mM NaCl, 2 mM Tris(2-carboxyethyl)phosphine (TCEP), 10% glycerol, pH 7.8) and concentrated in Millipore 30-kDa molecular weight cutoff filters. Protein concentrations were determined and 10- μ l aliquots were stored at –80 °C.

In vitro transcription and production of tRNA

A double-strand DNA template (TAATACGACTCACTATA-GGGCCAG TGGCGCAATGGATAACGCGTCTGACTACGGATCAGAAGATTCTAG GTTCGACTCTGGCTGGCTCGCCA) containing a T7 RNA polymerase recognition sequence (TAATACGACTCACTATA) and a 76-base tRNA^{arg} sequence⁵¹ was ordered from IDT. The in vitro transcription was carried out following the manufacturer's protocol on HiScribe T7 High Yield RNA Synthesis Kit (NEB, catalog no. E2040S) using 250 ng of DNA template for a 20- μ l reaction incubated at 37 °C for 4 h. DNase and its buffer were added (final volume 30 μ l) to remove DNA template by incubation for 30 min at 37 °C. The 170 μ l ddH₂O was added followed by 200 μ l of phenol:chloroform:isoamylalcohol pH 6.7 (Sigma, catalog no. 516726). The supernatant was added to a one-tenth volume of 3 M NaOAc and 2.5 volumes of cold 200-proof ethanol, incubated at –20 °C for 20 min and centrifuged at 4 °C. The pellet was air-dried until moist and resuspended in 20 μ l of ddH₂O. tRNA concentration was determined by NanoDrop, and 2- μ l aliquots were stored at –80 °C.

Preparation of whole-proteome peptide mixture

HEK293T (ATE1 KO) cell pellet collected from a 6-cm dish was added 200 μ l of ice-cold PBS, lysed by probe sonication (Fisherbrand Model 120 Sonic Dismembrator, settings: 1 s on, 2 s off, 20% energy, 60 cycles) to yield a homogeneous solution. Lysates were ultracentrifuged at 100,000g for 30 min at 4 °C to yield soluble and insoluble proteomes. To the insoluble fraction, 200 μ l of PBS was added and resuspended by probe sonication. Protein concentration was determined by DC assay (Bio-Rad). Then 100 μ g of soluble and insoluble proteomes were diluted in 40 μ l of 8 M urea in PBS, reduced by 5 mM of TCEP with 30 min incubation at 37 °C and alkylated by 15 mM of iodoacetamide with 30 min incubation at room temperature in the dark. The solution was diluted to 2 M urea by 50 mM ammonium bicarbonate in H₂O, digested by trypsin (sequence grade, Promega) at 1:50 trypsin:protein ratio (w:w) with overnight (–12 h) incubation at 37 °C. The resulting peptide solutions were acidified by formic acid (FA) at a final concentration of 5%, desalted and resuspended in H₂O for ATE1 assay.

Preparation of cell and tissue proteomes

Ice-cold PBS (200 μ l and 400 μ l for pellet from a 6-cm and 10-cm dish, respectively) was added to cell pellets, which were then lysed by probe sonication (Fisherbrand Model 120 Sonic Dismembrator, settings were 1 s on, 2 s off, 20% energy, 60 cycles) to yield a homogeneous solution.

A few milligrams (5–10 mg) of tissue pieces and samples (human or mouse) were transferred to tubes (Thermo Scientific, catalog no. 3468) and ice-cold PBS was added (200–400 μ l). A scoop of beads (GB05 beads for the brain and ZRB05 beads for other tissues, Next Advance, Inc.) was added to each tissue. Samples were homogenized in the cold room using FastPrep-24 Classic bead beating grinder (FastPrep-24, MP Biomedicals) at default settings (4.0 m s^{–1}, 20 s, 2 cycles). Lysates were ultracentrifuged at 100,000g for 30 min at 4 °C to yield soluble and insoluble proteomes. To the insoluble fraction, the same volume of PBS as the soluble fraction was added to resuspend by probe sonication until homogeneity (5–10 cycles). Protein concentration was determined by DC assay (Bio-Rad).

Affinity purification of Halo-tagged CALR

C-terminal Halo-tagged CALR was cloned into the pcDNA5/FRT vector (Invitrogen) for mammalian expression⁵². The CALR coding sequence was amplified from clone ID OHu23892 (GenScript). The C-terminal version of the HaloTag was amplified from pFC14A (Promega). 18R-CALR was generated by inserting an Arg codon after the signal peptide of CALR using mutagenesis. WT or ATE1^{–/–} HEK293T cells were transfected with plasmids expressing Halo-tagged WT or 18R-CALR using FuGENE HD (Promega). For experiments where ATE1 was coexpressed, various amounts of the plasmid expressing Flag-tagged ATE1 were also transfected together with CALR. After 2 days of transfection, cells were gathered and pellets were snap-frozen with liquid nitrogen. Halo purification was performed according to the manual of HaloTag Mammalian Pull-Down Systems (Promega). Briefly, thawed cell pellets were lysed with Mammalian Lysis Buffer with the addition of protease inhibitors (Promega). Crude lysates were centrifuged at 20,000g at 4 °C for 10 min, and supernatants were collected to bind with Magne HaloTag Beads (Promega). After rotating at 4 °C overnight, beads were washed with cold Wash Buffer five times, then eluted with AcTEV Protease (Invitrogen) at room temperature for 1 h with shaking. Eluates were subjected to sample processing for bottom-up mass spectrometry analysis.

Isotopic labeling by ATE1 arginylation assay

The ATE1 assay was adapted from a previous study²⁹. The assay was set up on ice by mixing a 20- μ l reaction containing 1 \times reaction buffer, 2 mM Arg (Arg¹⁰ and Arg⁰, respectively), 2 mM ATP, 3 μ M tRNA^{Arg}, 1 μ M RARS1, 3 μ M human ATE1 and 1 μ g μ l^{–1} substrate. The 5 \times reaction buffer consists of 50 mM HEPES, 30 mM KCl and 10 mM MgCl₂. The substrate can be peptide, peptide mixture, protein or proteome. A total of 20 μ g of substrate was always used unless the protein substrate was purified from overexpression and pull-down experiments where protein concentration was too low to quantify. The amount of commercial CALR was 3.5 μ g per reaction. The concentration of standard peptide was 100 μ M unless otherwise noted. Volumes of ddH₂O were adjusted every time depending on the volumes of other reagents. The pair reactions were incubated at 37 °C for 30 min for peptides, 1 h for proteins and 2 h for proteomes. The pair reactions were mixed 1:1 for sample preparation or –80 °C storage.

Sample preparation after ATE1 arginylation assay

The peptide sample was acidified by FA to a final concentration of 5%, desalted, dried by SpeedVac and reconstituted for LC–MS analysis. Whole-proteome peptide sample was acidified by FA to a final concentration of 5%, desalted, dried by SpeedVac and reconstituted for peptide fractionation. Protein or proteome sample was added accurately 20 mg urea (final concentration 6 M) to denature proteins, TCEP (final concentration 10 mM) for 30 min incubation at 37 °C to reduce disulfides and iodoacetamide (final concentration 30 mM, final volume 54 μ l) for 30 min at room temperature protected from light to alkylate reduced thiols. The solution was diluted to 2 M urea by 50 mM ammonium bicarbonate in H₂O, digested by trypsin (sequence grade,

Promega) at a 1:50 trypsin:protein ratio (w:w) with overnight (~12 h) incubation at 37 °C. The resulting peptide solution was acidified by FA at a final concentration of 5%, desalted, dried by SpeedVac and reconstituted for peptide fractionation.

Peptide desalting

Peptides were resuspended in 0.1% FA and desalted by in-house packed stage tips. Stage tips were self-prepared by sealing five disks of C18 material (cat. no. 2315, Empore, 3 M Company) at the bottom of a P200 tip. C18 disks were cut by sample corers (cat. no. 18035-01 for peptides <20 µg, cat. no. 18035-02 for 20–100 µg peptides, Fine Science Tools). Stage tips were equilibrated with 50 µl of methanol, 50 µl of 80% acetonitrile in H₂O containing 0.1% FA and 50 µl of water containing 0.1% FA by centrifugation (1,000g, ~1–2 min). The sample was loaded to the stage-tip and centrifuged to flow through, washed with 75 µl of water containing 0.1% FA. The stage-tip was transferred to a new collection tube, eluted by 2 × 75 µl of 80% acetonitrile in H₂O containing 0.1% FA. The sample was dried by SpeedVac (SAVANT SVC100H Refrigerated Condensation Trap) under vacuum for 30–60 min at room temperature, stored at –80 °C.

Peptide fractionation

Peptides (20 µg, half material from ATE1 assay) were fractionated before proteomics analysis. For high-pH fractionation, peptides were resuspended in 50 µl ammonium formate (pH 10). Stage tips were self-prepared by sealing five disks of C18 material (cat. no. 2315, Empore, 3 M Company) at the bottom of a P200 tip. C18 disks were cut by sample corers (cat. no. 18035-02, Fine Science Tools). Stage tips were equilibrated with 50 µl of methanol, 50 µl of 80% acetonitrile in H₂O containing 0.1% FA and 50 µl of ammonium formate pH 10 by centrifugation (1,000g, ~1–2 min). The sample was loaded to the stage-tip and centrifuged to flow through (fraction 1), then eluted by 20 µl of buffers consisting of MeCN (1–28% and 80%) and ammonium formate (pH 10) to yield an additional 29 fractions. The fractions were dried by SpeedVac (SAVANT SVC100H Refrigerated Condensation Trap) under vacuum for 30–60 min at room temperature, and stored at –80 °C. Each fraction was resuspended in 20 µl of 2% MeCN in 0.1% FA and 5 µl were injected (~167 ng per fraction in average) for analysis.

LC–MS for proteomics analysis

A Vanquish Neo UHPLC (for ultrahigh-performance LC) was coupled to an Exploris 240 or Orbitrap Ascend (Thermo Scientific). Peptide samples were maintained at 7 °C on the sample tray in the LC system. Separation of peptides was carried out on an Easy-Spray PepMap Neo nano-column (2 µm, C18, 75 µm × 150 mm) at room temperature with a mobile phase consisting of a linear gradient of A (0.1% FA in H₂O) and B (acetonitrile containing 0.1% FA) under the following conditions: 0 → 80 → 83 → 90 min, 0% → 27% → 100% → 100% B. The flow rate was 300 nl min^{–1}. The injection volume is 5 µl. The voltage applied to the nano-LC electrospray ionization source was 1.9 kV. The temperature of the ion transfer tube (ITC) was set at 275 °C. Spectra were collected in a data-dependent acquisition mode with MS1 scan range of *m/z* 350–2,000 in the Orbitrap. For Exploris, the top 20 most intense peaks from a single high-resolution (120,000) full MS spectrum of parent ions were fragmented for MS2 spectra. For Ascend, the time of a scan cycle is 3 s. Parent ions assigned as peptides in charge states +2–5 with an intensity higher than 1×10^4 were included for fragmentation. High-energy collisional dissociation-induced fragmentation (MS2) scans were recorded in the Orbitrap (scan from *m/z* 150–1,500). Dynamic exclusion was set as a repeat count of 1 within an exclusion time of 20 s. All other parameters were left as default values.

A Waters M5 UHPLC was coupled to a ZenoTOF 7600 (Sciex) in positive data-dependent mode. Peptide samples were maintained at 7 °C on the sample tray in LC. Separation of peptides was carried out on an Waters nanoEase M/Z Symmetry C18 Analytical Column

(5 µm, 100 Å, 300 µm × 150 mm) at room temperature with a mobile phase consisting of a linear gradient of A (0.1% FA in H₂O) and B (acetonitrile containing 0.1% FA) under the following conditions: 0 → 1 → 46 → 47 → 50 → 50.5 → 55 min, 2 → 2 → 32 → 80 → 80 → 2 → 2% B. The flow rate was set at 5 µl min^{–1}. The ZenoTOF 7600 system was operated using the OptiFlow TurboV ion source with a vertical microflow probe (1–50 µl min^{–1} electrode). The ionization window was 2.5–48.5 min. Source and gas parameters were 20 psi and 60 psi for ion source gases 1 and 2, respectively; 35 psi and 7 psi for curtain and collision gas, respectively; and 200 °C and 35 °C for temperature and column temperature, respectively. The spray voltage was 5 kV. The MS1 mass range was 350–1,500 Da with an accumulation time of 0.1 s, declustering potential of 80 V and collision energy of 10 V. The top 45 monoisotopic ions assigned as peptides in charge states +2–6 with an intensity higher than 100 cps were included for fragmentation. Dynamic exclusion was set as a repeat count of 1 within an exclusion time of 7 s. MS2 collision-induced dissociation fragmentation with Zeno trapping was carried out over a mass range of 120–1,600 Da with an accumulation time of 0.02 s, declustering potential of 80 V, time bins to a sum of 8 and a Zeno threshold of 1×10^5 cps. All other parameters were left as default values.

Search parameters using Byonic software

Each data file (in ‘raw’ format) was generated by the instrument (Xcalibur software), and searched using the Byonic software v.4.5.2 (Protein Metrics) against a reverse-concatenated, nonredundant database of the human proteome (20,398 proteins in total from UniProt). Peptides were required to have at least a tryptic C-terminus and up to one missed cleavage was allowed in the database search. The mass tolerances of precursor and product ions were set to 10 ppm and 20 ppm, respectively. Carbamidomethylation (+57.021464 Da) on cysteine residues was allowed for two per peptide as a ‘common’ variable modification in the Byonic parameter file. Oxidation (+15.99492 Da) on methionine residues was allowed for two per peptide as a ‘common’ variable modification. Total ‘common’ modifications were limited to up to four per peptide. Arginylation (+156.10111 Da as Arg⁰ modification or +166.10937 Da as Arg¹⁰ modification) on peptide N-terminal, Asp and Glu residues, arginylation after deamidation (+157.08513 Da as Arg⁰ modification or +167.09339 Da as Arg¹⁰ modification) on Asn and Gln residues, arginylation after di-oxidation (+188.09094 Da as Arg⁰ modification or +198.09920 Da as Arg¹⁰ modification) or tri-oxidation (+204.08585 Da as Arg⁰ modification or +214.09411 Da as Arg¹⁰ modification) on Cysteine residues was allowed for one per peptide as a ‘rare’ modification. Total rare modifications were limited to one per peptide. Peptides with two or more arginylation modifications were not included in the search. The false-positive rate was set at 1% or lower.

ArginylomePlot software determines arginylation sites after Byonic

Custom software based on RStudio (version 2023.06.0 Build 421) script, ArginylomePlot, was used to process all MS1 spectra of arginylated peptide and peptide IDs in Excel files generated from Byonic search. The software is publicly available to download from GitHub at <https://github.com/BeckyHan/Garcia-Lab/tree/main/ArginylomePlot>. A copy of the software together with a demo dataset is provided in the Supplementary Data 1. Briefly, .raw files from fractions of a sample were converted to .mzXML files using RawConverter (version 1.1.0.23) with monoisotopic selection (2015 released, publicly available at <http://fields.scripps.edu/rawconv>). The software first combines all peptide IDs into a single file, then extracts all coeluting MS2 spectra with a mass difference of 10.008269 Da indicating copresence of Arg¹⁰ and Arg⁰ modifications on the same peptide. Using the average retention time of all MS2 spectra from each arginylated peptide, the software then goes to a specific mzXML file and extracts their matching MS1 scans (doublet, quartet or sextet matching with a threshold of 10 ppm for

each m/z value) within a time window (± 1.25 min of average retention time). The H/L ratio was calculated based on each MS1 scan, and the number of MS1 scans of each peptide was exported with respective ratios (doublet, quartet, or sextet). Failure to detect matching doublet MS1 will result in the exclusion of an MS2 pair. A ratio summary in a box plot (duodecet, '6 + 6' peaks) was generated for each arginylation site with a unique peptide sequence, charge and raw. A summary table for each sample was generated with listed information on protein, unmodified peptide sequence, site, Arg¹⁰/Arg⁰ modified peptide ID, charge, raw, Byonic scores of Arg¹⁰/Arg⁰ modified peptides, H/L ratios of doublet, quartet or sextet MS1 scans, and their exact numbers of MS1 scans.

The Supplementary Methods section is provided in the Supplementary Information.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The mass spectra RAW data are publicly available online at <https://massive.ucsd.edu> under the accessions MSV000097195, MSV000097197 and MSV000097196. The RAW file index is listed in Supplementary Data 9. Arginylation site data are publicly available to view and download from www.arginylation.com. Source data are provided with this paper.

Code availability

Python codes for website construction are available via Zenodo at <https://doi.org/10.5281/zenodo.15770600> (ref. 53). R codes for ArginylomePlot software are available via Zenodo at <https://doi.org/10.5281/zenodo.15742920> (ref. 54). The arginylation website is publicly accessible at www.arginylation.com for data visualization and download.

References

- Gogakos, T. et al. Characterizing expression and processing of precursor and mature human tRNAs by hydro-tRNAseq and PAR-CLIP. *Cell Rep.* **20**, 1463–1475 (2017).
- Searfoss, R. M., Liu, X., Garcia, B. A. & Lin, Z. Top-down proteomics for the characterization and quantification of calreticulin arginylation. *Anal. Chem.* **97**, 15562–15569 (2025).
- Zhao, C. Source code of arginylation interactive website. *Zenodo* <https://doi.org/10.5281/zenodo.15770600> (2025).
- Han, B. ArginylomePlot. *Zenodo* <https://doi.org/10.5281/zenodo.15742920> (2025).

Acknowledgements

This work is partially supported by NIH grant nos. R01NS111997 (B.A.G.), R01HD106051 (B.A.G.), R01HL177113 (B.A.G., Z.L. and D.L.), R21CA292191 (Z.L. and D.L.), R35GM122505 (A.K.), R01NS102435 (A.K.), R01HL141086 (M.J.G.), R21CA286307 (D.L.) and R35GM150678 (Y.Z.), NSF grant no. CHE2127882 (B.A.G.), WUSM ICTS JIT grant no. JIT1181 (Z.L.), Children's Discovery Institute of Washington University

and St. Louis Children's Hospital grant no. PM-LI-2019-829 (M.J.G.), the American Heart Association grant no. 970198 (M.J.G.), Mays Cancer Center Early Career Pilot Award grant no. P30 CA054174 (D.L.) and Research Education Component (REC) through an NIA grant no. P30AG066444 (Z.L.). We thank N. Niemi (WashU) for guidance on the Seahorse assay, X. Liu and G. Peng (WashU) for providing mouse tissues, and Washington University Diabetes Research Center supported by NIH grant no. 5P30DK020579.

Author contributions

Z.L. and B.A.G. conceived the project. Z.L. performed arginylation assays and cellular functional studies and oversaw the experiments and manuscript revisions. Z.L. and Y.X. set up the experiment design and proteomics search parameters. J.G. performed peptide and HEK293T arginylation experiments. X. Liu performed CALR imaging and pull-down, and validated ATE1 KO by cloning. E.Z. ran fractionated samples. B.B.P. searched and organized proteomics data. D.H.R. performed ERO1A and SSBP imaging and pull-down. R.M.S. performed top-down experiments. F.N.V. performed JcM assay and IP-MS of ERO1A and SSBP mutants. F.N.V. and R.K. performed cell replicates on arginylation. G.P.D. produced tRNA and performed the arginylation assay on the peptide. B.M. and A.K. provided brain samples. X. Lan performed the in-bacteria assay, and D.F. prepared HeLa samples under the supervision of Y.Z. L.G. performed induced pluripotent stem cell culture under the supervision of M.J.G. K.J.L. provided heart tissue. C.Z. wrote Python code and established the website. X.H. wrote R codes and ArginylomePlot software. D.L. performed the sequence logo. Z.L. drafted the paper. B.A.G. and D.H.R. proofread the draft. All authors wrote, revised and approved the paper.

Competing interests

Z.L., D.L. and B.A.G. are cofounders of LasNova Therapeutics, LLC. B.A.G. is paid to be on the advisory board for Quantum Si. The remaining authors declare no competing interests.

Additional information

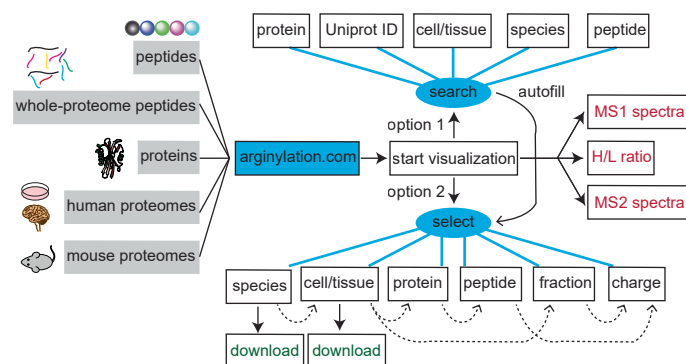
Extended data is available for this paper at <https://doi.org/10.1038/s41589-025-01996-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-025-01996-z>.

Correspondence and requests for materials should be addressed to Zongtao Lin or Benjamin A. Garcia.

Peer review information *Nature Chemical Biology* thanks Oleg Krokhin and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Design and flowchart of the arginylation database and its website. Arginylation data are from peptides, whole-proteome peptides, proteins, human proteomes, and mouse proteomes. Visualization includes

isotopic ratios and annotated isotopic MS1 and MS2 spectra. All indexed MS1 and MS2 scans are accessible to the public to download. A dashed arrow means a one-to-multiple inclusion relationship.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	LI-COR Biosciences v1.0.0.55 ZEN Blue v2.3 Wave v2.6.3
Data analysis	Byonic software v4.5.2 WebLogo v2.8.2 ZEN Blue v2.3 Skyline 64-bit v22.2.0.527 FLASHDeconv version OpenMS-3.1.0-pre-FDdevelop-2024-04-11-Wind64 ProsightLite v1.4 ArginylomePlot v12.3 (https://github.com/BeckyHan/Garcia-Lab/blob/main/ArginylomePlot/ArginylomePlot_v12.3.Rmd) RStudio version 2023.06.0 Build 421 Python v3.11.7 codes for website www.arginylation.com (https://github.com/ChenfengZhao/Arginylation)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The mass spectra RAW data are publicly available online at <https://massive.ucsd.edu> under the accessions MSV000097195, MSV000097197, and MSV000097196. The RAW file index is listed in Supplementary Dataset 9. Arginylation site data are publicly available to view and download from www.arginylation.com. Data supporting the findings in this study are available within the paper.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

All frozen postmortem tissues were lysed to solutions for arginylation site discovery, thus we did not consider sex and gender.

Population characteristics

Brains

INDDID Sex Age NPDx1
117849 Male 57 Normal
117504 Male 59 Normal
112568 Male 71 PD
108180 Male 59 PDD
110007 female 88 AD

Heart

TCBR ID Sex Age Diagnosis
T1223 Male 40 NICM/idiopathic DCM. No prior cardiac surgeries.

Only de-identified postmortem human tissue samples were used. Details of human samples are provided in Materials section in Supplementary Information.

Recruitment

NA

Ethics oversight

IRB# 829668 approved by University of Pennsylvania Institutional Review Board on 19/04/2018.
IRB# 201104172 approved by Washington University IRB on 12/20/2021

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical tests were performed to determine the sample size. Each type of sample was used for arginylation identification but not quantification.

Data exclusions

No data is excluded.

Replication

All replication information are provided in figure legend. Different sample types were used for arginylation identification but not quantification.

Randomization

Experiments were not randomized. Different sample types were used for arginylation identification but not quantification.

Blinding

NA. Different sample types were used for arginylation identification but not quantification.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

anti-FLAG antibody, Sigma-Aldrich, catalog: F3165, clone: M2 monoclonal, lot: 0000278731
 β-tubulin antibody, Cell Signaling, catalog: 2146, clone: NA, lot: 10
 IRDye 680RD goat anti-mouse antibody, LI-COR, catalog: 925-68070, clone: polyclonal, lot: D30419-15
 IRDye 800CW goat anti-rabbit antibody, LI-COR, catalog: 925-32211, clone: polyclonal, lot: D30425-15
 Anti-ATE1 antibody, Sigma-Aldrich, catalog: MABS436, clone: 6F11, lot: 2861669
 Anti-P4HB antibody, Abcam, catalog: ab2792, clone: RL90, lot: NA
 anti-Myc antibody, Sigma-Aldrich, catalog: 05-419, clone: 9E10 mouse monoclonal, lot: 4001527
 anti-HA antibody, Cell Signaling, catalog: 3724S, clone: C29F4, lot: NA
 anti-mouse IR 800CW antibody, LI-COR, catalog: 925-32210, clone: polyclonal, lot: D40206-15
 anti-rabbit IR 680RD antibody, LI-COR, catalog: 926-68071, clone: polyclonal, lot: D40109-05
 anti-mouse alexa fluor 647 antibody, Thermo Scientific, catalog: A-21235, clone: polyclonal, lot: 2674387
 anti-rabbit alexa fluor 568 antibody, Thermo Scientific, catalog: A-11011, clone: polyclonal, lot: 2782620

Validation

Anti-FLAG was validated for WB and IP in all species by manufacturer.
 anti-tubulin was validated for WB, IHC, IF, and F in human, mouse, rat, monkey, zebrafish, and bovine by manufacturer.
 anti-ATE1 was validated for WB and IC in human and mouse by manufacturer.
 anti-P4HB was validated for IP, EM, Flow Cyt, ELISA, WB, ICC/IF, Inhib, IHC-Fr, IHC-P and reacts with Mouse, Rat, Pig, Hamster, African green monkey, Monkey, Human, Dog samples.
 anti-Myc was validated for IC, IHC, IP and WB in all species by manufacturer.
 anti-HA was validated for WB, IP, IHC, IF, F and ChIP in all species by manufacturer.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HEK293T, MCF7, HCC2935, H1437, HeLa and A549 were purchased from ATCC.
 iPSC_CF, iPSC and iPSC_CM cells were derived from parant line (BJ fibroblast line CRL-2522, ATCC).

Authentication

Cell lines were purchased from ATCC and used directly. No authentication was performed.

Mycoplasma contamination

All cell lines were tested as mycoplasma-free.

Commonly misidentified lines (See [ICLAC](#) register)

No used cell line is among commonly misidentified lines.