Journal of proteome research

Article

pubs.acs.org/jpr

Opposite Electron-Transfer Dissociation and Higher-Energy Collisional Dissociation Fragmentation Characteristics of Proteolytic K/R(X)$_n$ and (X)$_n$K/R Peptides Provide Benefits for Peptide Sequencing in Proteomics and Phosphoproteomics

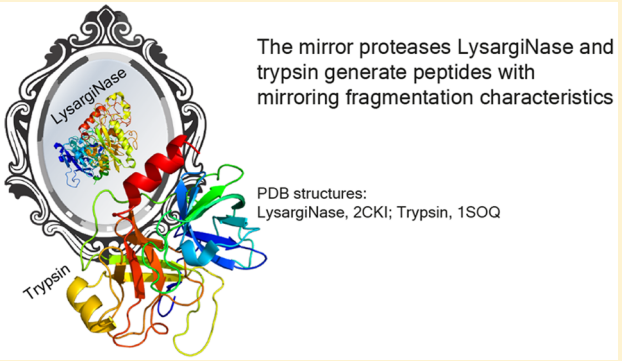
Liana Tsiatsiani,[†,‡] Piero Giansanti,[†,‡,∥] Richard A. Scheltema,[†,‡] Henk van den Toorn,[†,‡] Christopher M. Overall,[§] A.F. Maarten Altelaar,[†,‡] and Albert J.R. Heck*[,†,‡]

[†]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, and [‡]Netherlands Proteomics Centre, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[§]Centre for Blood Research, Department of Oral Biological and Medical Sciences, and Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver V6T 1Z3, BC, Canada

Supporting Information

ABSTRACT: A key step in shotgun proteomics is the digestion of proteins into peptides amenable for mass spectrometry. Tryptic peptides can be readily sequenced and identified by collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD) because the fragmentation rules are well-understood. Here, we investigate LysargiNase, a perfect trypsin mirror protease, because it cleaves equally specific at arginine and lysine residues, albeit at the N-terminal end. LysargiNase peptides are therefore practically tryptic-like in length and sequence except that following ESI, the two protons are now both positioned at the N-terminus. Here, we compare side-by-side the chromatographic separation properties, gas-phase fragmentation characteristics, and (phospho)proteome sequence coverage of tryptic (i.e., (X)$_n$K/R) and LysargiNase (i.e., K/R(X)$_n$) peptides using primarily electron-transfer dissociation (ETD) and, for comparison, HCD. We find that tryptic and LysargiNase peptides fragment nearly as mirror images. For LysargiNase predominantly N-terminal peptide ions (c-ions (ETD) and b-ions (HCD)) are formed, whereas for trypsin, C-terminal fragment ions dominate (z-ions (ETD) and y-ions (HCD)) in a homologous mixture of complementary ions. Especially during ETD, LysargiNase peptides fragment into low-complexity but information-rich sequence ladders. Trypsin and LysargiNase chart distinct parts of the proteome, and therefore, the combined use of these enzymes will benefit a more in-depth and reliable analysis of (phospho)proteomes.



KEYWORDS: *LysargiNase, trypsin, proteomics, phosphoproteomics, peptide fragmentation, electron-transfer dissociation, higher-energy collisional dissociation*

INTRODUCTION

Trypsin, the gold standard in shotgun proteomics,[1,2] has been extensively characterized and used in the generation of a tremendous amount of proteomics data.[3] However, in recent years, with the advent of new separation and peptide fragmentation methods,[4] novel proteases are being explored for applications in proteomics.[5−8] In this context, proteases with properties that would enable the complete unambiguous sequencing of proteins, peptides, and their modifications and generate peptides in a reproducible manner would further push the boundaries of MS-based proteomics.

In the past decade, a handful of new proteases have been explored for usage in proteomics including LysN,[9−11] OmpT,[12] α-lytic proteases,[13] and Sap9.[14] Of those, only LysN cleaves lysine residues at the amino side. Most recently, a thermophilic metalloprotease isolated from the Archaea species *Methanosarcina acetivorans*, LysargiNase (previously designated Ulilysin[15]), was explored for applications in proteomics.[16] Similar to LysN, LysargiNase is also a N-terminal protease, but it has the same preference for both lysine and arginine as trypsin.

Following electrospray ionization, tryptic peptides mostly attain two additional protons: one at the α-amine of the N-terminus and one at the basic side chain of arginine or lysine at the peptide C-terminus (Figure 1a). Due to their relatively short length, these tryptic peptides are highly amenable to collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD) techniques.[17] Backbone amide bond

Received: September 13, 2016
Published: November 21, 2016

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

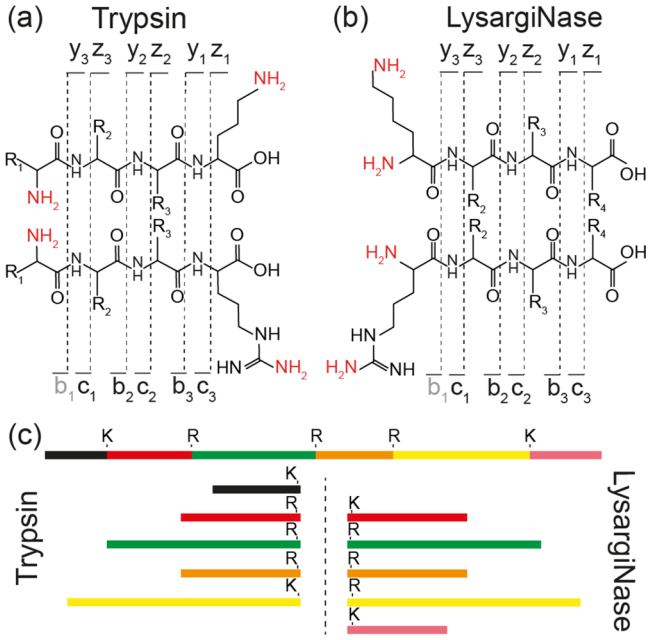


Figure 1. Roepstorff nomenclature[19] used in the gas-phase fragmentation of peptide ion precursors, for (a) tryptic and (b) LysargiNase peptides, with the proton locations in $[M + 2H]^{2+}$ ions highlighted in red. For both LysargiNase and tryptic peptides, the illustrations depict a lysine-cleaved (top) and an arginine-cleaved (bottom) peptide. The b_1 ions are depicted in gray because these species are never observed in peptides with free N-termini. (c) A protein is cleaved by trypsin and LysargiNase at the opposite N- and C-terminal sides of arginine and lysine residues, forming mirror peptides.

cleavages, with a preference to labile bonds, lead to the formation of many b and y fragment ions and rich peptide sequence data, wherefore the fragmentation rules are well-understood.[18] Mirroring trypsin, LysargiNase cleaves arginine and lysine residues exclusively at the N-terminal side. As a result, during ESI, doubly charged LysargiNase peptides predominantly attain two protons as well but are now both located at the N-terminus (Figure 1b), and collisional activation of these peptides results mostly in b fragment ions, which can be interpreted successfully in proteomics worksflows.[16] Because of their digestion patterns, LysargiNase and trypsin can be regarded as mirror proteases, generating in length and in sequence almost identical peptides, however, with swapped basic residues at the peptide termini (Figure 1c).

We previously showed that peptides that carry two protons at the N-terminus fragment predominantly into N-terminal c fragment ions during electron-transfer dissociation (ETD), facilitating straightforward peptide sequencing even when no a priori knowledge of the protein sequence is available.[9,20] Here, we analyzed thousands of peptides generated by LysargiNase or trypsin from a Jurkat cell lysate using ETD or HCD to investigate whether LysargiNase represents a suitable protease for ladder sequencing of peptides. We describe the fragmentation characteristics of these $K/R(X)_n$ (i.e., LysargiNase) and $(X)_nK/R$ (i.e., trypsin) peptides during chromatographic separation and gas-phase fragmentation and show the consequences of mirroring the location of the basic residue on peptide fragmentation, with the aim of probing the broader usability of LysargiNase in (phospho)proteomics applications. Even though LysargiNase and trypsin generate peptides of equal length and nearly equal sequence, our data reveal that these proteases chart different parts of the (phospho)proteome and lead to different observations in label-free quantitative proteomics.

EXPERIMENTAL METHODS

Sample Preparation and Phosphopeptide Enrichment

A detailed description of the method is included in the Supplementary Data section. In short, the lysate of Jurkat T lymphoma cells was prepared in a single step by combining cell lysis, protein reduction, and alkylation as previously described[21] with minor modifications. The cleared lysate was 10-fold diluted, either with 10 mM CaCl2 containing LysargiNase or 5 mM ammonium bicarbonate (Sigma-Aldrich) containing Trypsin Gold (Promega) in a 1:50 (enzyme-to-protein w/w) ratio and incubated for 12 h at 37 °C. At this point, the digestion was split in two, and one half was quenched with trifluoroacetic acid (Sigma-Aldrich) at a final concentration of 1% (v/v), while the other half was quenched to 5% formic acid (Merck, Germany). The aliquot that was quenched with formic acid was used for phosphopeptide enrichment using Ti^{4+}-IMAC, as previously described,[22] while the other half (quenched with TFA) was lyophilized and stored at −20 °C until liquid chromatography−tandem mass spectrometry (LC−MS/MS) analysis.

Mass Spectrometry, Protein Identification, and Data Analysis

A detailed description of the experimental methods is included in the Supplementary Data section. In short, for the analysis of the Jurkat cell lysates and the phosphopeptide enriched samples, an Agilent 1290 Infinity UHPLC system was coupled online to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) via a two-column vented setup[23] consisting of (1) a double frit 20 × 0.1 mm ID trapping column (Reprosil C18, 3 μm; Dr. Maisch) and (2) a single frit 50 × 75 μm ID analytical column (Poroshell 120 EC-C18, 2.7 μm; Dr. Maisch), both constructed in-house. Each sample was analyzed in triplicate by separating 1 μg of peptides over a 3 h gradient using 0.1% formic acid (Merck, Germany) (solvent A) and 80% acetonitrile (Biosolve, The Netherlands) with 0.1% formic acid (solvent B). In a data-dependent manner, the most intense peptide ions with a charge of z = 2+ and fitting within a 3 s cycle were selected for either HCD or ETD fragmentation using the Top Speed setting in the Orbitrap Fusion mass spectrometer and a dynamic exclusion time of 18 s. Full MS spectra were acquired in the Orbitrap at a resolution of 60 000 (full width half maximum, fwhm), while fragment spectra were acquired in the Orbitrap at a resolution of 15 000 (fwhm). ETD was supplemented with 30% normalized collision energy, while for HCD, normalized collision energy was set to 35%. For the phosphopeptide-enriched samples, the gradient length was reduced to 2 h, and the most-intense precursors were selected for HCD fragmentation with 35% normalized collision energy. Dynamic exclusion was set to 18 s and fragment ions were read out in the ion trap mass analyzer.

Tandem mass spectrometry (MS/MS) results were searched against the *Homo sapiens* Uniprot protein database (version 2015_04, 145 766 sequences) using the SEQUEST-HT[24] peptide search engine embedded in Proteome Discoverer (version 2.1). Percolator was used for validation of peptide spectrum matches (PSMs) with a 1% false discovery rate (FDR) based on q values.[25] The phosphorylation site

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

Table 1. Identification Statistics for Doubly Charged LysargiNase and Trypsin Peptides[a]

fragmentation	enzyme	protein groups	peptides	PSMs	input spectra	ID rate %[b]
ETD	LysargiNase	2380	10 379	11 485	27 956	41
	trypsin	2516	11 845	13 247	28 561	46
HCD	LysargiNase	3086	12 219	15 939	78 794	20
	trypsin	4535	23 308	32 262	86 998	37

[a]Identified protein groups, peptides, and PSMs for each combination of enzyme and fragmentation method. Experiments were performed in technical triplicates with an average standard error of 3%. [b]Identification (ID) rate is the percentage of PSM vs input spectra, i.e., MS2 events.

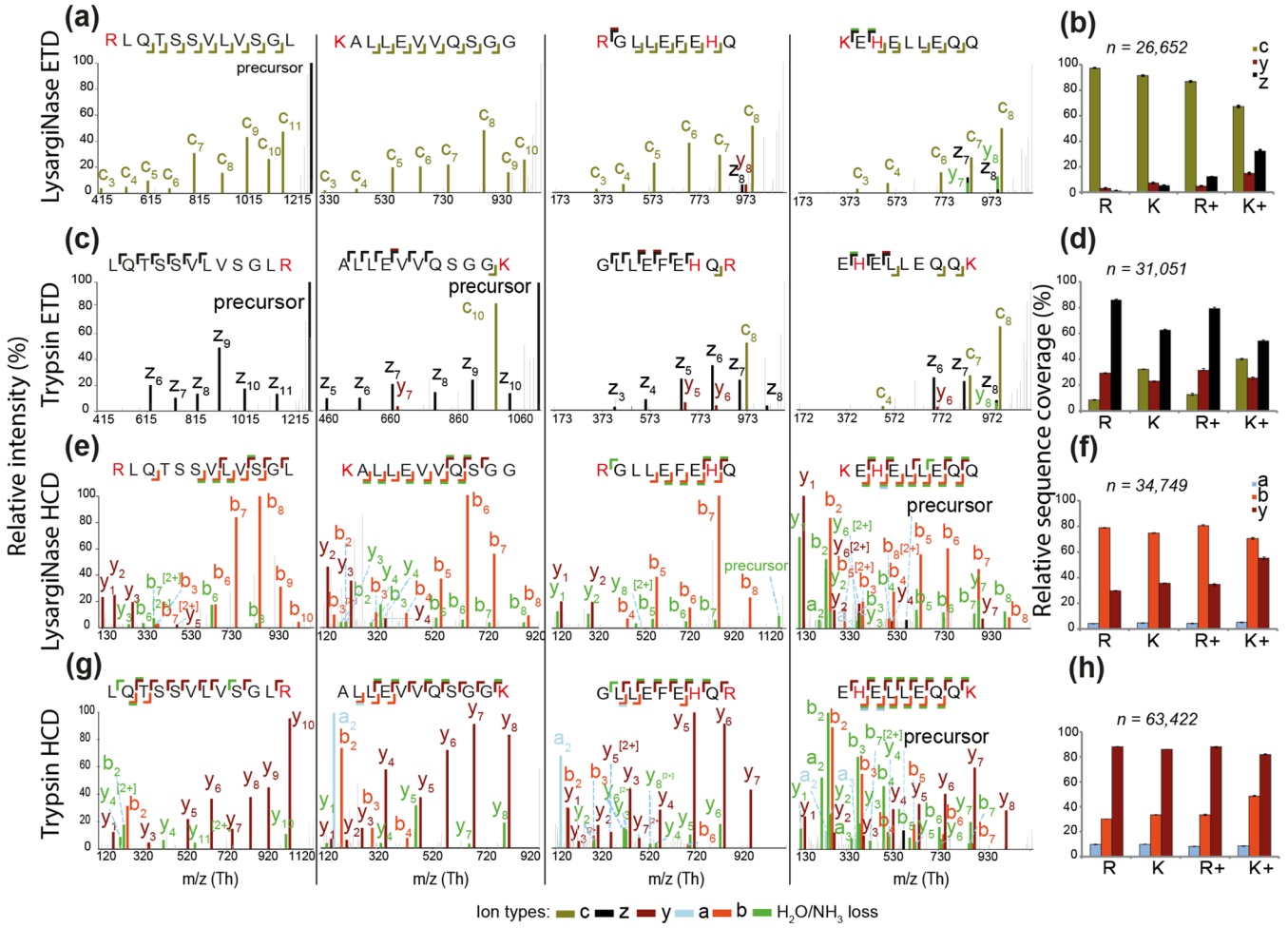


Figure 2. Fragmentation characteristics of proteolytic K/R(X)$_n$ and (X)$_n$K/R peptides. N-terminal or C-terminal protons drive the formation of opposite but complementary ion patterns for the alike peptides during ETD and HCD, with basic residues depicted in red. (a) ETD MS/MS spectra of LysargiNase peptides with a single basic residue or with multiple basic residues. (b) Average relative peptide sequence coverage (%) after proteome-wide ion counting in LysargiNase peptides and ETD. (c) ETD MS/MS spectra of tryptic peptides with a single basic residue or with multiple basic residues. (d) Average relative peptide sequence coverage (%) after proteome-wide ion counting in tryptic peptides and ETD. (e) HCD MS/MS spectra of LysargiNase peptides with a single basic residue or with multiple basic residues. (f) Average relative peptide sequence coverage (%) after proteome-wide ion counting in LysargiNase peptides with HCD. (g) HCD MS/MS spectra of tryptic peptides with a single basic residue or with multiple basic residues. (h) Average relative peptide sequence coverage (%) after proteome-wide ion counting in tryptic peptides with HCD. Peptides clusters represent arginine-cleaved peptides (R), lysine-cleaved peptides (K), arginine-cleaved peptides containing additional basic residues in the peptide sequence (R+), and lysine-cleaved peptides containing additional basic residues in the peptide sequence (K+). Fragment ions in green have undergone water or ammonia loss. Error bars represent the 95% confidence interval. Data are combined from three replicate experiments.

localization of the identified phosphopeptides was performed by phosphoRS algorithm (ptmRS),[26] also embedded in Proteome Discoverer. All of the raw data files have been deposited to the ProteomeXchange Consortium via the PRIDE[27] partner repository with the data set identifier PXD004447.

Peptide fragment ion annotation, in a proteome-wide scale, was performed with in-house tooling software capable of extracting spectral data from the RAW files file utilizing MSFileReader (Thermo Fisher Scientific) that performs peak detection, isotope deconvolution, and peptide fragment annotation based on the peptide sequence extracted from Proteome Discoverer. For the analysis of MS/MS peak

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

intensities, raw spectral data were extracted from the RAW files with the DLL MSFileReader (Thermo Scientific). These data were then further processed by centroiding, deisotoping, and filtering for basepeak intensity percentage 3% to come to the final fragmentation spectrum. The output from Proteome Discover (PSM table) was used for peptide identities to annotate all spectra with fragment ions (HCD: a/b/y + neutral losses; ETD: c/y/z + neutral losses). The intensities of the annotated peaks were located according to their peptide sequence position. Visualizations were made with the function "heatmap.2" in the R scripting and statistical environment[28] extended with "gplots".

RESULTS

Peptide Fragmentation Characteristics

We analyzed the fragmentation behavior of thousands of tryptic and LysargiNase peptides in a proteome-wide scale (Jurkat cell lysate) using ETD and HCD. Despite the shared amino acid preference of the two enzymes and the fact that the technical reproducibility (i.e., overlap in identifications) of the performed mass spectrometric runs was on average more than 65% (Supplementary Table 1); after removal of the N- or C-terminal cleavage sites, there was only a 30−35% overlap in peptide identifications between the two digests (Supplementary Figure 1). In fact, there were about 5400 unique peptide identifications (28%) in the LysargiNase ETD data set and 6277 (19%) in the HCD data set that were not identified by using trypsin (Supplementary Figure 1). This already reveals that as for CID,[16] and also with HCD and ETD, there is a large degree of orthogonality in proteome coverage using LysargiNase and trypsin.

The number of PSMs, peptide groups and protein identifications for both LysargiNase and tryptic proteolytic preparations was comparable, revealing that LysargiNase and trypsin perform equally well for ETD-based shotgun proteomics (Table 1). About 28 000 tandem mass spectra were acquired for each enzyme during ETD fragmentation, and upon peptide to spectrum matching, this resulted in 10 379 and 11 845 unique peptide identifications for LysargiNase and trypsin, respectively, at 1% FDR. Irrespective of the enzyme used, during HCD, the number of acquired MS/MS spectra was approximately three times higher than with ETD. This is due to the required higher ion load and increased duty cycle in ETD, as previously reported.[29] Even though the number of HCD MS/MS spectra did not differ substantially between the two proteolytic digests (Table 1), the identification rate for LysargiNase (20%) was about half that of trypsin (37%). To determine which factors, other than cleavage specificity, influence the identification of the different peptide sets with MS/MS, we examined the fragment ion composition of LysargiNase ETD and HCD spectra with respect to trypsin in a proteome-wide scale.

Electron-based dissociation methods, such as electron-capture dissociation (ECD)[30] or electron-transfer dissociation (ETD),[31] provide alternative and complementary methods to CID, which generally are accepted to be better suited for multiply charged, longer peptides, or intact proteins and for the analysis of labile peptide modifications.[32] Considering the charge arrangement, fragmentation of LysargiNase peptides should, in theory, lead to intense ions that retain the charge at the N-terminus (i.e., c-ion types), while tryptic peptides should fragment in mixed populations of N- and C-terminal charge-retaining ions (i.e., c- and z-ions) (Figure 1a,b).

We selectively analyzed the fragmentation of only doubly charged LysargiNase and tryptic peptide precursor ions, as these are the most abundant species in our data sets and due to the particular property of doubly charged peptides with N-terminal charges to fragment solely in c-ions during ETD. We saw that, irrespective of whether LysargiNase peptides contained an arginine or lysine at the N-terminus, ETD spectra were almost exclusively composed of c fragment ions (Figure 2a), mapping more than 90% of the annotated LysargiNase peptide sequences (Figure 2b). Such behavior has previously been described for lysine-initiated LysN peptides;[9] however, as shown here, LysargiNase extends this behavior to arginine-initiated peptides. The formation of the c fragment ions for LysargiNase peptides is driven by the clustering of the two protons at the N-terminus. This unique specificity is somewhat lost when additional basic amino acids are present in the remainder of the peptide sequence, either due to the missed cleavage of arginine or lysine or the presence of histidine residues, causing the possible formation of complementary z-ions (Figure 2a,b). In lysine-cleaved peptides with additional basic residues (K+), the complementary z-ions were more frequent than in arginine-cleaved peptides (R+), likely due to the lower basicity of lysine (compared to arginine) in the gas phase (Figure 2b). Complementary c/z-ions were almost always observed in the tryptic peptide data sets irrespective of the number of basic residues in the peptide sequence (Figure 2c,d). This is because in tryptic peptides, the protons are captured at either the N-terminal amine or the C-terminal base (arginine or lysine) and lead to the formation of complementary c/z fragment ion populations. Nevertheless, as the side chain of arginine exerts higher proton affinity in the gas-phase than lysine or the N-terminal amine,[33] arginine-ending tryptic spectra were dominated by C-terminal z-ions as clearly observed in our data (Figure 2c,d). Lysine-ending peptides, however, formed more varied fragment ion populations, consisting of about 70% z-ions and 30% c-ions (Figure 2d). As before, the complementarity of ions increased to 60% z-ions and 40% c-ions when additional basic residues were present in the remainder of the peptide sequences.

In accordance with the fragmentation pattern of LysargiNase peptides with CID,[16] the HCD spectra of LysargiNase peptides consisted mostly of b-ions (80%) (Figure 2e,f), whereas the majority of tryptic fragment ions were y-ions (90%) (Figure 2g,h). Again, more-varied ion compositions were observed for lysine-cleaved peptides containing more than one basic residue (Figure 2e−h). As for ETD, the dominant basicity of arginine over lysine in peptides with a single basic residue led to higher sequence coverage by N- or C-terminal ion fragments, depending on the position of the arginine residue. In this manner, under HCD conditions, arginine-initiated peptide spectra contained more b-ions than the lysine-initiated peptides in the LysargiNase digestions (Figure 2e,f) and the same was observed for y-ions and tryptic peptides (Figure 2g,h). A larger selection of illustrative spectra is provided in Supplementary Figures 4−19 accompanying this manuscript, and ion frequencies are provided in Supplementary Table 2.

All data presented in Figure 2 are based on the frequency of occurrence and not on the intensity of each particular fragment ion. Therefore, next we constructed ion fragmentation heat maps by incorporating the intensity of fragment ions as a function of fragment ion size (amino acid residue number)

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

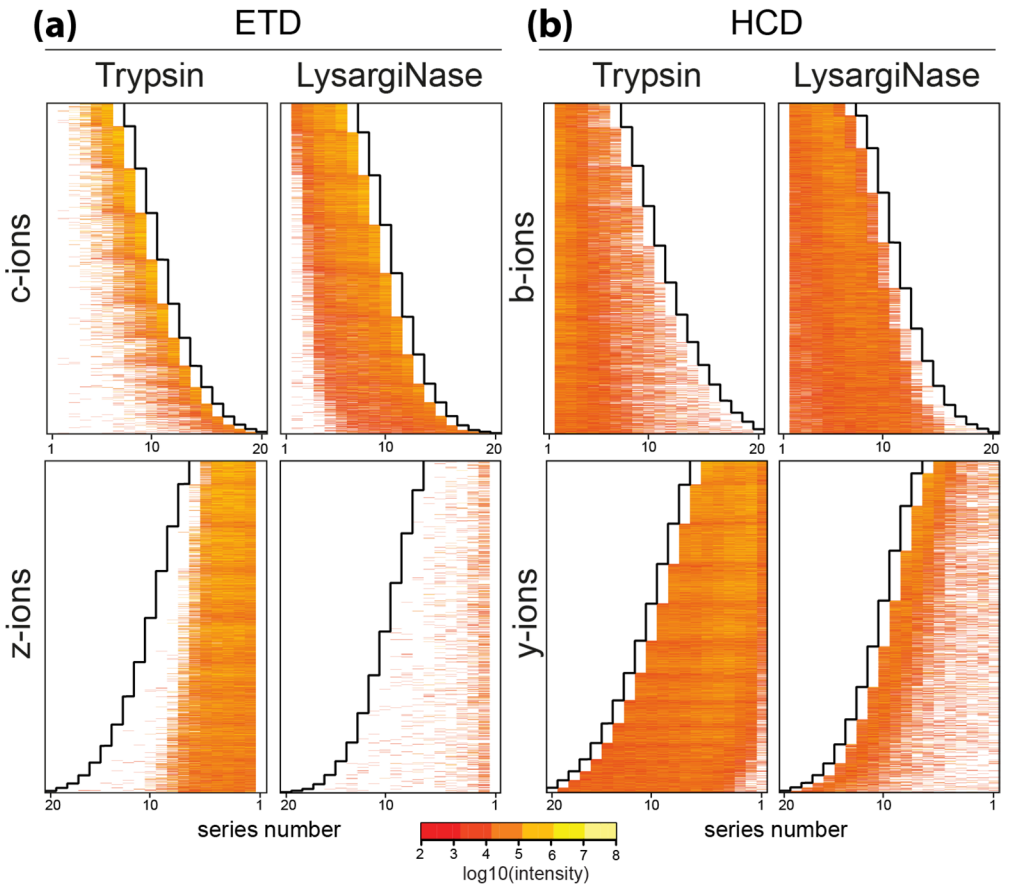


Figure 3. Fragmentation heat maps based on ion intensities for $K/R(X)_n$ and $(X)_nK/R$ peptides during HCD and ETD. Normalized relative intensity values were calculated for peptides 6−20 amino acids long. Series numbers are matched to the sequence orientation.

(Figure 3). As we previously described, during ETD, tryptic peptide backbones fragmented into homologous z- and c-ions and, most frequently, z-ions. In fact, peptide sequences were annotated by low-order z-ions and complementary high-order c-ions reflecting the higher probability for initial electron transfer to the charged C-terminal site. No obvious differences in ion intensities were observed between c- and z-ion types (Figure 3a). Once again, almost complete sequence coverage of LysargiNase peptides was achieved with c-ions, and as expected, very few, and mostly never very intense, z-ions were formed. Therefore, we were not able to compare the respective intensities. With HCD of tryptic peptides, almost complete y-ion series were formed, consistent with the presence of a localized proton at the C-terminal basic side chain, and medium-sized b-ions formed by the transfer of a mobile proton[18] alongside the peptide backbone originating from the peptide N-terminus. Consequently, the degree of ion complementarity was high, and amino acid residues were frequently annotated by both b- and y-ion types (Figure 3b). In contrast, for peptides generated by LysargiNase, the localized proton on the N-terminal basic side chain drove the formation of full b-ion series that were only partially complemented by high-order y-ions. Although y-ions seemed to be slightly more intense in short $(X)_nK/R$ peptides, we did not observe any obvious differences in overall HCD ion intensity within ion types or digests.

To further address the issue of the substantial lower HCD identification rates in LysargiNase searches (Table 1), we compared the total number of detected ions in LysargiNase MS/MS spectra to those in the trypsin data sets. Interestingly, the distribution of ions in the two digests differed significantly and fewer ions were generated and detected in the LysargiNase HCD fragmentation events (Supplementary Figure 2a). Considering the mobile-proton model,[18] charge sequestration at the N-terminus of $K/R(X)_n$ peptides hampers random proton mobility, and thus, HCD−CID fragmentation is limited to the peptide regions with higher charge states. The model corroborates the lower number of detected ions in LysargiNase HCD MS/MS spectra and may explain the lower HCD identification rates. LysargiNase performs suboptimal with the default settings used in most of the current HCD workflows because possibly higher collision energy is necessary to induce adequate peptide backbone fragmentation of such peptides. With ETD the number of detected ions was comparable between the two digests (Supplementary Figure 2b) because ETD settings were optimized prior analysis with 30% supplemental collision energy.

Retention Characteristics of $K/R(X)_n$ and $(X)_nK/R$ Peptides during LC Separation

In shotgun proteomics of complex mixtures, peptide identification heavily relies on chromatographic separation. Amino acid composition, peptide length,[34] and sequence-dependent effects[35] have been shown to affect peptide retention on chromatography columns. Because local side-chain interactions of amino acids with their immediate neighbors may affect peptide retention, we reasoned that differential positioning of the basic cleavage site in LysargiNase and tryptic peptides could affect their retention during reverse-phase ultrahigh-pressure

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

liquid chromatography (RP-UPLC). Therefore, we analyzed LysargiNase and tryptic peptides that were derived from the digestion of bovine serum albumin (BSA) with LC−MS/MS. To examine the effect of the proton location on C18 retention, we compared peptides with identical sequences only differing at the terminal side by the basic residue. To enable the selection of these peptides, the basic cleavage sites were in silico removed from the sequence. For each peptide, the retention time (t_R) was extracted from Proteome Discoverer and plotted against the retention of its variant carrying the basic cleavage site at the opposite terminus (Figure 4). A pair of additional BSA tryptic

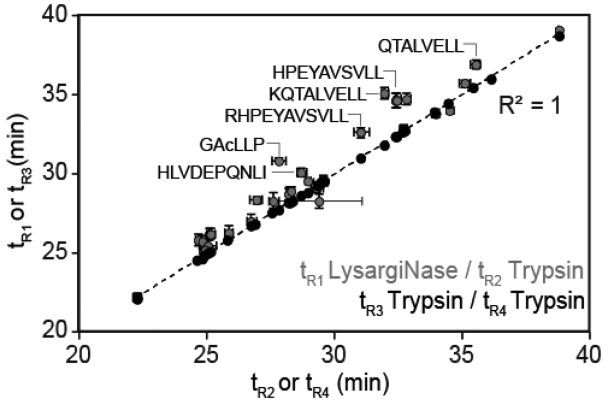


Figure 4. Retention time (t_R) plot for alike LysargiNase and tryptic peptides. LysargiNase peptides (t_{R1}) are plotted against their tryptic variants (t_{R2}) in the gray series. Control peptides, in this case tryptic peptides, exhibit identical and reproducible retention times (t_{R3} vs t_{R4}) within consecutive chromatographic separations (trend line in black series). The N- or C-terminal cleavage sites are removed from the LysargiNase and tryptic peptide sequences, respectively, to enable the selection of identical sequences that only differ on the location of the basic residue. Error bars represent the standard deviation of the mean (n = 3).

analyses were acquired as a measure for the retention behavior of identical peptides. As expected, identical tryptic peptides exhibited reproducible retention times with high correlation (R^2 = 1). On the contrary, most of the LysargiNase peptides were retained longer on C18 than their tryptic counterparts, sometimes even up to 3 min (Figure 4 and Supplementary Table 3). This observation confirms that the location of the basic residue in the peptide sequence indeed affects the elution profile of peptides and in fact, LysargiNase peptides are retained somewhat longer on C18. Interestingly, this effect was most pronounced in LysargiNase peptides bearing C-terminal hydrophobic binding domains,[36] such as GACLLP, RHPEYAVSVLL, HPEYAVSVLL, KQTALVELL, DDSPDLP, or YICDNQDTISSKL. Instead, in the tryptic peptide variants where the C-terminal hydrophobicity was counteracted by the hydrophilic lysine or arginine, binding to the C18 stationary phase was weaker, resulting in shorter retention times (Figure 4 and Supplementary Table 3).

LysargiNase and Trypsin in Protein Quantitation

As shown here, LysargiNase and tryptic peptides are intrinsically different, albeit sharing, to a large extent, identical amino acid compositions. We hypothesized that this may also affect their identification probability in MS-based proteomics. A major generic aim in MS-based proteomics is protein quantitation, whereby label-free quantitation (based, for instance, on peptide identification and total number of spectral counts per protein) has recently gained interest.[37−39] However, we have shown before that estimation of protein concentration may be affected when samples are digested with different proteases and that protein abundances correlate better when proteases of similar specificity are used for sample preparation.[7] To investigate whether the shared specificity of LysargiNase and trypsin leads to comparable estimated concentrations for a given protein, we estimated the concentration of the 1000 most abundant Jurkat cell proteins using a well-established spectral count based approach. In a correlation matrix, we observed high reproducibility for the normalized protein spectral counts across technical replicates and experiments ($r^2 \approx 0.9$) when a single protease was used (Figure 5 and Supplementary Table

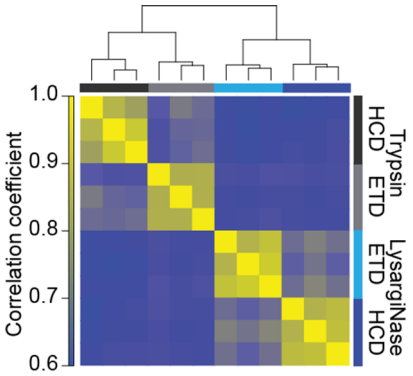


Figure 5. Correlation in label-free quantitation of proteins in data sets gathered by using LysargiNase and trypsin in combination with ETD or HCD. Protein levels for the 1000 most-abundant human proteins are estimated based on normalized spectral counts.

4). High correlation was also observed among samples of the same protease that were analyzed with either HCD or ETD ($r^2 \approx 0.8$). However, comparison of protein abundance extracted from the LysargiNase versus the tryptic samples notably diminished the correlation to less than 0.7. Thus, we show that despite the shared similar specificity of trypsin and LysargiNase toward arginine and lysine, protein quantification leads to different results due to the aforementioned differences in fragmentation, identification, and elution of the same peptide variants.

LysargiNase Bias in Phosphoproteome Analysis

A number of strategies have been explored to improve the identification, quantification, and site-localization of phosphorylated peptides.[40] Determination of the true phosphorylation sites can be even more challenging than identifying the phosphopeptide itself, which is caused by the necessity of specific diagnostic backbone fragment ions to confidently localize the phosphorylated residue.[41] Currently, several scoring methods exist to process large-scale phosphoproteomics data sets. However, these rely on probability-based algorithms that do not cope easily with closely located phosphorylated residues. Ambiguous or erroneous assignment of site localization can occur especially when spectra are noisy or when backbone fragmentation is not complete, and only a few specific ions can pinpoint the exact phosphorylation site.[42]

As we described above for unmodified LysargiNase peptides, phosphopeptides also attain the protons at the N-terminus. As a consequence, ETD of doubly charged LysargiNase phosphopeptides results in spectra that consist almost exclusively of c-type fragment ions and are largely simplified

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

over HCD or CID spectra (Supplementary Figure 3), as they also do not suffer from abundant losses of the neutral phosphate moiety. Therefore, LysargiNase, used in combination with ETD, provides a clear advantage over carboxyl side proteases for spectral interpretation because the generated simple (phospho)peptide sequence ladders are very easy to interpret. Because generally, nearly full ion series are generated, the exact site of phosphorylation can be readily determined.

To further investigate the applicability of LysargiNase in phosphoproteomics applications, we executed selective analysis of enriched phosphopeptides. In short, we performed three independent Ti^{4+}-IMAC enrichments on the LysargiNase and tryptic Jurkat cell digests to specifically enrich for phosphopeptides, using just 200 μg of peptide material per enrichment. The data from the three independent enrichments revealed that a high enrichment selectivity (85%) for phosphopeptides could be achieved from the LysargiNase digest, on par with the enrichment efficiency obtained for trypsin.[8] Combined use of both proteases increased the phosphoproteome coverage by about a quarter (23%) (Figure 6a). We recovered 12 280 unique phosphorylation sites (from 19 072 unique phosphopeptides) in the tryptic digest, whereas around 7000 unique phosphosites (from 10 258 unique phosphopeptides) were recovered from the LysargiNase digests (Supplementary Table 5). Again, the lower number of LysargiNase phosphopeptides is likely due to the fact that the analysis of the phosphopeptide-enriched samples was done with HCD and less-favorable collision settings.

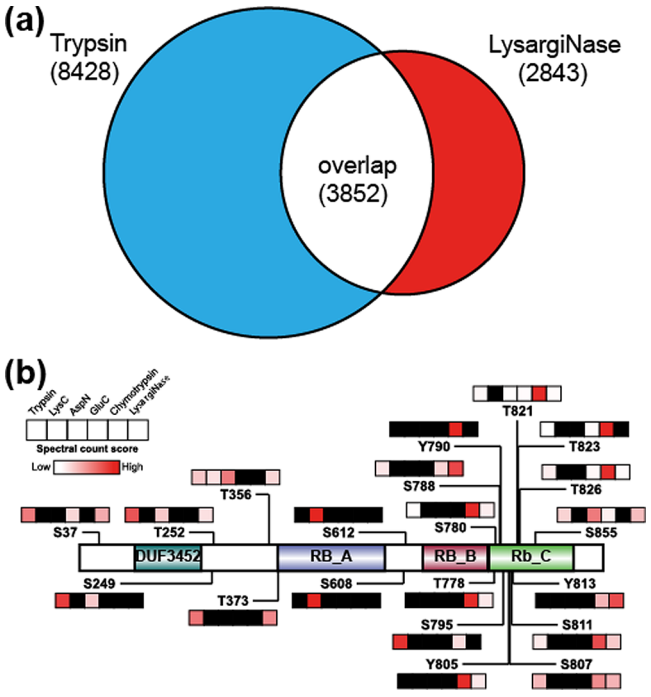


Figure 6. Complementary contribution of LysargiNase to the coverage of the phosphoproteome. (A) Venn diagram displaying the overlap in detected unique phosphosites between the trypsin and LysargiNase data sets. Only 3852 were shared between the two data sets, indicative of the high orthogonality of the two proteases. (B) Domain structure of an illustrative highly phosphorylated oncoprotein (RB1). A spectral count score depicts the detectability of each phosphosite by the different proteolytic preparations. Black color indicates not detected sites. LysargiNase is the optimal protease to detect for instance S788 and Y813.

In phosphoproteomics, certain protein phosphorylation events are consistently better-identified by certain proteases over others.[8] To investigate the potential contribution of the new LysargiNase phosphoproteome data set, we incorporated the here-acquired phosphoproteomics data into PhosphoDB (http://phosphodb.hecklab.com). Our analysis corroborates previous findings[16] and clearly reveals that LysargiNase is a protease that would substantially facilitate the detection of thousands of phosphosites by MS. In Figure 6b, we selected a representative phosphoprotein, the tumor suppressor RB1, which is involved in the regulation of cell division.[43] Throughout MS-based phosphoproteomics, a multitude of phosphorylation sites have been reported for this protein (60 according to the latest release of PhosphoSitePlus). Ideally, all the sites should be monitored when performing a targeted or untargeted phosphoproteomics experiment, as some of them are known to modulate the activity of the protein. Whereas the majority of them can be detected by the two most commonly used proteomics-grade enzymes, trypsin and LysC, the remaining phosphosites remain inaccessible, and for particular sites (e.g., Y813 and S788), LysargiNase seems to be the protease of choice for optimal detection.

DISCUSSION

Compared to other proteases used for proteomics,[3,44] LysargiNase exhibits unique properties generating peptides of identical length and similar sequence content as trypsin. We investigated, in the context of their use in proteomics, the characteristics of these peptides $K/R(X)_n$ and $(X)_nK/R$, as generated by LysargiNase and trypsin. We described the ETD and HCD fragmentation characteristics of these peptide variants and extended our observations with respect to protein quantitation and post-translational modification analysis. Our main conclusion is that although LysargiNase and trypsin share the same residue specificity, their digestion products have different chemical properties and generate different but complementary proteomics data sets.

Of all of the proteomics data deposited in public depositories, more than 90% are based on workflows using trypsin.[3] The characteristics of tryptic peptides and their fragmentation rules under HCD and ETD are thus also very well-known and used in database search algorithms. If we want to use LysargiNase equally prominently in proteomics workflows, we need to know equally well the fragmentation characteristics of LysargiNase peptides so that instrument settings as well as search algorithms can be adapted accordingly. Based on the fact that the basic residues are switched over to the N-terminus in LysargiNase peptide precursors, b-ions are predominantly formed, being a clear unique characteristic of LysargiNase peptides (Figure 2e,f). For ETD, LysargiNase spectra are easy to interpret as they nearly exclusively contain c ions in both arginine- and lysine-cleaved peptides (Figure 2a,b). When additional basic residues are dispersed in the peptide sequence, this c-ion specificity is somewhat lost and c/z-ions are concurrently present. Further, our data revealed that the degree of complementarity in ion types depends on the basicity of the cleaved residue irrespective of its positioning. This way, fragment spectra of arginine-cleaved peptides were more homogeneous than lysine-cleaved peptides, even when additional basic residues (lysine or histidine) were included in the peptide sequence. Under the tested ETD conditions, the

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

sequencing performance of LysargiNase and trypsin are alike despite the opposite yet complementary fragmentation patterns.

Using HCD fragmentation, we observed that LysargiNase performed substantially weaker than trypsin. The low identification rate that we observed in this data set could be possibly explained by the fact that b-ions, the predominant ions formed by HCD fragmentation of LysargiNase peptides, are relatively less-stable than y-ions during CID.[45] In principle, this could explain why tryptic peptides were annotated with y-ions up to 90%, while the equivalent b-ion annotation of LysargiNase peptides reached 10% lower (80%) (Figure 2b,d). However, in $K/R(X)_n$ peptides, the proton mobility, the catalyst for peptide backbone fragmentation under HCD conditions,[18] is inhibited by the strong charge sequestering at the peptide N-terminus. Even so, we did not observe any notable difference in spectra quality between the two digests, and this is because low-quality spectra were filtered out by the search engine, which returned less LysargiNase hits (Table 1). This issue can possibly be dealt with by using higher collision energies. However, to set the optimal fragmentation settings, caution needs to be taken to avoid possible secondary fragmentation events. With ETD, we did not observe any significant difference in the performance of LysargiNase and trypsin because ETD was supplemented with additional 30% collision energy, and this was sufficient for the identification of $K/R(X)_n$ peptides.

Additionally, we analyzed the phosphoproteome originating from LysargiNase and trypsin digests. As initially reported[16] as well, our data reveal that through digestion with these proteases, we chart different parts of the phosphoproteome. Even though there were fewer identified LysargiNase phopshopeptides than tryptic phopshopeptides, we were able to enrich distinct phopshopeptides (Figure 6). In addition, when LysargiNase was used in combination with ETD, the spectra consisted of easy-to-interpret sequence ladders and, in most cases, nearly-full ion series, allowing for the clear determination of the phosphorylated residue.

The LysargiNase and tryptic peptide sets also varied with respect to their chromatographic properties. Obviously, removal of the basic side chain of arginine or lysine from the C-terminus and sequestration of the peptide's polarity to only one of its ends leads to stronger binding on the stationary phase. Again, the N-terminal proton sequestration in LysargiNase peptides allows for the occurrence of hydrophobic stretches in peptides, a phenomenon that in tryptic peptides is obscured by the diametric positioning of protons at both peptide ends.

CONCLUSIONS

Despite the apparent similarities between LysargiNase and tryptic peptides, the opposite localization of the basic terminal residue, leads to distinct peptide fragmentation and LC separation and, consequently, also to different peptide identifications, which in turn bias protein quantitation and (phopsho)proteome coverage. The generated peptides are chemically very distinct and require different settings for optimal MS/MS identification. The two "mirror proteases" complement each other, and superior results can be achieved with their combined use. Conclusively, the use of LysargiNase may extend the boundaries of the proteomics field further, especially if these advances are supported by dedicated LysargiNase-amended search algorithms. To this end, the detailed fragmentation characteristics of LysargiNase peptides presented here will contribute to the design of improved theoretical spectra and fragmentation scoring models. As new proteases are continuously sought for in proteomics applications, development of new computational methods readily tailored toward the attributes of each protease would allow a better classification of the new enzymes and expand our proteomics toolbox.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00825.

Additional details on experimental procedures, technical reproducibility within LC−MS/MS runs, peptide retention times on the C18 column, Pearson correlation values, Venn diagrams of peptide identifications, frequency distributions of total ions in MS/MS spectra, ETD and HCD MS/MS spectra of LysargiNase phosphopeptides, and a series of illustrative ETD and HCD spectra of unmodified tryptic and LysargiNase peptides. (PDF)

Fragment ion frequencies in LysargiNase and tryptic peptides. (XLSX)

List of identified phosphopeptides. (XLSX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: A.J.R.Heck@uu.nl. Phone: +31-30-2536797.

ORCID

A.F. Maarten Altelaar: 0000-0001-5093-5945

Albert J.R. Heck: 0000-0002-2405-4404

Present Address

[‖]Proteomics and Bioanalytics Group, Technical University of Munich, 85354 Freising, Germany

Author Contributions

A.J.R.H. conceived the idea and, together, with A.F.M.A. supervised this project. L.T. and P.G. designed and performed the experiments and analyzed the data. R.A.S. contributed to the data analysis and, together with H.V.D.T., developed the computational applications for spectra analysis. C.M.O. supplied us with LysargiNase and provided input to the study. L.T., P.G., and A.J.R.H. wrote the paper. All authors have read and commented on the paper content.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. F. Xavier Gomis Ruth for the purification of LysargiNase. We thank our colleagues in the Biomolecular Mass Spectrometry and Proteomics group for fruitful discussions and technical support. L.T. was supported by EMBO with a Long-Term Fellowship (ALTF 776-2013). A.J.R.H. and A.F.M.A acknowledge support through the European Union Horizon 2020 program FET-OPEN project MSmed, project no. 686547. This work forms part of the Roadmap Initiative Proteins@Work (project no. 184.032.201)

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

financed by The Netherlands Organization for Scientific Research (NWO).

■ ABBREVIATIONS

CID, collision-induced dissociation; HCD, higher-energy collisional dissociation; ETD, electron-transfer dissociation; TIC, total ion current; FDR, false discovery Rate; NSC, normalized spectral count; PSM, peptide-to-spectrum match; ESI, electrospray ionization; RP-UPLC, reverse-phase ultra-high-pressure liquid chromatography; BSA, bovine serum albumin

■ REFERENCES

(1) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2012**, *14* (1), 35−48.
(2) Smits, A. H.; Vermeulen, M. Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol.* **2016**, *34*, 825.
(3) Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond trypsin. *FEBS J.* **2015**, *282* (14), 2612−2626.
(4) Boersema, P. J.; Kahraman, A.; Picotti, P. Proteomics beyond large-scale protein expression analysis. *Curr. Opin. Biotechnol.* **2015**, *34*, 162−170.
(5) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* **2010**, *9* (3), 1323−1329.
(6) Guo, X.; Trudgian, D. C.; Lemoff, A.; Yadavalli, S.; Mirzaei, H. Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics. *Mol. Cell. Proteomics* **2014**, *13* (6), 1573−1584.
(7) Peng, M.; Taouatas, N.; Cappadona, S.; van Breukelen, B.; Mohammed, S.; Scholten, A.; Heck, A. J. Protease bias in absolute protein quantitation. *Nat. Methods* **2012**, *9* (6), 524−525.
(8) Giansanti, P.; Aye, T. T.; van den Toorn, H.; Peng, M.; van Breukelen, B.; Heck, A. J. R. An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Rep.* **2015**, *11* (11), 1834−1843.
(9) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* **2008**, *5* (5), 405−407.
(10) Taouatas, N.; Heck, A. J. R.; Mohammed, S. Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions. *J. Proteome Res.* **2010**, *9* (8), 4282−4288.
(11) Raijmakers, R.; Neerincx, P.; Mohammed, S.; Heck, A. J. R. Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. *Chem. Commun. (Cambridge, U. K.)* **2009**, *46* (46), 8827−8829.
(12) Wu, C.; Tran, J. C.; Zamdborg, L.; Durbin, K. R.; Li, M.; Ahlf, D. R.; Early, B. P.; Thomas, P. M.; Sweedler, J. V.; Kelleher, N. L. A protease for "middle-down" proteomics. *Nat. Methods* **2012**, *9* (8), 822−824.
(13) Meyer, J. G.; Kim, S.; Maltby, D. A.; Ghassemian, M.; Bandeira, N.; Komives, E. A. Expanding proteome coverage with orthogonal-specificity α-lytic proteases. *Mol. Cell. Proteomics* **2014**, *13* (3), 823−835.
(14) Laskay, U. A.; Srzentić, K.; Monod, M.; Tsybin, Y. O. Extended bottom-up proteomics with secreted aspartic protease Sap9. *J. Proteomics* **2014**, *110*, 20−31.
(15) Tallant, C.; García-Castellanos, R.; Seco, J.; Baumann, U.; Gomis-Rüth, F. X. Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *J. Biol. Chem.* **2006**, *281* (26), 17920−17928.
(16) Huesgen, P. F.; Lange, P. F.; Rogers, L. D.; Solis, N.; Eckhard, U.; Kleifeld, O.; Goulas, T.; Gomis-Rüth, F. X.; Overall, C. M. LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification. *Nat. Methods* **2014**, *12* (1), 55−58.
(17) Mitchell Wells, J.; McLuckey, S. A. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol.* **2005**, *402*, 148−185.
(18) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35* (12), 1399−1406.
(19) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.
(20) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database independent proteomics analysis of the ostrich and human proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407−412.
(21) Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **2014**, *11* (3), 319−324.
(22) Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J. R.; Zou, H.; Mohammed, S. Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **2013**, *8* (3), 461−480.
(23) Cristobal, A.; Hennrich, M. L.; Giansanti, P.; Goerdayal, S. S.; Heck, A. J. R.; Mohammed, S. In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst* **2012**, *137* (15), 3541−3548.
(24) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976−989.
(25) Käll, L.; Storey, J. D.; Noble, W. S. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics* **2008**, *24* (16), i42−8.
(26) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **2011**, *10* (12), 5354−5362.
(27) Vizcaíno, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **2016**, *44* (D1), D447−56.
(28) Ihaka, R.; Gentleman, R. R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* **1996**, *5* (3), 299−314.
(29) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **2011**, *10* (5), 2377−2388.
(30) Zubarev, R.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A···. *J. Am. Chem. Soc.* **1998**, *120* (16), 3265−3266.
(31) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528−9533.
(32) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1811−1822.
(33) Hunter, E. P. L.; Lias, S. G. Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *J. Phys. Chem. Ref. Data* **1998**, *27* (3), 413.
(34) Mant, C. T.; Burke, T. W.; Black, J. A.; Hodges, R. S. Effect of peptide chain length on peptide retention behaviour in reversed-phase chromatography. *J. Chromatogr.* **1988**, *458*, 193−205.
(35) Tripet, B.; Cepeniene, D.; Kovacs, J. M.; Mant, C. T.; Krokhin, O. V.; Hodges, R. S. Requirements for prediction of peptide retention time in reversed-phase high-performance liquid chromatography: hydrophilicity/hydrophobicity of side-chains at the N- and C-termini

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861

of peptides are dramatically affected by the end-groups and location. *J. Chromatogr. A* **2007**, *1141* (2), 212−225.
(36) Zhou, N. E.; Mant, C. T.; Hodges, R. S. Effect of preferred binding domains on peptide retention behavior in reversed-phase chromatography: amphipathic alpha-helices. *Pept. Res.* **1990**3 (1), 8−20.
(37) Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villén, J.; Haas, W.; Sowa, M. E.; Gygi, S. P. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* **2010**, *143* (7), 1174−1189.
(38) Malmström, J.; Beck, M.; Schmidt, A.; Lange, V.; Deutsch, E. W.; Aebersold, R. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **2009**, *460* (7256), 762−765.
(39) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473* (7347), 337−342.
(40) Riley, N. M.; Coon, J. J. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal. Chem.* **2016**, *88* (1), 74−94.
(41) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24* (10), 1285−1292.
(42) Marx, H.; Lemeer, S.; Schliep, J. E.; Matheron, L.; Mohammed, S.; Cox, J.; Mann, M.; Heck, A. J. R.; Kuster, B. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat. Biotechnol.* **2013**, *31* (6), 557−564.
(43) Murphree, A. L.; Benedict, W. F. Retinoblastoma: clues to human oncogenesis. *Science* **1984**, *223* (4640), 1028−1033.
(44) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* **2016**, *11* (5), 993−1006.
(45) Waldera-Lupa, D. M.; Stefanski, A.; Meyer, H. E.; Stühler, K. The fate of b-ions in the two worlds of collision-induced dissociation. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834* (12), 2843−2848.

DOI: 10.1021/acs.jproteome.6b00825
J. Proteome Res. 2017, 16, 852−861