

Rapid Commun. Mass Spectrom. 2014, 28, 2735–2743
(wileyonlinelibrary.com) DOI: 10.1002/rcm.7069

Enzymatic generation of peptides flanked by basic amino acids to obtain MS/MS spectra with 2× sequence coverage

H. Alexander Ehardt^{1*}, Jie Nan¹, Steven G. Chaulk², Richard P. Fahlman^{2,3} and Ruedi Aebersold^{1,4,5}

¹Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Auguste-Piccard-Hof 1, 8093 Zürich, Switzerland

²Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada

³Department of Oncology, University of Alberta, Edmonton, Alberta, Canada

⁴Competence Center for Systems Physiology and Metabolic Diseases, Zürich, Switzerland

⁵Faculty of Science, University of Zürich, Zürich, Switzerland

RATIONALE: Tandem mass (MS/MS) spectra generated by collision-induced dissociation (CID) typically lack redundant peptide sequence information in the form of e.g. b- and y-ion series due to frequent use of sequence-specific endopeptidases cleaving C- or N-terminal to Arg or Lys residues.

METHODS: Here we introduce arginyl-tRNA protein transferase (ATE, EC 2.3.2.8) for proteomics. ATE recognizes acidic amino acids or oxidized Cys at the N-terminus of a substrate peptide and conjugates an arginine from an aminoacylated tRNA^{Arg} onto the N-terminus of the substrate peptide. This enzymatic reaction is carried out under physiological conditions and, in combination with Lys-C/Asp-N double digest, results in arginylated peptides with basic amino acids on both termini.

RESULTS: We demonstrate that *in vitro* arginylation of peptides using yeast arginyl tRNA protein transferase 1 (yATE1) is a robust enzymatic reaction, specific to only modifying N-terminal acidic amino acids. Precursors originating from arginylated peptides generally have an increased protonation state compared with their non-arginylated forms. Furthermore, the product ion spectra of arginylated peptides show near complete 2× fragment ladders within the same MS/MS spectrum using commonly available electrospray ionization peptide fragmentation modes. Unexpectedly, arginylated peptides generate complete y- and c-ion series using electron transfer dissociation (ETD) despite having an internal proline residue.

CONCLUSIONS: We introduce a rapid enzymatic method to generate peptides flanked on either terminus by basic amino acids, resulting in a rich, redundant MS/MS fragment pattern. © 2014 The Authors. *Rapid Communications in Mass Spectrometry* published by John Wiley & Sons Ltd.

Peptide identification using electrospray ionization mass spectrometry (ESI-MS) is based on a series of ionization and fragmentation events. First, the peptide secures multiple protons during the gas-phase ionization process via basic sites of the analyte. These sites are the N-terminal's primary amine and basic amino acid side chains, mainly Arg, Lys, and His. Then, using various fragmentation modes, e.g. collision-induced dissociation (CID) or electron transfer dissociation (ETD), the ionizing proton(s) transfer to the peptide backbone to trigger fragmentation in a stochastic manner. Throughout the process, the *m/z* values of the

precursor and product ions are determined by the mass analyzer(s). The mobile proton hypothesis is an attempt to rationalize the empiric observation of predominant y-fragment ions when tryptic peptides are fragmented using CID.^[1] These y-fragment ions result frequently in fragment ladders, which are used by *in silico* MS/MS spectra annotation algorithms to match MS/MS spectra to a peptide database (also known as peptide spectrum matches: PSM).^[2–4] However, incomplete or non-stochastic cleavage of precursors can be reasons for missing PSM annotations. To overcome these challenges, precursor ion fragmentation has been aided by chemical labeling strategies, e.g. iTRAQ (isobaric tags for relative and absolute quantitation) or TMT (tandem mass tags).^[5,6] As Arg requires the greatest dissociation energy to mobilize the captured proton for subsequent peptide linkage cleavage, we hypothesize that precursor ion fragmentation would be even further improved if the peptide was flanked by basic amino acids on both termini.

Arginylation is an enzymatic reaction in which arginyl-tRNA protein transferase 1 (ATE1; E.C. 2.3.2.8) conjugates a single arginyl moiety from aminoacylated tRNA^{Arg} onto a target poly-peptide.^[7] ATE1 generally recognizes N-terminal

* Correspondence to: H. A. Ehardt, Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Auguste-Piccard-Hof 1, 8093 Zürich, Switzerland.
E-mail: ehhardt@imsb.biol.ethz.ch

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

aspartate (Asp), glutamate (Glu) and oxidized cysteine (Cys*).^[8] The N-terminal arginyl moiety serves as a degron (degradation signal) for ubiquitin ligases to poly-ubiquitinate the arginylated protein.^[9] The poly-ubiquitinated proteins enter the N-End Rule pathway.^[10]

Here we establish arginylation for *in vitro* labelling of peptides with N-terminal acidic amino acids. Consistent with prior knowledge, arginylated peptides flanked by basic amino acids result in rich redundant MS/MS product ion spectra using various precursor ion fragmentation modes. Arginylation carried out by yATE1 is a fast method for labelling peptides with exclusive specificity for N-terminal acidic amino acids *in vitro*. Sequence-specific proteolytic digest is best carried out using a double digest of proteins using Lys-C and Asp-N. Upon arginylation, very short peptides are detectable.

EXPERIMENTAL

yATE1 purification

Saccharomyces cerevisiae ATE1 (yATE1) was cloned into the pMAL-c2x vector (New England Biolabs, Ipswich, MA, USA) and expressed (25 °C) in BL21 Gold cells (Stratagene, now Agilent Technologies, Inc., Santa Clara, CA, USA) for 16 h after induction with 300 mM isopropyl β -D-1-thiogalactopyranoside. Pelleted cells were lysed (30 min, on ice) with lysozyme (25 mg/L bacterial culture) and subsequent sonification in lysis buffer (50 mM tris (hydroxymethyl)aminomethane, pH 7.5, 150 mM NaCl, 5 mM β -mercaptoethanol). The yATE1-maltose binding protein fusion protein was purified by affinity chromatography (4 °C) on amylose beads using lysis buffer (New England Biolabs, as per manufacturer's protocols). The eluted fusion was then purified with Q-Sepharose (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) ion-exchange chromatography with a gradient from 0.1 to 1.0 M NaCl in lysis buffer (50 mM tris (hydroxymethyl)aminomethane, pH 7.5, 150 mM NaCl, 5 mM β -mercaptoethanol). Repeat purification of yATE1 in different laboratories yielded yATE1 with similar rates of activity.

tRNAs

tRNA^{Arg} were purchased from Chemical Block (Moscow, Russia): arginine specific, lyophilized, from *Escherichia coli* (MRE 600), and arginine acceptor activity: approx. 1400 pmol per A260 unit. The amount of purified tRNA was determined by UV-spectrometry (absorbance at 260 nm).

yATE1 enzyme kinetics

As amino-acylated tRNA are labile at neutral pH *in vitro*, the aminoacylation reaction was carried out *in situ* by including the arginyl-tRNA synthase (ArgRS, EC 6.1.1.19), ATP and arginine to yield Arg-tRNA^{Arg}. Enzymatic assays were carried out in a 100 μ L reaction volume as described in detail in Ehardt *et al.*^[11] with only minor changes: the aminoacylation reaction using CCA-adding tRNA nucleotidyltransferase (EC 2.7.7.72), Arg-RS and arginine was carried out at 37 °C for 5 min in a chilling/heating block with 96-well adapter (Cole-Parmer, Vernon Hills, IL, USA). The substrate peptide for yATE1 was EPGLCTWQSLR with an initial concentration

of 2.24 mM as determined by amino acid analysis. Substrate and product peptides were labelled using (d5)-bromoethane^[11] resulting in a *heavy substrate* (*m/z* 1323) and a *light product* (*m/z* 1474). Prior to adding the yeast enzyme, the temperature of the chilling/heating block was reduced to 30 °C. 10 μ g of yATE1 was added to a 100 μ L reaction volume representing t_0 and individual time points were taken by withdrawing 5 μ L aliquots and quenching with an equal volume of quench solution (10 μ L of 1 mg/mL bovine serum albumin, 10 μ L of acetonitrile (ACN), and 2.5% trifluoroacetic acid (TFA) to a final volume of 100 μ L). All the time points were spotted in duplicate using α -cyano-4-hydroxycinnamic acid (CHCA) as matrix.

Quantification

Matrix-assisted laser desorption/ionization (MALDI) analysis was performed using an UltraFlex III time-of-flight mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) and the obtained spectra were analyzed using Bruker flexAnalysis software (version 4.2). The peak detection algorithm was SNAP (signal-to-noise ratio: 2, quality factor threshold: 30) and the peak area was plotted as a function of time. During the arginylation reaction the *heavy substrate* will be converted into a *heavy product* by ATE1 and the ratio of *light* and *heavy product* peptides serves as a measure of product appearance. Technically, adding the *light product* peptide within the reaction mixture, as opposed to adding the *light product* reference peptide to the quench, will eliminate pipetting irregularities using multiplex pipettors. This is because the ratios of both *light* and *heavy product* peptides are determined and both will be affected equally by pipetting irregularities. The ionization properties of the *light* and *heavy product* peptides are identical and so their peaks can be used for quantification.^[11]

Externally calibrated peptides

Substrate peptides were synthesized using Fmoc solid-phase technology, purified by high-performance liquid chromatography (HPLC) to >97% and quantified using amino acid analysis (purchased in 5 μ M aliquots from Thermo Scientific, Ulm, Germany as HeavyPeptide AQUA QuantPro). All peptides were stable isotope labelled on their C-terminal Arg (¹³C₆, ¹⁵N₄) or Lys (¹³C₆, ¹⁵N₂) with the following peptide sequences in the NH₂-to-COOH direction: DGQCTLVSSLDSTLR, DIHVNGLPAAEEISGK, DLALQLLHK, DPSLLLHK, DLSI-NATNIK, EAQISSAIVSSVQSK, EASGLSADSLAR, ELGSST-NALR, ENQIPEEAGSSGLGK, EPVQLETLSIR, EQLHLYDTR, EVSSATNALR, EVTIVVLGNK.

Liquid chromatography

One-dimensional (1D) chromatographic separations of peptides were performed by a nanoLC ultra 2Dplus system (Eksigent, AB SCIEX Germany GmbH, Darmstadt, Germany) coupled to a 15 cm fused-silica emitter, 75 μ m inner diameter, packed with a Magic C18 AQ 5 μ m resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded on the column from a cooled (4 °C) nanoLC-AS2 autosampler (Eksigent) and separated in a 60-min linear gradient of ACN (5–40%) and water containing 0.2% formic acid at a flow rate of 300 nL/min.

Mass spectrometer settings

Direct infusion was performed using a hybrid ion trap-Orbitrap mass spectrometer with dual-pressure linear ion trap technology (LTQ Orbitrap Velos, Thermo Fisher Scientific Inc., Waltham, MA, USA). MS/MS spectra were acquired in the Orbitrap with a resolution of 60 000 at m/z 400 after accumulation to a target value of 1×10^5 for ETD and ion-trap-type CID fragmentation. The normalized collision energy was set to 35% for ion-trap-type CID fragmentation. The target number of reagent anions to be injected into the trap to perform ETD $= 1 \times 10^5$, the charge-state dependent ETD reaction times was enabled, setting a reference value of 80 ms for doubly charged peptides, and the isolation width for ETD was set to 3 u. The peptide concentration was 1 μ M for direct infusion. Nano-LC/MS was performed with an LC system with the above-mentioned LC settings coupled to a hybrid ion trap-Orbitrap analyzer mass spectrometer (LTQ Orbitrap XL, Thermo Scientific). The mass spectrometer was set to a resolution of 60 000 (m/z 350–1600) and the six most intense ions were selected for CID fragmentation with a normalized collision energy of 35 eV, minimum signal required: 150, and minimum charge state 2.

UPS1 protein standard

The UPS1 standard was purchased from Sigma-Aldrich (St. Louis, MO, USA) and contains 5 pmol per 50 partially truncated human proteins (see Supporting Information). The UPS1 proteins were resuspended in lysis buffer (8 M urea, 0.1 M ammonium bicarbonate, 0.1% RapiGest™ SF surfactant), reduced with 5 mM final concentration of TCEP (tris-2-carboxyethyl phosphine) at 37 °C for 10 min, and alkylated with 10 mM final concentration iodoacetamide at 25 °C for 30 min. Prior to enzymatic digestion, the buffer was adjusted to 2 M urea using 0.1 M ammonium bicarbonate. Each vial was digested at pH 8 and 37 °C with Asp-N (endoproteinase Asp-N 2 μ g, sequencing grade, 11054589001; Roche Diagnostics, Rotkreuz, Switzerland) for 3 h. For Lys-C/Asp-N digest, a vial of UPS1 was first digested with Lys-C (endoproteinase Lys-C 5 μ g, 11047825001; Roche Diagnostics) for 30 min at 37 °C, followed by the addition of Asp-N for 3 h at 37 °C. Following digestion, the reaction mixture was supplemented with buffer and enzymes as described above (*yATE1 enzyme kinetics*). Then, 90 min after adding *yATE1*, the reaction was quenched using 10% TFA to obtain a pH of 2 to 3 (as judged by pH paper). The peptides were purified by a standard C18 clean-up procedure^[12] using a vacuum manifold (Sep-Pak Vac 1 cc (50 mg) tC18 cartridges, WAT054960; Waters Ltd, Elstree, UK).

RESULTS AND DISCUSSION

The mobile proton hypothesis states that protonation of peptides during gas-phase ionization is mainly governed by the basic sites of the analyte. Hence, a typical tryptic peptide harboring a single basic amino acid is doubly charged (with the second proton secured on the primary amine of the N-terminus). We directly infused chemically synthesized peptide EPGLCTWQSLR into a hybrid ion trap-Orbitrap mass spectrometer using ESI-MS and observed a doubly

charged precursor, as seen in Fig. 1(A). In accordance with the mobile proton hypothesis, adding an additional Arg to the N-terminus of a tryptic peptide can result in doubly and triply charged precursors. Formation of the $[M+2H]^{2+}$ ion is rationalized by sequestering protons on both basic amino acids, while the $[M+3H]^{3+}$ ion is rationalized by further securing a proton via the primary amine of the N-terminus. Indeed, directly infusing REPGLCTWQSLR into a hybrid ion trap-Orbitrap mass spectrometer results in $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions, as seen in Fig. 1(B). Furthermore, the triply charged precursor ion could be selected for subsequent ETD. Despite there being a Pro residue in the peptide sequence, a complete c-fragment ion ladder was observed in the resulting MS/MS spectrum together with a complete y-fragment ion ladder and several other product ions (see Fig. 1(C)). Hence, arginylation allows for ETD fragmentation of relatively short peptides and the resulting MS/MS spectra contain at least 2 \times sequence coverage for unambiguous assignment of peptides. Based on these results we decided to further pursue the application of arginylation for proteomics and to determine the robustness of the *yATE1* enzyme.

Enzyme kinetics of *yATE1*

Arginylation is an enzymatic reaction that involves two substrates: aminoacylated tRNA^{Arg} and (poly-)peptides with N-terminal acidic amino acids (see Fig. 2(A)).^[10] To determine the Michaelis-Menten rate constants of *Saccharomyces cerevisiae* arginyl-tRNA protein transferase 1 (*yATE1*), we used a multiplexed quantitative method based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS) and ethyl-(D5)-labelled peptide standards (short *pentadeutero method* for the remainder of the manuscript).^[11,13,14] As MALDI-MS is inherently non-quantitative, i.e. the signal observed from an analyte is poorly related to the amount of analyte in the sample, quantification of the *pentadeutero method* is based on differential labelling of substrate and product peptides containing an internal Cys.^[13,14] The substrate peptide EPGLCTWQSLR was labelled using bromoethane-D5, resulting in a deuterated S-ethylcysteine (*heavy substrate*, m/z 1323). The product peptide with the sequence REPGLCTWQSLR was labelled using isotopically light bromoethane and served as a reference standard (*light product*, m/z 1474). Both the *heavy substrate* and the *light product* were added to the arginylation reaction, and aliquots were taken as a function of time, quenched with TFA and directly spotted onto a t-plate for mass spectral analysis. Due to the low complexity peptide mixture with known substrate and product peptide m/z values, no chromatographic separation was required prior to MALDI spotting. Figure 2(B) shows an example of the primary MALDI data: the *light product* (m/z 1474) is kept constant at 1, while the *heavy product* (m/z 1479) increases as a function of time.

To determine Michaelis-Menten rate constants for *yATE1*, the arginylation reaction was carried out in triplicate with an *in situ* aminoacylation reaction to regenerate aminoacylated Arg-tRNA^{Arg} (see Experimental section for details). The peptide product appearance measurements^[15] were carried out over a range of peptide substrate concentrations ranging from 4.48 μ M to 51.5 μ M, while the *yATE1* and tRNA^{Arg}

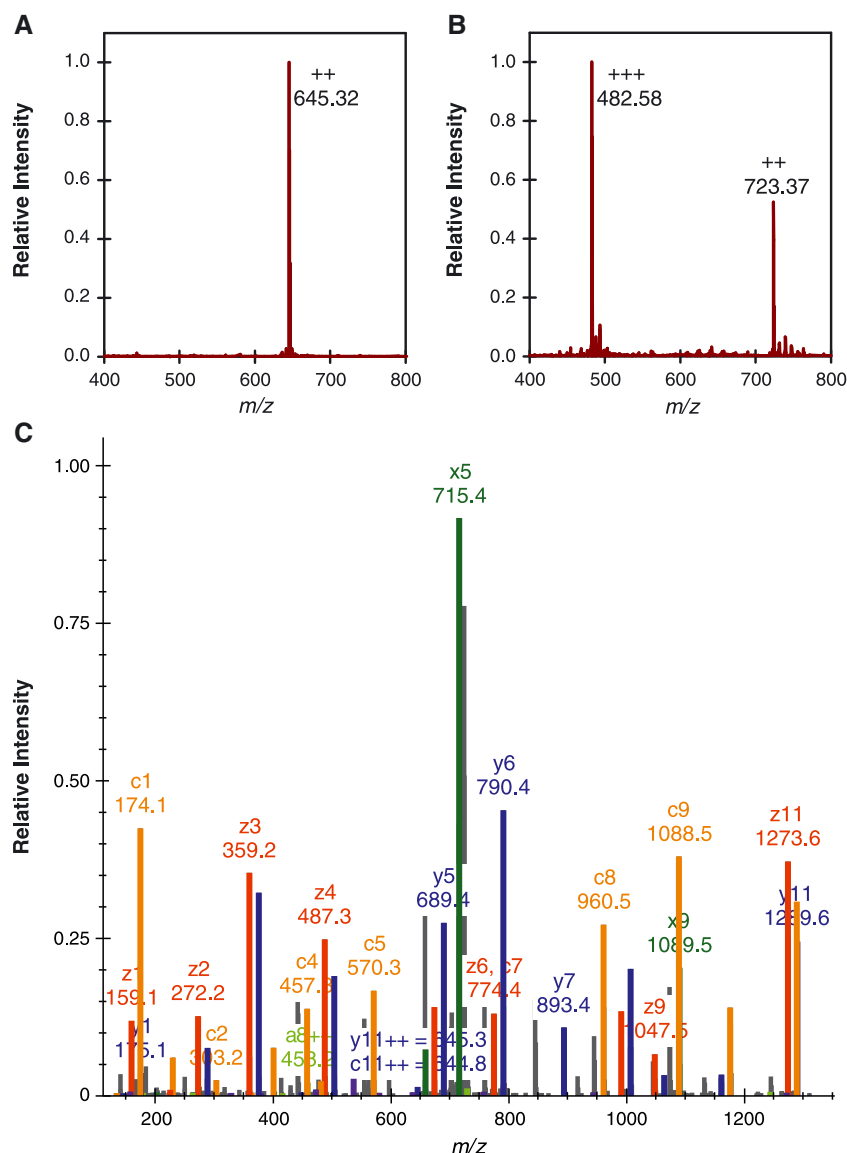


Figure 1. Direct infusion ESI-MS of peptide (R)EPGLCTWQSR. (A) MS1 scan of direct infusion of EPGLCTWQSR resulting in a doubly charged precursor ion $[M+2H]^2+$. (B) MS1 scan of direct infusion of REPLCTWQSR resulting in doubly and triply charged precursor ions. (C) ETD fragmentation of REPLCTWQSR $[M+3H]^3+$ results in complete c- and y-ion series as well as several other peptide fragments.

concentrations were kept constant. The product appearance, plotted as a function of initial *heavy substrate* concentration, and the initial rate constants were determined. The end point of the initial linear velocity was determined to be 6 min. Prism (GraphPad Software, La Jolla, CA, USA) was used to plot all kinetic curves and calculate kinetic parameters. The v_{\max} was calculated to be $2.2 \mu\text{M} \cdot \text{min}^{-1}$, $K_m = 39.3 \mu\text{M}$ and $k_{\text{cat}} = 16.1 \text{ min}^{-1}$ in our initial yATE1 purification (see Table 1 and Fig. 2(C)). In this experiment 220 pmol of EPGLCTWQSLR (MW 1289.49) in are converted per minute or 2 μg per 7 min. As a typical amount of peptide mixture analyzed per LC/MS experiment is 2–5 μg , we conclude that these enzymatic parameters are suitable to further pursue the application of arginylation at a scale that is compatible with common proteomic workflows.

Extending arginylation to 13 chemically synthesized and calibrated peptides

The *pentadeutero method* is based on MS1 m/z value measurements using MALDI-MS and requires internal Cys for labelling. For more complex peptide mixtures, MS1 m/z values are not sufficient to unequivocally identify peptide substrates and products. Furthermore, not all peptides contain an internal Cys. Hence, to apply arginylation beyond a single peptide substrate, we chose 13 chemically synthesized and amino acid analysis calibrated peptides with N-terminal aspartate or glutamate. These 13 substrate peptides were arginylated in varying concentrations ranging from 5 to 405 pmol per enzyme reaction and a single 90 min end point was analyzed by ESI-LC/MS (see Experimental section, and

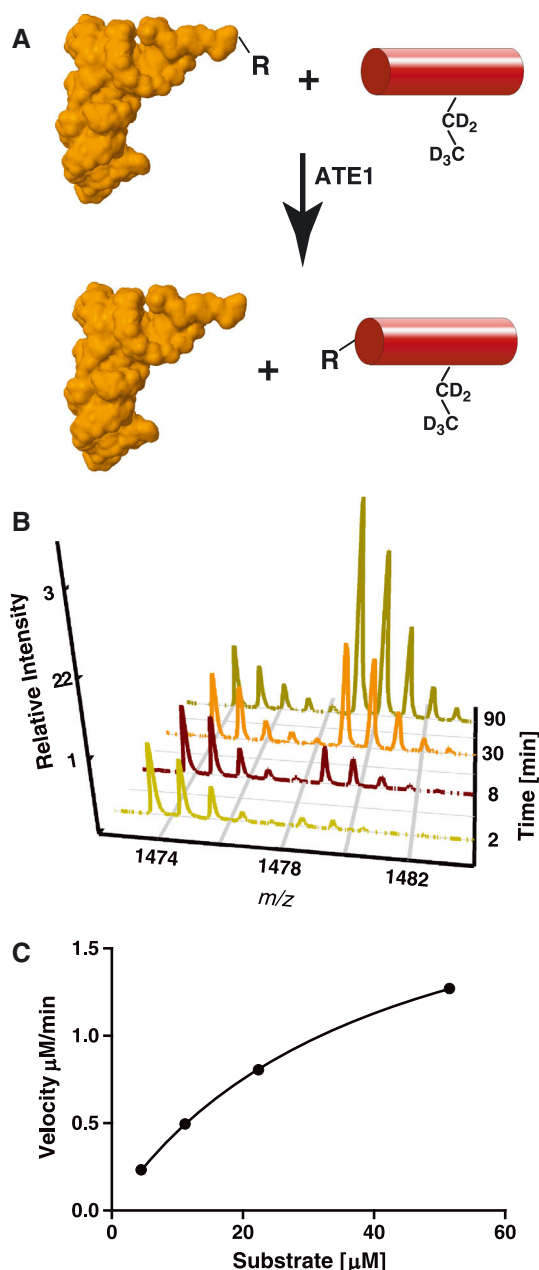


Figure 2. Enzyme reaction mechanism and quantitation using quantitative MALDI assay. (A) Arginylation reaction mechanism: ATE1 conjugates an arginyl moiety from an aminoacylated tRNA^{Arg} onto a peptide (depicted as red cylinder) containing an ethylated Cys for quantification. The tRNA structure was based on work from Konno and colleagues (ZZUF).^[27] (B) The *light* reference product with a monoisotopic m/z 1474 ion was added to the reaction mixture as well as the *heavy* substrate of m/z 1323 in excess (not shown). As a function of time, the *heavy* product with a monoisotopic ion of m/z 1479 appears and is quantified relative to the spiked-in reference peptide (m/z 1474). (C) Initial enzyme velocity as a function of initial substrate concentration. In black circles are data points for yATE1 enzyme kinetics.

Supporting Information, 13-substrate-peptides^[28]). Due to the conjugation of the arginyl moiety onto the N-terminus of the substrate peptide, there was a clear retention time shift between the substrate peptide and the arginylated peptide, on

Table 1. Initial rates and kinetic parameters of yeast arginyl-tRNA protein transferase 1 (yATE1)-catalyzed peptide bond formation

Substrate peptide concentration (μM)	Initial rate ($\mu\text{M min}^{-1}$)	Rate constants	
yATE1			
4.48	0.243 \pm 0.037	V_{max}	2.238
11.2	0.496 \pm 0.170	K_{m}	39.34
22.4	0.805 \pm 0.210	k_{cat}	16.08
51.5	1.271 \pm 0.455	$k_{\text{cat}}/K_{\text{m}}$	0.408

average 10% earlier in the linear gradient on a reversed-phase C18 column. Overall, the arginylation reaction of 13 substrate peptides was near quantitative at 5 and 45 pmol substrate peptide. Once the arginylation reaction was challenged with 405 pmol of peptide per reaction, the arginylation efficiency after 90 min was no longer quantitative (see Figs. 3(A) and 3(B)). For example, peptide DAGVVCTDETR elutes at 13.3 min and its arginylated version elutes at 11.7 min. Overall, the peptide has a very high degree of arginylation (Fig. 3(A)). This is in contrast to the substrate peptide EAQISSAIVSSVQSK, reacted under the same conditions, where a large fraction of substrate peptide (eluting at 23.3 min) was still unreacted (Fig. 3(B)). To further investigate this substrate bias using a more complex sample, we performed the arginylation reaction using a proteolytic digest of the Universal Protein Standard 1 (UPS1) comprised of 50 proteins in equimolar ratio (5 pmol each per vial).

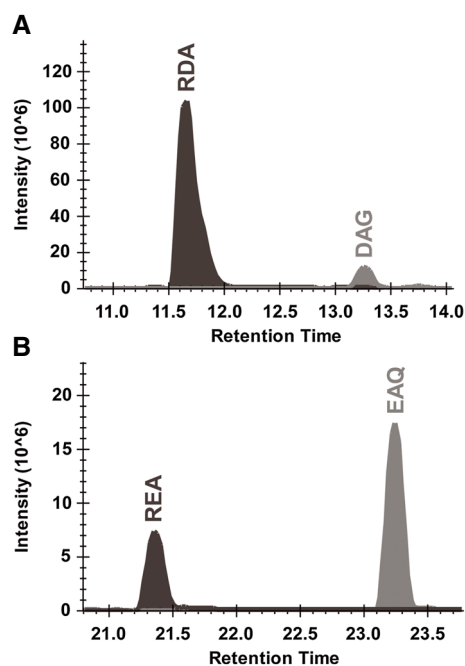


Figure 3. yATE1 arginylates substrate peptides over a wide range of substrate concentrations (from 5 to 405 pmol of substrate peptides). At the highest substrate concentration, product formation takes place, but is no longer quantitative. Ion chromatogram of MS1 trace of arginylated (dark grey) and non-arginylated (light grey) peptides as a function of time for (A) DAGVVCTDETR and (B) EAQISSAIVSSVQSK (both at 405 pmol substrate).

Protease digestion optimization and labelling bias investigation

To obtain suitable substrate peptides for arginylation, the UPS1 proteins were digested using Asp-N, resulting in N-terminal aspartate or glutamate, but unspecific C-termini. The proteolytic peptides were arginylated for 90 min and subsequently desalted using a standard C18 clean-up procedure. Using LC/MS, 149 peptides with N-terminal aspartate or glutamate were detected as being arginylated. In some cases, where no arginylation of peptides with acidic amino acids was detected, manual inspection of the MS1 ion chromatograms using Skyline^[16] revealed many pairs of substrate and product peptide precursor ion masses corresponding to the theoretical values (see Supporting Information, UPS1_AspN_digest^[28]). In some cases, only substrate peptides eluted very early in the LC gradient. As arginylation will result in a chromatographic shift (see Fig. 3), we assume that some arginylated peptides might be too polar to bind to the C18 column material used. Hence, we conclude that all the substrate peptides were arginylated and that the stochastic sampling approach of shotgun MS/MS or chromatographic performance is probably the reason why MS/MS spectra for the arginylated version of some substrate peptides are lacking. Examining the Asp-N digest of 50 UPS1 peptides reveals that only 35 proteins were detected.

To increase the number of detectable proteins, the single Asp-N protease digest was further optimized by using a double digest, e.g. Lys-C. A double digest of Lys-C and Asp-N results in substrate peptides for arginylation and peptides with C-terminal Lys. In addition, these peptides would allow for a direct comparison between MS/MS fragmentation of non-arginylated and arginylated peptides. Hence, the UPS1 protein standard was first digested using Lys-C followed by an Asp-N digest. The resulting peptide mixture was arginylated and peptides analyzed by LC/MS. Subsequent *in silico* annotation of the MS/MS spectra resulted in experimental evidence for 45 proteins, an increase of 20% over the mono Asp-N digest. On the peptide level, 156 peptides with an N-terminal acidic amino acid were detected. The majority of these peptides were quantitatively arginylated or both arginylated and non-arginylated forms were detected (see Supporting Information, UPS1_LysC_AspN_digest^[28]).

To further illustrate the influence of arginylation on peptide fragmentation, we selected peptides from the UPS1 Lys-C/Asp-N double digest. As can be seen in Fig. 4(A), peptide DDHFLF precursor $[M+2H]^2+$ ion fragments into six annotated product ions. However, upon arginylation of the peptide, the $[M+2H]^2+$ precursor ion fragments and the number of annotated ions doubles (Fig. 4(B)) although only a single amino acid was added to the N-terminus of the peptide. Figure 4(C) shows the MS/MS spectrum of $[DDTVCLAK+2H]^2+$ with enhanced cleavage C-terminal to two Asp residues. This enhanced cleavage is overcome upon arginylation, resulting in a more balanced fragmentation of the $[RDDTVCLAK+2H]^2+$ ion. The resulting MS/MS spectrum is comprised of b- and y-fragment ion ladders (Fig. 4(D)). A third example illustrating the contribution of arginylation to precursor ion fragmentation is peptide EGVVGAWEK. The $[EGVVGAWEK+2H]^2+$ precursor ion fragmentation results in very few annotated product ions (Fig. 4(E)) while, upon arginylation, complete b- and y-fragment

ion ladders were annotated (Fig. 4(F)). As all the MS/MS spectra are charge state matched, it is possible to study the influence of arginylation on peptide fragmentation. Non-arginylated peptides (Figs. 4(A), 4(C) and 4(E)) secure two protons, one to the primary amine of the N-terminus and the second to the side chain of a basic amino acid within the peptide sequence. This is in contrast to arginylated peptides (Figs. 4(B), 4(D) and 4(F)) securing two protons, where both reside with the side chain of basic amino acids. The side chain sequestered protons allow for a more stochastic fragmentation of the peptide backbone to generate fragment ion ladders. Upon introduction of an additional proton, presumably on the primary amine of the peptide's N-terminus, the product ions are again distributed unevenly. Peptide EPISVSSEQVLK was detected three times in the UPS1 double digest dataset: non-arginylated $[M+2H]^2+$ (Fig. 5(A)), arginylated $[M+2H]^2+$ (Fig. 5(B)) and arginylated $[M+3H]^3+$ (Fig. 5(C)). The most dominant product ion in $[EPISVSSEQVLK+2H]^2+$ (Fig. 5(A)) is $y7^+$ generated by peptide backbone cleavage between Val₅ and Ser₆. Upon arginylation, in the formation of the $[REPISVSSEQVLK+2H]^2+$ ion (Fig. 5(B)) each amino acid side chain of Arg and Lys sequesters a proton resulting in an evenly distributed fragmentation pattern of b-fragment and y-fragment ions during the ion trap CID. However, upon adding an additional proton to the primary amine of the peptide's N-terminus, the $[REPISVSSEQVLK+3H]^3+$ ion (Fig. 5(C)) fragments again at preferred sites, e.g. between Glu and Gln. From these results (and others found in the Supporting Information, UPS1_LysC_AspN_digest^[28]), we conclude that arginylation enhances peptide fragmentation and allows for the detection of relatively short peptides.

No mid-chain arginylation detectable on the peptide level

There are reports not only that N-terminal acidic amino acids are substrates of ATE1, but also that mid-chain Asp or Glu could be arginylated through their side-chain carboxyl groups.^[17,18] To test for mid-chain arginylation of peptides, we re-analyzed the 90 min end-point reactions of the UPS1 experiments using the Mascot error tolerant search option (Matrix Science, London, UK), which is especially useful for detecting unexpected chemical or post-translational modifications. As expected, an additional mass of +156.10 (equivalent to an Arg moiety) was only detected at N-terminal residues, but not mid-chain residues (see Supporting Information, error-tolerant-search.csv^[28]). We cannot rule out that on the protein level ATE1 might arginylate mid-chain acidic amino acids, but in our peptide level experiments $yATE1$ only arginylated N-terminal residues.

CONCLUDING REMARKS

We have developed a method using $yATE1$ which will generate peptides flanked on both termini with basic amino acids under physiological conditions. We determined the Michaelis-Menton kinetic parameters of $yATE1$ using an *in situ* aminoacylation tRNA^{Arg} enzyme reaction system. We conclude that a suitable amount of substrate peptides encountered for a typical proteomics experiment is arginylated within less than 10 min. Furthermore, $yATE1$

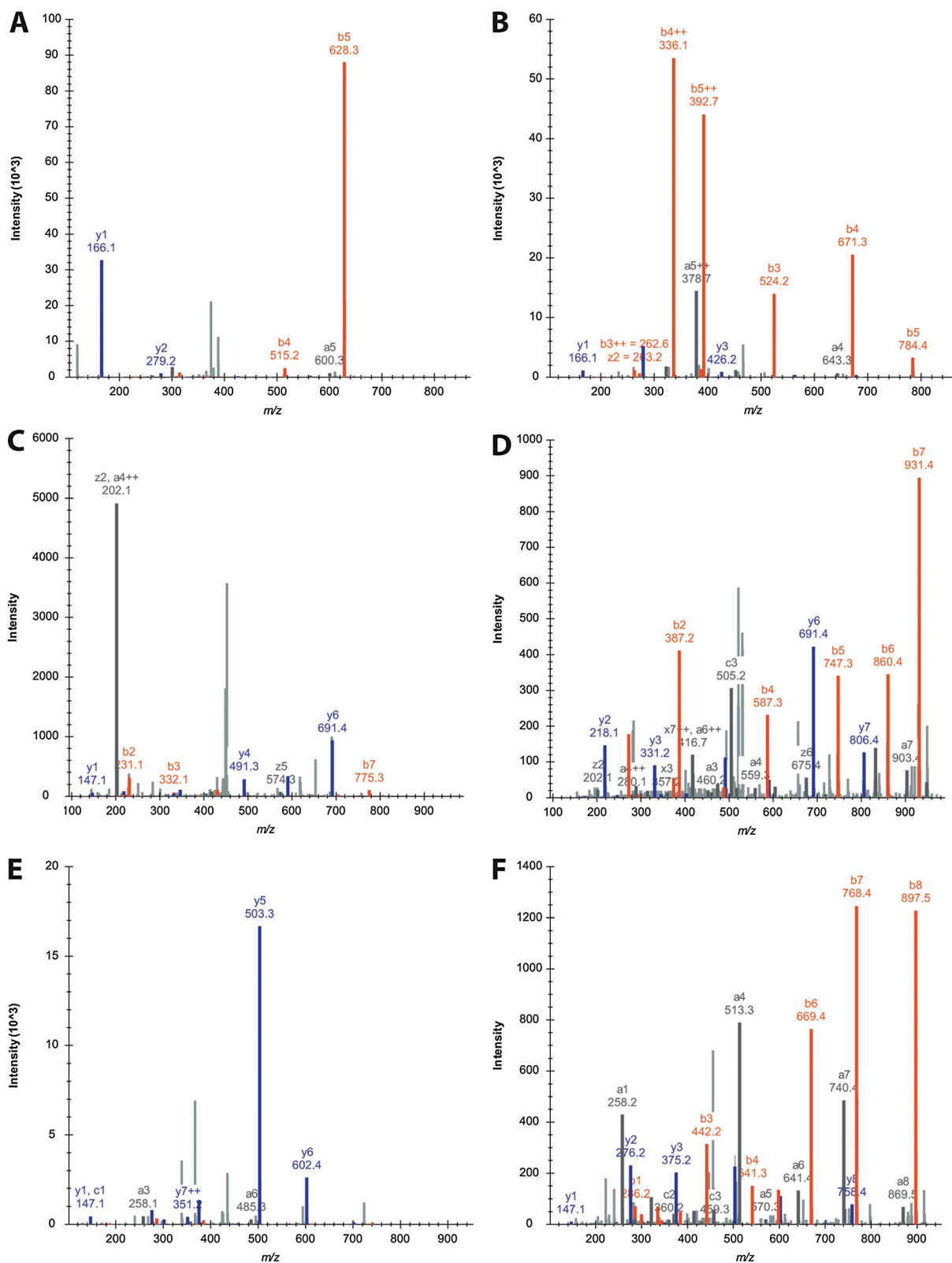


Figure 4. Contribution of arginylation to enrich MS/MS product ion spectra using CID fragmentation. b-ions are coloured in orange while y-ions are coloured in blue. (A) Peptide DDHFLF [M+2H]²⁺ fragments and very few product ions are detectable. Upon arginylation of DDHFLF, the number of detectable product ions doubles as shown in the MS/MS spectrum of [M+2H]²⁺ in (B). (C) Peptide DDTVCLAK [M+2H]²⁺ generates very few b- and y-ions, but, upon arginylation, the doubly charged precursor fragments and two sequence ladders comprised of b- and y-ions are detectable (as seen in (D)). The amount of product ions of the doubly charged precursor ion triples upon arginylation of peptide EGVVGAIVEK as seen in (E) non-arginylated versus (F) arginylated.

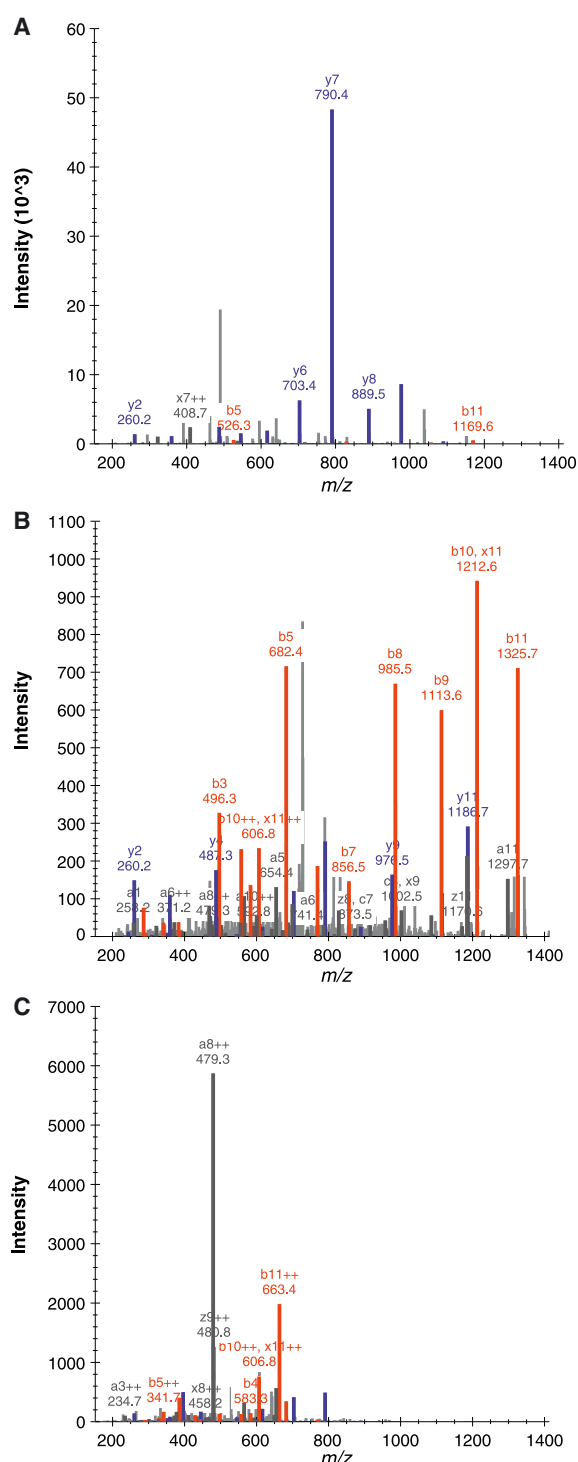


Figure 5. Contribution of the proton secured on the N-terminal primary amine. The peptide EPISVSSEQVLK was detected as $[M+2H]^{2+}$ and the resulting MS/MS spectrum is shown in (A). Upon arginylation, $[REPIVSSEQVLK+2H]^{2+}$ generates 2 \times fragment ladders as both protons are secured on the basic amino acid side chains in (B). Upon securing an additional proton to the N-terminus of the peptide, $[REPIVSSEQVLK+3H]^{3+}$ generates MS/MS spectra with preferred cleavage sites as seen in (C). (Note: the MS/MS spectra were imported into Skyline and arginylation was set as a variable post-translational modification of Asp or Glu. Hence, the numbering nomenclature remains identical to the unmodified peptide sequence, e.g. $a8^{2+} = m/z$ 479.3.)

only arginylates peptides on their N-terminal acidic amino acid residues and no mid-chain reaction activity was detected in our experiments carried out on the peptide level. We optimized the proteolytic parameters to digest proteins and conclude that a double digest using Lys-C/Asp-N will generate peptides with N-termini as substrates for arginylation while generating sequence-specific C-termini with a basic amino acid. The direct comparison of charge-state-matched MS/MS spectra of non-arginylated versus arginylated peptides clearly shows the positive effect of the N-terminal Arg residue on peptide fragmentation to generate redundant sequence information within the same MS/MS spectrum. Currently, a limitation is the substrate specificity of yATE1 for N-terminal acidic amino acids. However, there are prokaryotic^[13,14] and viral^[19] peptidyl transferases exhibiting various preferences for N-terminal amino acids recognized and generation of neo-termini, so that one could entertain the idea of using these in protein engineering or *in vitro* selection^[20] towards arginyl transferases with a wider N-terminal amino acid recognition repertoire.

To our knowledge, this is the first report of consistent 2 \times peptide sequence coverage within a single MS/MS spectrum. This high coverage is especially useful for *de novo* peptide sequence determination from organisms without precisely annotated proteomes,^[21] environmental samples containing unknown organisms,^[22] or peptides originating from non-sequence-specific protease cleavage such as major histocompatibility complex (MHC) peptidome^[23] surveys, especially of HLA-A26 subtype A allele which accommodate acidic amino acids on their N-termini.^[24] The boost in available proton(s) will benefit amide backbone cleavage of peptides and contribute to the detection of product ions. In addition, arginylation leads to an increase in precursor charge which in turn leads to the selection of the precursor for MS/MS fragmentation.^[25] Furthermore, we envision arginylation being applied as an alternative tool for the 100% sequence coverage problem, in which the *m/z* values of all amino acids of a given protein are determined to elucidate precise protein sequence maps including sites of post-translational/post-transcriptional modifications.^[26]

Supplementary data^[28]

REPGLCTWQSLR_ETD_3p_2014-09-17_09-04-50.sky.zip: contains MS/MS spectrum as annotated by Mascot v.2.4.1 of ETD fragmented REPGLCTWQSR $[M+3H]^{3+}$.

13-substrate-peptides_2014-09-19_12-48-53.sky.zip: contains MS1 and MS/MS spectra of 13 externally calibrated peptides which were arginylated.

UPS1_AspN_digest_2014-09-19_13-21-30.sky.zip: contains MS1 extracted ion chromatograms together with MS/MS spectra of peptides generated by digest of protein standard UPS1 using Asp-N endopeptidase.

UPS1_LysC_AspN_digest_2014-09-21_12-11-45.sky.zip: contains MS1 extracted ion chromatograms together with MS/MS spectra of peptides generated by double digest of protein standard UPS1 using Lys-C and Asp-N endopeptidases.

error-tolerant-search.csv: mascot error tolerant search of Lys-C/Asp-N digested UPS1.

ups1.fasta: truncated protein sequences of UPS1.

Acknowledgements

This work was supported by a research grant from the Canadian Breast Cancer Foundation (CBCF) to RPF and a Marie Curie International Incoming Fellowship to HAE. The direct infusion experiments were performed at the Functional Genomics Center Zurich, Switzerland.

REFERENCES

- [1] V. H. Wysocki, G. Tsaprailis, L. L. Smith, L. A. Breci. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399.
- [2] B. Domon, R. Aebersold. Challenges and opportunities in proteomics data analysis. *Mol. Cell. Proteomics* **2006**, *5*, 1921.
- [3] V. Dancik, T. A. Addona, K. R. Clauser, J. E. Vath, P. A. Pevzner. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999**, *6*, 327.
- [4] A. M. Frank, M. M. Savitski, M. L. Nielsen, R. A. Zubarev, P. A. Pevzner. De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* **2007**, *6*, 114.
- [5] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlett-Jones, F. He, A. Jacobson, D. J. Pappin. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3*, 1154.
- [6] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. Mohammed, C. Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895.
- [7] E. Balzi, M. Choder, W. N. Chen, A. Varshavsky, A. Goffeau. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **1990**, *265*, 7464.
- [8] R. G. Hu, C. S. Brower, H. Wang, I. V. Davydov, J. Sheng, J. Zhou, Y. T. Kwon, A. Varshavsky. Arginyltransferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. *J. Biol. Chem.* **2006**, *281*, 32559.
- [9] A. Varshavsky. The N-end rule pathway of protein degradation. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* **1997**, *2*, 13.
- [10] A. Varshavsky. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **2011**, *20*, 1298.
- [11] H. A. Eberhardt, Z. Xu, A. W. Fung, R. P. Fahlman. Quantification of the post-translational addition of amino acids to proteins by MALDI-TOF mass spectrometry. *Anal. Chem.* **2009**, *81*, 1937.
- [12] H. A. Eberhardt, E. Sabido, R. Huttenhain, B. Collins, R. Aebersold. Range of protein detection by selected/multiple reaction monitoring mass spectrometry in an unfractionated human cell culture lysate. *Proteomics* **2012**, *12*, 1185.
- [13] A. W. Fung, H. A. Eberhardt, K. S. Krishnakumar, J. Moore, Z. Xu, P. Strazewski, R. P. Fahlman. Probing the leucyl/phenylalanyl tRNA protein transferase active site with tRNA substrate analogues. *Protein Peptide Lett.* **2014**, *21*, 603.
- [14] A. W. Fung, H. A. Eberhardt, H. Abeyundara, J. Moore, Z. Xu, R. P. Fahlman. An alternative mechanism for the catalysis of peptide bond formation by L/F transferase: substrate binding and orientation. *J. Mol. Biol.* **2011**, *409*, 617.
- [15] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Butterworths, London, Boston, **1979**.
- [16] B. Schilling, M. J. Rardin, B. X. MacLean, A. M. Zawadzka, B. E. Frewen, M. P. Cusack, D. J. Sorensen, M. S. Bereman, E. Jing, C. C. Wu, E. Verdin, C. R. Kahn, M. J. Maccoss, B. W. Gibson. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol. Cell. Proteomics* **2012**, *11*, 202.
- [17] S. Saha, A. Kashina. Posttranslational arginylation as a global biological regulator. *Dev. Biol.* **2011**, *358*, 1.
- [18] J. Wang, X. Han, C. C. Wong, H. Cheng, A. Aslanian, T. Xu, P. Leavis, H. Roder, L. Hedstrom, J. R. Yates 3rd, A. Kashina. Arginyltransferase ATE1 catalyzes midchain arginylation of proteins at side chain carboxylates in vivo. *Chem. Biol.* **2014**, *21*, 331.
- [19] E. Graciet, R. G. Hu, K. Piatkov, J. H. Rhee, E. M. Schwarz, A. Varshavsky. Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3078.
- [20] C. Tuerk, S. MacDougal, L. Gold. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6988.
- [21] S. E. St Pierre, L. Ponting, R. Stefancsik, P. McQuilton, C. FlyBase. FlyBase 102 – advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* **2014**, *42*, D780.
- [22] S. Yooseph, G. Sutton, D. B. Rusch, A. L. Halpern, S. J. Williamson, K. Remington, J. A. Eisen, K. B. Heidelberg, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C. S. Miller, H. Li, S. T. Mashiyama, M. P. Joachimiak, C. van Belle, J. M. Chandonia, D. A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B. J. Raphael, V. Bafna, R. Friedman, S. E. Brenner, A. Godzik, D. Eisenberg, J. E. Dixon, S. S. Taylor, R. L. Strausberg, M. Frazier, J. C. Venter. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **2007**, *5*, e16.
- [23] S. Walter, T. Weinschenk, A. Stenzl, R. Zdrojowy, A. Pluzanska, C. Szczylik, M. Staehler, W. Brugger, P. Y. Dietrich, R. Mendrzyk, N. Hilf, O. Schoor, J. Fritsche, A. Mahr, D. Maurer, V. Vass, C. Trautwein, P. Lewandrowski, C. Flohr, H. Pohla, J. J. Stanczak, V. Bronte, S. Mandruzzato, T. Biedermann, G. Pawelec, E. Derhovanessian, H. Yamagishi, T. Miki, F. Hongo, N. Takaha, K. Hirakawa, H. Tanaka, S. Stevanovic, J. Frisch, A. Mayer-Mokler, A. Kirner, H. G. Rammensee, C. Reinhardt, H. Singh-Jasuja. Multipetide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival. *Nat. Med.* **2012**, *18*, 1254.
- [24] T. Dumrese, S. Stevanovic, F. H. Seeger, N. Yamada, Y. Ishikawa, K. Tokunaga, M. Takiguchi, H. Rammensee. HLA-A26 subtype A pockets accommodate acidic N-termini of ligands. *Immunogenetics* **1998**, *48*, 350.
- [25] A. Michalski, J. Cox, M. Mann. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 1785.
- [26] B. Meyer, D. G. Papatotiriou, M. Karas. 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2011**, *41*, 291.
- [27] M. Konno, T. Sumida, E. Uchikawa, Y. Mori, T. Yanagisawa, S. Sekine, S. Yokoyama. Modeling of tRNA-assisted mechanism of Arg activation based on a structure of Arg-tRNA synthetase, tRNA, and an ATP analog (ANP). *FEBS J.* **2009**, *276*, 4763.
- [28] Available: <https://daily.panoramaweb.org/labkey/arg-ms.url>.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website.