Open access • Journal Article • DOI:10.1038/NMETH.3177

LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification — Source link

Pitter F. Huesgen, Philipp F. Lange, Lindsay D. Rogers, Nestor Solis ...+5 more authors

Institutions: University of British Columbia, Monash University, Clayton campus

Published on: 01 Jan 2015 - Nature Methods (Nature Research)

Topics: Trypsin, Methanosarcina acetivorans, Arginine, Lysine and Methylation

LysargiNase mirrors trypsin for protein C-termini and methylation site identification

Pitter F. Huesgen[1-4], Philipp F. Lange[1-3], Lindsay D. Rogers[1-3], Nestor Solis[1-3], Ulrich Eckhard[1-3], Oded Kleifeld[5], Theodoros Goulas[6], F. Xavier Gomis-Rüth[6] & Christopher M. Overall[1-3]

[1]Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada
[2]Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, BC, Canada
[3]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada
[4]Zentralinstitut für Engineering, Elektronik und Analytik, ZEA-3, Forschungszentrum Jülich, Jülich, Germany
[5]Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia
[6]Department of Structural Biology, Molecular Biology Institute of Barcelona, Spanish Research Council, Barcelona, Spain

Correspondence should be addressed to C.M.O. (chris.overall@ubc.ca)

Editorial summary:
The *Archaea* metalloproteinase LysargiNase increases proteome coverage, identifies more C-terminal peptides from proteins and improves methylated peptide identification.

To improve proteome coverage and protein C-termini identification we characterized the *Methanosarcina acetivorans* thermophilic proteinase LysargiNase, which cleaves before lysine and arginine up to 55°C. Unlike trypsin, LysargiNase-generated peptides have N-terminal lysine or arginine residues and fragment with b-ion dominated spectra. This improves protein C-terminal peptide identification and several arginine-rich phosphosite assignments. Uniquely, cleavage also occurs at methylated or dimethylated lysine and arginine, facilitating detection of these epigenetic modifications.

Proteomic identification of proteins and posttranslational modifications requires proteolytic digestion of proteins into peptides for fragmentation by tandem mass spectrometry and identification through spectra-to-sequence matching. The cleavage specificity of the digestion enzyme determines peptide length, *mass to charge ratio,* and sequence and charge position. Trypsin is the most commonly used digestion enzyme in proteomics[1]. Cleaving after arginine and lysine, trypsin generates peptides having a C-terminal positively charged residue in addition to an N-terminal positive charge from the α-amine. This charge distribution results in sequence-informative collision-induced dissociation (CID) fragmentation spectra dominated by y-type ions[1,2]. However, not all peptides are identified following trypsin digestion: Protein C-terminal peptides inherently lack basic residues and are therefore rarely identified without selective enrichment[3,4] leading to loss of biologically important information on protein structure and function[5]. In addition, cleavage after basic residues in the most common phosphosites, *e.g.* (R or K)^n(X)(S or T), can result in ambiguous phosphosite localization. Tryptic cleavage is also blocked at modified (*e.g.* methylated) basic residues and such missed cleavages increase peptide length and charge resulting in complex fragmentation spectra that hinders identification of the peptide and methylation sites[2]. Lysine and arginine methylation is an important epigenetic and regulatory mechanism for protein-protein interactions and protein function in transcription, DNA repair, and signaling[6]. Hence, new approaches are needed to overcome the problems in detecting C-terminal peptides, protein methylation and certain phosphosites.

New proteases to complement trypsin are much sought after for identification of alternate splice forms and the 3,844 “missing” proteins that lack proteomic evidence for their occurrence in human tissues[7]. One such protease is LysN, which selectively hydrolyses peptidyl-lysine bonds and so leaves a lysine at the peptide N-terminus. This results in strong b-ion series during CID[2] and c-ion series during electron transfer dissociation (ETD) fragmentation, enabling *de novo* ladder sequencing of peptides[8]. LysN also identifies peptides missed in tryptic digests[9],

allows for enrichment of posttranslational modified peptides based on peptide charge[10], and cuts before methylated lysines[11]. However, these benefits are restricted to those LysN-generated peptides not containing internal arginines or peptides without lysine miscleavage, which is reported to be 13% (ref. 12). Moreover, only 71% of all LysN cleavage sites contain an N-terminal lysine pointing to poor specificity compared with trypsin[12]. Thus, we sought a protease to address these needs and report the characterization of LysargiNase, a metalloproteinase from the *Archaea* species *Methanosarcina acetivorans*. We changed the original name Ulilysin[13] to reflect its exquisite N-terminal selectivity for lysine in addition to its previously described peptidyl-arginine specificity[13,14]. LysargiNase mirrors trypsin specificity and overcomes the limitations of LysN in controlled N-terminal placement of basic residues, with the added advantage of also cleaving N-terminal to methylated forms of arginine and lysine so generating shorter, less complex peptides.

We compared LysargiNase with trypsin digests of human breast cancer MDA-MB-231 cell proteome using high-resolution MS/MS. We identified cleaved peptides in a stringent and unbiased manner using non-specific database searches, i.e. spectra-to-sequence matching was not restricted by prior information on preferred cleavage sites. 1,096 unique peptides from LysargiNase-digests and 1,789 unique tryptic peptides were identified (n=4) (**Supplementary Tables 1**). By extraction of the sequence surrounding the peptide termini we derived 1,917 and 3,235 unique LysargiNase and trypsin cleavage sites, respectively. LysargiNase showed exquisite sequence specificity: 93% of LysargiNase cleavage sites were before arginine (40%), as previously reported[13,14], but importantly also before lysine (52%) (**Fig. 1a**). In contrast, 91% of tryptic cleavages were after arginine (39%) or lysine (52%). Thus, LysargiNase mirrors trypsin cleavage specificity. The number of peptide identifications was improved when we applied the new specificity information to restrict the search space during spectrum-to-sequence assignment with 2,956 LysargiNase peptides (**Supplementary Table 2**) and 4,203 tryptic peptides identified (**Supplementary Table 3**). Similar to trypsin[15], basic residues in missed LysargiNase cleavage sites were frequently preceded or followed by proline (**Supplementary Fig. 1**). LysargiNase fidelity also compared favorably to the lysine content (71%) at cleavage sites reported for LysN[12]. However, as LysN peptides frequently contain internal arginine or miscleaved lysines experimentally observed LysN peptides are longer on average, 18.0 residues in length[12], versus 14.2 residues for LysargiNase.

Confirming the applicability of LysargiNase for proteomics we found that LysargiNase was

active in a variety of conditions and solvents commonly used in sample preparation and at dilutions down to 1:100 enzyme:proteome ratios (**Supplementary Fig. 2**). We used 1:50 since it represents a similar molar ratio as 1:100 trypsin, enabling a fair comparison. LysargiNase showed notable thermophilicity at 55°C retaining 100% activity compared to 25°C, which aids digestion by maintaining denaturation (**Supplementary Fig. 2f**). LysargiNase exhibited tri-, di- and moderate carboxypeptidase activity at lysine or arginine residues using synthetic peptides (**Supplementary Fig. 3**). Yet, when trypsin-digested BSA was used as substrate only 2 of 10 peptides were processed at their carboxy terminus and no amino- or diaminopeptidase activity was observed (**Supplementary Fig. 4**). LysargiNase proved suitable for in gel digestion of β-casein and bovine serum albumin (**Supplementary Fig. 5**).

LysargiNase aided peptide identification and hence proteome coverage by several means. Compared with trypsin, very different LysargiNase-fragment ion sequences retain the basic residue after CID ion trap fragmentation and so exhibited strong b-ion series (**Fig. 1b, Supplementary Fig.6a**). Corresponding spectra of tryptic peptides identified with identical cleavage site basic residues were dominated by y-type fragment ions (**Fig. 1b, Supplementary Fig. 7**). This led to different coverage of the precursor peptide sequence and resulted in different search engine scores. Secondly, corresponding peptides can differ where the cleavage site basic residue is different (**Supplementary Fig. 8**), which can also alter peptide ionization and fragmentation[1]. However, fragment ions of the y-series are more stable than b-ions[16], resulting in more tryptic peptide identifications than from LysargiNase (**Supplementary Fig. 6b**). Indeed, we observed extensive a-ion fragments in a LysargiNase proteome digest analyzed by beam-type CID fragmentation (**Supplementary Fig. 9a**). Including these a-ions in Comet XCORR scoring significantly increased quality (P<0.0001), but not the number of LysargiNase peptide identifications (**Supplementary Fig. 9b**). Hence, complementary use of both proteases increased proteome coverage—1,347 unique peptide sequences were identified in LysargiNase digests that were not identified following trypsin cleavage (**Supplementary Fig. 6b**). A considerable overlap in peptide identifications resulted from the complementary fragmentation spectra (n=1,558), increasing confidence in identification of these sequences. Finally, LysargiNase peptides included significantly more protein C-terminal peptides in four independent experiments by now having a lysine or arginine at their N-terminus compared with those generated by trypsin, where all basic residues are lost (P<0.001, **Fig. 1d**, **Supplementary Fig. 10**, **Supplementary Tables 4** and **5**).

We TiO_2-enriched SILAC-labeled phosphopeptides from MDA-MB-231 cells. The identified phosphosites were stringently filtered for those with a MaxQuant localization site probability >0.75 (ref. 17) and phosphorylation motifs were determined using MotifX[18]. Compared with trypsin, LysargiNase phosphopeptides showed a higher representation of RxS and RxxSP sites and phosphorylation motifs containing proline (**Fig. 1e** and **Supplementary Figures 11** and **12**). Thus, LysargiNase benefits detection of sites phosphorylated by certain kinases e.g. protein kinase C, ERK and CDK5, which prefer RxS and RxxSP motifs[17]. After 5 and 20 min stimulation with 12-*O*-tetradecanoylphorbol-13-acetate, phosphopeptides with these sites accumulated (**Supplementary Fig. 11**) in agreement with stimulation of the protein kinase C family[19].

We verified the LysargiNase cleavage specificity by PICS (proteomic identification of protease cleavage sites), which determines cleavage sites after incubation with database-searchable proteome-derived peptide libraries generated by chymotrypsin (**Fig. 2a**) and GluC- (**Fig. 2b**) digestion and subsequent dimethylation of lysine ε-amines[20]. The dimethyl-lysine modification restricted trypsin cleavage to arginine residues as reported[20], whereas LysargiNase also cleaved before the dimethylated lysines. Using MALDI-TOF-MS we verified this flexible specificity by demonstrating cleavage of synthetic peptide substrates containing dimethylated lysine, but not acetylated lysine (**Fig. 2c**). Further confirmation came from LysargiNase digestion of dimethylated *E. coli* proteome (n=2) in which 49% and 43%, respectively, of the beam type CID-identified peptides were cleaved before dimethylated lysine. In comparison, only 5% and 6% of tryptic peptides had a C-terminal dimethylated lysine (**Supplementary Table 6**). In MDA-MD-231 cell proteomes we also identified natural mono-methylated (**Supplementary Fig. 13**) and dimethylated (**Supplementary Fig. 14**) N-terminal lysines following LysargiNase digestion. Thus, LysargiNase accepts unmodified, mono-methylated and dimethylated lysine at P1' as shown in the structural model (**Supplementary Fig. 15**), but kinetically prefers unmodified lysine.

This unusual tolerance of methylated lysines is noteworthy as it is challenging to identify such peptides due to the presence of highly charged and often long tryptic peptides arising from missed cleavages[6]. Therefore, we posited that LysargiNase would also cut at methylated arginine. We found that LysargiNase cleaved mono-methylated and dimethylated arginines in MDA-MD-231 proteomes (**Supplementary Figures 13** and **14**). This was further characterized using synthetic peptides containing mono-methylated and asymmetric or symmetric dimethylated arginine (**Fig. 2d**), whereas trypsin cut essentially only after unmodified arginine.

Moreover, we observed a significantly higher number of peptides containing a methylated (P<0.05) or dimethylated (P<0.001) arginine or lysine in the LysargiNase-digested MDA-MD-231 proteome as compared to trypsin (Supplementary **Fig. 15b**). Overall, the thermophilic LysargiNase is a powerful complement to trypsin in proteomics.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

Accession codes. The mass spectrometry raw data and associated peptide identification details are deposited in ProteomeXchange with the dataset identifiers <PXD001113, PXD001114, PXD001121, PXD001122, PXD001378, and PXD001379>.

ACKNOWLEDGEMENTS

We thank N. Stoynov and J. Rogalski for mass spectrometry data acquisition, N. Scott and A. Frankel (both University of British Columbia) for stimulating discussions and providing the R1 peptides. We thank G. Butler for valuable editorial input. The German Academic Exchange Service (DAAD) and the Michael Smith Foundation for Health Research (MSFHR) supported P.F.H., P.F.L. was supported by the Alexander von Humboldt Foundation, the Breast Cancer Society of Canada and the MSFHR, L.D.R. was supported by the Canadian Institutes for Health Research (CIHR) and the MSFHR, and U.E. was supported by the MSFHR. This work was supported by a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology (C.M.O.), a grant from the CIHR as well as with an Infrastructure Grant from MSFHR and the Canada Foundations for Innovation (C.M.O.). Further support was provided by the European Union FP7 program, the Consolider Program of the Spanish Ministry of Science and Technology and the State Plan for Research in Science, Technology and Innovation of the Spanish Ministry of Economy and Competitiveness (F.X.G.-R.).

AUTHOR CONTRIBUTIONS

P.F.H. and C.M.O. conceived the experiments. T.G. and F.X.G.-R. produced the recombinant LysargiNase and structural models, P.F.H. performed all shotgun and PICS analyses, P.F.H. and L.D.R. performed the phosphosite experiments, N.S. and U.E. performed in-gel digests,

U.E. performed exopeptidase experiments. O.K. provided access to a QExactive instrument and performed mass spectrometer analysis of phosphopeptides, P.F.H., P.F.L., L.D.R., N.S. and C.M.O. analyzed the data, P.F.H., P.F.L. and C.M.O. designed the figures, L.D.R., N.S. and U.E. contributed to supplementary figures, P.F.H. and C.M.O. wrote the paper, and all authors edited the manuscript. C.M.O. supervised the project and provided grant support.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Tabb, D. L., Huang, Y., Wysocki, V. H. & Yates, J. R. *Anal. Chem.* **76,** 1243–1248 (2004).
2. Hohmann, L. *et al. J. Proteome Res.* **8,** 1415–1422 (2009).
3. Schilling, O., Barré, O., Huesgen, P. F. & Overall, C. M. *Nat. Methods* **7,** 508–511 (2010).
4. Van Damme, P. *et al. Nat. Methods* **7,** 512–515 (2010).
5. Gomis-Rüth, F. X. *Crit. Rev. Biochem. Mol. Biol.* **43,** 319–345 (2008).
6. Snijders, A. P. L., Hung, M.-L., Wilson, S. A. & Dickman, M. J. *J Am. Soc. Mass Spectrom.* **21,** 88–96 (2010).
7. Lange, P. F., Huesgen, P. F., Nguyen, K. & Overall, C. M. *J. Proteome Res.* **13,** 2028–2044 (2014).
8. Taouatas, N., Drugan, M. M., Heck, A. J. R. & Mohammed, S. *Nat. Methods* **5,** 405–407 (2008).
9. Gauci, S. *et al. Anal. Chem.* **81,** 4493–4501 (2009).
10. Taouatas, N. *et al. Mol. Cell. Proteomics* **8,** 190–200 (2009).
11. Taouatas, N., Heck, A. J. R. & Mohammed, S. *J. Proteome Res.* **9,** 4282–4288 (2010).
12. Raijmakers, R., Neerincx, P., Mohammed, S. & Heck, A. J. R. *Chem. Commun.* **46,** 8827–8829 (2010).
13. Tallant, C., Garcia-Castellanos, R., Seco, J., Baumann, U. & Gomis-Ruth, F. X. *J. Biol. Chem.* **281,** 17920–17928 (2006).
14. Tallant, C. *et al. Biological Chemistry* **388,** 1243–1253 (2007).
15. Vandermarliere, E., Mueller, M. & Martens, L. *Mass Spectrom. Rev.* **32,** 453–465 (2013).
16. Waldera-Lupa, D. M., Stefanski, A., Meyer, H. E. & Stühler, K. *Biochim. Biophys. Acta* **1834,** 2843–2848 (2013).
17. Rogers, L. D., Brown, N. F., Fang, Y., Pelech, S. & Foster, L. J. *Sci. Signaling* **4,** rs9–rs9 (2011).
18. Schwartz, D. & Gygi, S. P. *Nat. Biotechnol.* **23,** 1391–1398 (2005).
19. Mochly-Rosen, D., Das, K. & Grimes, K. V. *Nat. Rev. Drug Discov.* **11,** 937–957 (2012).
20. Schilling, O. & Overall, C. M. *Nat. Biotechnol.* **26,** 685–694 (2008).

Figure 1 | LysargiNase cleavage specificity and application in shotgun and phosphoproteomics. (**a**) Frequency distribution plots of cleavage sites identified by ion trap CID from human proteomes (n = 4) digested with LysargiNase (n = 1,917) and trypsin (n = 3,235). .Sequences were aligned at cleavage sites between P1 and P1' and normalized to the natural human amino acid abundances. (**b**) Exemplary spectra of the peptide KAAIDWFDG (precursor $[M+2H]^{2+}$

511.7507 m/z) identified in a LysargiNase digested proteome and the corresponding peptide AAIDWFDGK (precursor $[M+2H]^{2+}$ 511.7507 m/z) from a tryptic digest. (**c**) Sequence coverage by b-type and y-type fragment ions observed in 8,463 ion trap CID peptide spectra matches for trypsin and 5,474 for LysargiNase (FDR < 0.01). The centerlines show the medians, box limits indicate the 25th and 75th percentiles, whiskers extend to the 5th and 95th percentiles, and outliers are represented by dots. (**d**) Protein C- and N-terminal as a proportion of all peptides identified by shotgun proteomics after trypsin (white) or LysargiNase (grey) cleavage of proteomes (n = 4). (**e**) Phosphorylation motifs in trypsin and LysargiNase-digested proteomes identified by motif-x, number and proportion of matching high confidence phosphopeptides. having a localization probability > 0.75 identified by MaxQuant at an FDR < 0.01.. Phosphosites matching selected kinase specificity as extracted from the human protein reference database a43 shown as heatmaps. Significance of differences was tested using the two-tailed Student's t-test (*P < 0.05, **P < 0.01, ***P < 0.001).

Figure 2 | LysargiNase cleavage at methylated and dimethylated Arginine and Lysine. (**a**) Frequency distribution plots of LysargiNase (n = 126) and trypsin (n = 54) cleavage sites identified in chymotrypsin-generated *E. coli* proteome peptide libraries by PICS. K* designates dimethylated lysine. (**b**) LysargiNase sites (n = 316) and trypsin cleavage sites (n = 190) identified from *E. coli* GluC libraries by PICS. Sequences were aligned at cleavage sites between P1 and P1' and normalized to the natural *E. coli* amino acid abundances. (**c**) MALDI-TOF MS LysargiNase cleavage assays of RSGTLTYEAV↓KQTTD: peptide K1 (unmodified lysine and N-terminus, theoretical $[M+H]^+$ 1,669.83 m/z), peptide K1 with dimethylated lysine and N-terminus (K1-dm, $[M+H]^+$ 1,725.89 m/z), and peptide K1 with acetylated lysine and N-terminus (K1-ac, $[M+H]^+$ 1,753.85 m/z). The N-terminal LysargiNase cleavage products are RSGTLTYEAV (unmodified K1, $[M+H]^+$ 1,096.56 m/z; K1-dm, $[M+H]^+$ 1,124.59 m/z; and K1-ac, $[M+H]^+$ 1,138.59 m/z). The C-terminal cleavage product KQTTD is not observed. (**d**) MALDI-TOF MS cleavage assays of WGGYS↓RGGYGGW by LysargiNase and WGGYSR↓GGYGGW by trypsin: peptide R1 ($[M+H]^+$ 1,302.53 m/z), peptide R1 with mono-methylated arginine (R1-MMA, $[M+H]^+$ 1,316.53), peptide R1 with asymmetric arginine dimethylation (R1-aDMA, $[M+H]^+$ 1,330.56), and peptide R1 with symmetric arginine dimethylation (R1-sDMA, $[M+H]^+$ 1,330.56). C-terminal LysargiNase cleavage products are RGGYGGW (R1, $[M+H]^+$ 752.34; R1-MMA, $[M+H]^+$ 766.36 m/z; R1-aDMA and R1-sDMA, $[M+H]^+$ 780.37) and N-terminal tryptic cleavage products are WGGYSR (R1, $[M+H]^+$ 725.33; R1-MMA, $[M+H]^+$ 739.35 m/z; R1-aDMA and R1-sDMA, $[M+H]^+$ 753.36). The corresponding N-terminal LysargiNase cleavage product WGGYS

and the C-terminal tryptic fragment GGYGGW were not detected. For each peptide, overlaid MALDI-TOF MS spectra show control reactions (grey curve; bottom layer), LysargiNase digest (orange; top layer) and tryptic digest (light blue; middle layer). Substrate ion peak *m/z* is indicated in grey, product ion peaks in orange and light blue for cleavage by LysargiNase and trypsin, respectively. Spectra are normalized to the peak intensity of the most intense ion peak.

ONLINE METHODS

LysargiNase purification and activation. Wild-type LysargiNase (alias ulilysin) from *Methanosarcina acetivorans* was expressed and purified as the zymogen proform as described for the C269A mutant[13]. Aliquots of proLysargiNase were activated by stepwise addition 10 mM $CaCl_2$ buffered in 50 mM HEPES, pH 7.5 and incubated for 12 to 16 h at 20 °C.

Characterization of LysargiNase activity. LysargiNase activity was assayed with resorufin-labeled casein (Universal Protease Substrate, UPS) at an enzyme: substrate ratio of 1:400 in 50 mM HEPES, 10 mM $CaCl_2$, pH 7.5, with varying concentrations of chemicals and solvents commonly used in proteomics added as indicated (**Figures 1 and 2**). Assays were performed for 30 min at 37 °C or at the temperatures shown in the figure and stopped by addition of 5% TCA, followed by neutralization with 1.5 M Tris-HCl, pH 9.0. Fluorescence was measured using an excitation wavelength 544 nm and emission wavelength 590 nm on a fluorescence microplate reader (Molecular Devices fmax). These assay conditions were designed to generate less than 50% of the maximal fluorescence in the control reactions without added chemicals (data not shown) so as to quantify the effect of the different conditions. For proteome digestion efficiency assays, human MDA-MB-231 cell proteome was digested with LysargiNase (enzyme:substrate ratio 1:100) at 37 °C, 18 h in 50 mM HEPES,10 mM $CaCl_2$, pH 7.5, and the chemicals as indicated in the figure. Digestion efficiency was visualized by 12% SDS-PAGE and Coomassie brilliant blue R250 staining.

MALDI-TOF based exopeptidase activity assay. Synthetic peptides were dissolved and diluted 1:100 in HPLC grade water to a final concentration of 10 ng/µl. A tryptic BSA library was prepared by incubating 100 µg of BSA (Sigma) with 1.0 µg of Trypsin Gold (Promega) in 50 µl of 10 mM HEPES, 10 mM $CaCl_2$, pH 7.5, 37 °C. The digestion was stopped before completion after 2 h by adding the serine protease inhibitor AEBSF (Sigma) to a final concentration of 1.0 mM. LysargiNase was incubated with synthetic peptides (10 ng) or 1 µg of the tryptic BSA mixture in 10 mM HEPES, 10 mM $CaCl_2$, pH 7.5 for 18 h at 37 °C. Reactions were stopped by

adding trifluoroacetic acid (TFA) to a final concentration of 0.1%. One microliter of sample was spotted in duplicate on a MALDI target plate and mixed with the equal volume of alpha-cyano-4-hydroxycinnamic acid (CHCA) (5 mg/ml in 70% acetonitrile and 0.1% TFA). Samples were desalted on-plate by two washes with 1.0 µl of ice cold 0.1% TFA. Samples were analyzed using a Voyager-DE STR MALDI-TOF mass spectrometer (Applied Biosystems). Baseline reduction and noise removal were performed using Data Explorer™ version 4.5.

In vitro assays with synthetic peptide substrates. To test cleavage at modified lysine residues, synthetic peptide K1 (sequence RSGTLTYEAVKQTTD, obtained from Genescript) was modified by reductive dimethylation using isotopically light formaldehyde and sodium cyanoborohydride[21] or acetylated using NHS-activated acetate (Pierce). Unmodified or modified K1 peptides were incubated with LysargiNase or trypsin at a molar ratio of 1:100 in 50 mM HEPES, 5 mM $CaCl_2$ for 3h at 37°C. Cleavage at modified arginine residues was tested with peptide R1 (sequence WGGYSRGGYGGW), obtained with unmodified, monomethylated, asymmetric dimethylated or symmetric dimethylated arginine, was incubated with LysargiNase or trypsin at a molar ratio of 1:50 in 50 mM HEPES, 5 mM $CaCl_2$, pH 7.5 for 16 h at 37 °C. All peptide cleavage reactions were terminated dilution with 25 x excess of 5 mg/ml CHCA dissolved in 80% ACN, 0.1% TFA. Cleavage assays with synthetic peptide substrates were spotted on a MALDI target (1 µl of reaction mixture in 5 mg/ml CHCA in 80% ACN, 0.1% TFA), air-dried and analyzed using a 4700 series MALDI-TOF/TOF mass spectrometer (Applied Biosystems) in positive ion reflector mode.

Modeling of protein structure. Structural models of the LysargiNase are based on the experimental crystal structure of the enzyme with an arginine-valine dipeptide occupying the S_1' and S_1' pockets (Protein Data Bank access code 2CKI)[13]. Models were generated using the UCSF Chimera software suite[22] with LysargiNase displayed in standard orientation[23].

Cell culture and proteome preparation. For shotgun experiments, MDA-MB-231 cells (ATCC), previously tested for mycoplasma contaminations, were cultured in DMEM (Gibco) supplemented with 10% Cosmic Calf Serum, 10,000 U/ml penicillin G and 100 µg/ml streptomycin and non-essential amino acids (Gibco). Cells were detached using Versene and gentle scraping, collected by centrifugation and lysed in 100 mM HEPES, 150 mM NaCl, 5 mM EDTA, 5 mM EDTA, pH 7.5 supplemented with protease inhibitors (Complete Protease Inhibitors, Roche) and phosphatase inhibitors (PhosStop, Roche). Protein concentrations were

determined using the BioRad protein assay (BioRad). For the 12-*O*-tetradecanoylphorbol-13-acetate stimulation experiments, MDA-MB-231 cells were SILAC labeled with L-arginine (21 mg/liter) and L-lysine (36.5 mg/liter) for light-labeled (Arg0, Lys0) cells (Sigma-Aldrich), L-arginine-$^{13}C_6$ (21.75 mg/liter) and L-lysine-D_4 (37.5 mg/liter) for medium-labeled (Arg6, Lys4) cells, L-arginine-$^{13}C_6{}^{15}N_4$ (22.25 mg/liter) and L-lysine-$^{13}C_6{}^{15}N_2$ (38.5 mg/liter) or heavy-labeled (Arg10, Lys8) cells (Cambridge Isotope Laboratories) as previously described for HeLa cells[17]. Labeled cells were serum starved for 16 h (six 15 cm dishes per condition; 95% confluent cells) before stimulation. Light-labeled cells were treated with 0.1% DMSO in serum-free medium for 0 min (replicates 1 and 2) or 5 min (replicates 3 and 4) and medium- and heavy- labeled stimulated with 200 ng/ml 12-*O*-tetradecanoylphorbol-13-acetate, 0.1% DMSO in serum-free medium for 5 min and 20 min, respectively.

***E. coli* proteome preparation and digestion.** *E. coli* strain K12 was grown in LB broth before harvesting by centrifugation. Cells were washed with twice with 50 mM HEPES, pH 7.5 before snap freezing in liquid nitrogen. Cells were lysed by thawing in 50 mM HEPES with added protease inhibitor cocktail (100 µM PMSF, 1 µM 1,10-Phenantroline, 1 µM E-64, 1 µM pepstatin), pH 7.5, followed by sonication and centrifugation to remove cell debris. The soluble *E. coli* proteome was reduced and denatured in 2 M Guanidine hydrochloride and 1 mM TCEP (60 min, 60 °C), followed by alkylation with 5 mM iodoacetamide (45 min, 60 °C, dark). Two aliquots of the *E. coli* proteome were further modified by reductive dimethylation with 25 mM sodium cyanoborohydride and 25 mM light formaldehyde[24]. Both modified and unmodified *E. coli* proteomes were digested with LysargiNase and trypsin in 50 mM HEPES, 10 mM $CaCl_2$, pH 7.5 at an enzyme:proteome ratio of 1:100 (w/w), 37 °C, 18 h.

Shotgun proteome preparation. Generally, proteome aliquots of approximately 1 mg were denatured by addition of 3.5 M guanidinium-HCl, 10 mM DTT (30 min, 65 °C) and alkylated using 50 mM iodoacetamide (30 min, 20 °C, in the dark) and purified by chloroform/methanol precipitation[25]. For shotgun proteomics experiments, precipitated proteins were dissolved in 50 mM NaOH, immediately diluted 10-fold and neutralized by addition of water and buffer to a final concentration of 50 mM HEPES, 10 mM $CaCl_2$, pH 7.5. Proteins were digested with LysargiNase (proteome:enzyme ratio 50:1 w/w) or trypsin gold (Promega, 100:1 w/w) at 37 °C for 16 h. Note that this corresponds to an approximately equal mole enzyme/substrate ratio since the molecular mass of pro-LysargiNase (39 kDa) is almost twice that of trypsin (23.4 kDa). Digests were clarified by centrifugation, and an aliquot of 20 µg digested protein desalted using

C18 STAGE tips[26].

Phosphopeptide enrichment. For phosphopeptide analysis, protein pellets were dissolved in 8 M urea, 2 M thioruea to avoid loss of phosphorylation at high pH, diluted 5-fold with 50 mM HEPES, 10 mM $CaCl_2$, pH 7.5 and digested with LysargiNase (50:1 w/w) or trypsin gold (Promega, 100:1 w/w) for 16 h at 30 °C. Digests were clarified by centrifugation, peptides acidified and desalted with a C18 cartridge (SepPak, Waters). Phosphopeptides were enriched from four independent unstimulated MDA-MB-231 cell proteomes and from cell proteomes of four independent 12-*O*-tetradecanoylphorbol-13-acetate stimulation experiments at 5 min and 20 min. At each point, two replicates were enriched from two independent MDA-MB-231 cell proteomes using custom made TiO_2 tips as described[17] and two replicates were enriched using commercial TiO_2 phosphopeptide enrichment kit (Sigma). All peptide samples were desalted using C18 STAGE-tips[26] and stored on-tip until mass spectroscopy analysis.

Proteomic Identification of protease cleavage site specificity (PICS). PICS was performed as described[27]. In brief, *E. coli*-derived proteome peptide libraries were generated by digestion with GluC or chymotrypsin followed by amine-modification by reductive dimethylation. Libraries (200 µg) were incubated with trypsin or LysargiNase at an enzyme:peptide library ratio of 1:100 (w/w). Free N-terminal amine groups of fragment peptides, generated by substrate peptide cleavage, were tagged with an N-terminal modification with a redox-cleavable biotin moiety and enriched by streptavidin capture, eluted and desalted with C18 OMIX tips (Varian).

Mass spectrometry. For shotgun and phosphoproteomics experiments, peptides were eluted from STAGE tips in mobile phase 80% acetonitrile (ACN), 0.1 % formic acid, SpeedVac concentrated to near-dryness and dissolved in approximately 20 µl mobile phase 2% ACN, 0.1% formic acid. Ten microliters were loaded on the column. Shotgun experiments were analyzed on a nanoHPLC systems (Thermo Scientific) coupled to an LTQ-Orbitrap hybrid mass spectrometer (LTQ-Orbitrap XL, Thermo Scientific, operated by the UBC Centre for High Throughput Biology) using a nanospray ionization source consisting of a fused silica trap column (length 2 cm, inner diameter 100 µm, packed with 5- µm diameter Aqua C-18 beads; Phenomenex), fused silica fritted analytical column (length 20 cm, inner diameter 50 µm, packed with 3 µm diameter Reprosil-Pur C-18-AQ beads; Dr. Maisch GmbH) and a silica gold coated spray tip (20 µm inner diameter, 6 µm diameter opening, pulled on a P-2000 laser puller; Sutter Instruments; coated on EM SCD005 Super Cool Sputtering Device; Leica Microsystems).

Buffer A consisted of 0.5% acetic acid, and buffer B consisted of 0.5% acetic acid and 80% acetonitrile. Gradients were run from 0% B to 15% B over 15 min, then from 15% B to 40% B in the next 65 min, then increased to 100% B over 10 min period, held at 100% B for 30 min. The LTQ-Orbitrap was set to acquire a full-range scan at 60,000 resolution from (*m/z* 350–1800) in the Orbitrap and to simultaneously fragment the top five peptide ions in each cycle in the LTQ (minimum intensity 200 counts). Parent ions were then excluded from MS/MS for the next 180 sec. The Orbitrap was continuously recalibrated against protonated $(Si(CH_3)_2O)_6$; at *m/z* = 445.120025 using the lock-mass function[28].

Enriched phosphopeptides were analyzed by nanoHPLC-MS/MS using the LTQ-Orbitrap XL setup as above or with a hybrid quadrupole-Orbitrap mass spectrometer (QExactive, Thermo Scientific) coupled online with a RSLC nanoHPLC (Ultimate 3000, Thermo Scientific). Samples were loaded on a 100 µm, 2 cm nanoviper pepmap100 trap column (Thermo Scientific) in 2% ACN, 0.1% formic acid at a flow rate of 15 µl/minute. Peptides were eluted and separated at a flow rate of 300 µl/minute onto a RSLC nanocolumn 75 µm x 15 cm, pepmap100 C18, 3 µm 100 Å pore size (Thermo Scientific), with a linear ACN gradient from 2% to 24% in 0.1% formic for 25 minutes acid followed by a linear increase to 30% ACN in 0.1% formic acid over 5 minutes and additional increase up to 80% ACN in 0.1% formic acid over 5 minutes, followed re-equilibration at 2% ACN. The eluent was nebulized and ionized using a nanoelectrospray source (Thermo Scientific) with a distal coated fused silica emitter (New Objective) with a capillary voltage of 1.8-2.2 kV. The QExactive instrument was operated in the data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (*m/z* 350–1850) were acquired in the Orbitrap with 70,000 resolution (*m/z* 200) after accumulation of ions to a 3×10^6 target value with maximum injection time of 120 ms. Dynamic exclusion was set to 30 sec. The 10 most intense multiply charged ions ($z \geq 2$) were sequentially isolated and fragmented in the octopole collision cell by HCD with a fixed injection time of 60 ms 17,500 resolution and AGC target of 1×10^5 counts. A 2.7 Da isolation width was chosen. Underfill ratio was at 10% and dynamic exclusion was set to 30 sec. Typical mass spectrometric conditions were as follows: spray voltage, 2 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275 °C; normalized HCD collision energy 27%.

E. coli proteome digests and PICS assays were analyzed using a quadrupole time-of-flight mass spectrometer (QStar XL, Applied Biosystems, operated by the UBC Centre for Blood Research Mass Spectrometry Suite) coupled on-line to an LC Packings capillary LC system

(Dionex). Peptide samples were diluted in 0.3% formic acid, loaded onto a column packed with Magic C18 resin (Michrom Bioresources) and eluted using a 0–80% gradient of organic phase over 95 min. Buffer A was 2% ACN with 0.1% formic acid and buffer B was 80% ACN, 0.1% formic acid. ESI ionization tip ionization voltage was 25,000 V. MS data were acquired automatically using the Analyst QS software, v1.1 (Applied Biosystems) with information-dependent acquisition based on a 1s MS survey scan from 350 m/z to 1500 m/z, followed by up to 3 MS/MS scans of 2 s each. Nitrogen was used as the collision gas.

In gel digestion. The standard proteins BSA and β-casein were resolved by 12% SDS-PAGE. Bands were stained with Coomassie G-250 and then excised, destained thrice with 60% acetonitrile, 20 mM ammonium bicarbonate, washed in 100% acetonitrile and lyophilized. Gel bands were rehydrated with either 10 µL trypsin in 20 mM ammonium bicarbonate (12 ng/µL) or 10 µL LysargiNase (40 ng/µL) by passive diffusion for 1 h, 4 °C. Excess solution was removed, and 10 µL 20 mM ammonium bicarbonate 10 mM $CaCl_2$ was added to the gel plugs and digested overnight at 37 °C. Digests (1 µL on column) were analyzed by LC-MS/MS using an Agilent G4240A ChipCube interfaced directly to an Agilent G6550A Q-TOF mass spectrometer (Agilent Technologies). Peptides were diluted with buffer A (0.1% formic acid) and resolved by a 35 min linear gradient on a reversed-phase chip (75 µm x 150 mm, Zorbax 300SB-C18) from 3% buffer B (99.9% acetonitrile, 0.1% formic acid) to 60% buffer B at 300 nL/min. Peptides were ionized by ESI (1.8 kV) and mass spectrometry analysis was performed in positive polarity, with precursor ions detected from 350-1500 m/z. The top three ions per scan were selected for CID using a narrow (1.3 amu) exclusion window and an MS/MS scan rate of 2 spectra/second. Precursor ions were then excluded from further CID for 30 s. Data files were converted into .mgf files using Agilent MassHunter Qualitative Analysis B.06.00 and the files searched using Mascot v2.4 (Matrix Science). Search parameters included 0.3 Da tolerance for MS and MS/MS, UniProt-SwissProt database with 'all entries' enabled, variable oxidation of methionine, variable propionamide of cysteine, and a maximum of 2 missed cleavages.

Spectrum-to-sequence matching. Shotgun and phosphoproteomics data acquired with the Orbitrap XL and QExactive spectra where matched to peptide sequences in the human UniProt protein database (October 2013) with appended standard laboratory and common contamination protein entries and reverse decoy sequences (in total 177,324 entries) using the Andromeda algorithm[29] as implemented in the MaxQuant software package v1.4.12 (ref. 30), using an peptide FDR of 0.01. Search parameters included a mass tolerance of 5 ppm for the

parent ion, 0.5 Da for the fragment ions in LTQ-Orbitrap XL data and 20 ppm for fragment ions in QExactive data, carbamidomethylation of cysteine residues (+57.021464 Da), variable N-terminal modification by acetylation (+42.010565 Da), and variable methionine oxidation (+15.994915 Da). Non-specific searches of proteome digest were performed to probe the cleavage site specificity of LysargiNase. Specific searches of the same datasets used {X|KR} and {KR|X} as simple cleavage site rules for LysargiNase and trypsin, respectively, additionally considered Arg and Lys methylation (+14.01565 Da) and dimethylation (+28.031300 Da) as variable modifications and allowed up to three missed cleavages. Phosphoproteome data analysis additionally considered variable phosphorylation (+79.966331 Da) at Ser, Thr and Tyr, and in the case of 12-*O*-tetradecanoylphorbol-13-acetate stimulation experiments triple SILAC labeling (Arg +6.020129, Lys +4.025107 and Arg+10.008269, Lys +8.014199 for medium and heavy labels, respectively).

Peak lists of QSTAR XL data were created using the Analyst QS software v1.1 (Applied Biosystems) and converted to the mzXML format using msConvert. Spectra were matched to peptides sequences in the UniProt *E. coli* K12 proteome database (downloaded Nov 22, 2013) with appended standard laboratory and common contamination protein entries and reverse decoy sequences (9,134 entries total) using X!Tandem[31] (*E. coli* proteome data) or Mascot v2.3 and X!Tandem[31] (for PICS data) in conjunction with PeptideProphet[32] and iProphet[33] as implemented in the Trans Proteomic Pipeline v4.6(ref. 34) at an estimated FDR of 0.05. Search parameters included a mass tolerance of 200 ppm for the parent ion and 0.2 Da for the fragment ions, semi-specific restriction enzyme cleavages for chymotrypsin or GluC (PICS) or specific enzyme cleavage with LysargiNase or trypsin (*E. coli* proteome) with up to two missed cleavages. *E. coli* proteome data searches considered the following peptide modifications: Carbamidomethylation of cysteine residues (+57.021464 Da), dimethylation of lysine ε-amines (+28.031300 Da) and variable methionine oxidation (+15.994915 Da). PICS data searches also considered thioacylation of peptide N-termini (+87.998285 Da).

For analysis of a-ions in beam-type fragmentation data, QStar .wiff files were converted to .mzXML using msConvert and used as input for the TransProteomic Pipeline v4.7 rev1. Data was searched with Comet[35] as part of TPP using 0.2 Da tolerance for MS and MS/MS, LysargiNase specificity {X|KR}, variable oxidation of methionine (+15.994915 Da), 2 missed cleavages and scoring for y-ions and b-ions. Additionally, an independent search was used with a-ions also included as part of the scoring. XCORR scores for peptide spectrum matches were

extracted from pepXML files and compared to determine the a-ion scoring effect.

Data analysis. For shotgun proteomics peptide identifications the sequence context surrounding the N- and C-termini of the identified peptides was extracted from the database. For determination of the cleavage site specificity, all peptide termini identified by database searches without enzyme constraints were considered as LysargiNase or trypsin cleavage sites unless they mapped to protein termini. This resulted in the definition of {X|KR} as cleavage specificity of LysargiNase that was used in all subsequent searches. b-ion and y-ion counts were extracted from the MaxQuant result files, and fragment coverage calculated by dividing the number of matched ions by the sequence length.

Phosphorylation motives were extracted using the motif-x webservice[18,36]. Phosphorylated peptide identifications obtained in all experiments after digestion with trypsin or LysargiNase were searched separately using the following settings: foreground format: pre-aligned; central character: S; width: 15; occurrences: 20; significance: 0.000001; background: IPI human proteome. The motifs identified by were merged and for each motif the number of occurrences within the pool of phosphorylated peptides identified in the LysargiNase and trypsin datasets counted by regular expression matching. For comparison of the proportion of each motif in the LysargiNase and trypsin datasets, the occurrence was calculated as percent difference and significance tested using Pearson's chi-squared p-value. Only motifs describing at least 5% of the phosphorylated peptides identified by trypsin or LysargiNase digestion are reported.

To link observed phosphorylation motifs to kinase activity we retrieved known kinase sequence specificities from HRPD v9(ref. 37). For each identified phosphorylation motif all corresponding phosphorylation sites in our joint trypsin and LysargiNase datasets were counted, matched to kinase specificity by regular expression search and the percentage of sites potentially targeted by a specific kinase calculated. Only kinases matching at least one of the phosphorylation site motifs identified in the dataset are reported.

Protease cleavage sites were inferred by database searches from the peptides identified by mass spectrometry, which represent the prime side cleavage products of substrate cleavages, using a web-based tool (available at http://clipserve.clip.ubc.ca/pics) as described[27]. Cleavage site specificity logos were generated with the iceLogo web tool[38] using standard settings. Boxplots were generated using the BoxPlotR web server[39].

Online references

21. Kleifeld, O. *et al. Nat. Protoc.* **6,** 1578–1611 (2011).
22. Pettersen, E. F. *et al. J. Comput. Chem.* **25,** 1605–1612 (2004).
23. Gomis-Rüth, F. X., Botelho, T. O. & Bode, W. *Biochim. Biophys. Acta* **1824,** 157–163 (2012).
24. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. *Nat. Protoc.* **4,** 484–494 (2009).
25. Wessel, D. & Flügge, U. I. *Anal. Biochem.* **138,** 141–143 (1984).
26. Rappsilber, J., Mann, M. & Ishihama, Y. *Nat. Protoc.* **2,** 1896–1906 (2007).
27. Schilling, O., Huesgen, P. F., Barré, O., Auf dem Keller, U. & Overall, C. M. *Nat. Protoc.* **6,** 111–120 (2011).
28. Olsen, J. V. *et al. Mol. Cell. Proteomics* **4,** 2010–2021 (2005).
29. Cox, J. *et al. J. Proteome Res.* **10,** 1794–1805 (2011).
30. Cox, J. & Mann, M. *Nat. Biotechnol.* **26,** 1367–1372 (2008).
31. Craig, R. & Beavis, R. C. *Bioinformatics* **20,** 1466–1467 (2004).
32. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. *Anal. Chem.* **74,** 5383–5392 (2002).
33. Shteynberg, D. *et al. Mol. Cell. Proteomics* **10,** M111.007690–M111.007690 (2011).
34. Keller, A. & Shteynberg, D. *Methods Mol. Biol.* **694,** 169–189 (2011).
35. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. *Proteomics* **13,** 22–24 (2013).
36. Chou, M. F. & Schwartz, D. *Curr. Protoc. Bioinformatics* **Chapter 13,** Unit 13.15–24 (2011).
37. Keshava Prasad, T. S. *et al. Nucl. Acids Res.* **37,** D767–72 (2009).
38. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. *Nat. Methods* **6,** 786–787 (2009).
39. Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. *Nat. Methods* **11,** 121–122 (2014).