

Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin

Therese Dau, Giulia Bartolomucci, and Juri Rappsilber*



Cite This: *Anal. Chem.* 2020, 92, 9523–9527



Read Online

ACCESS |



Metrics & More

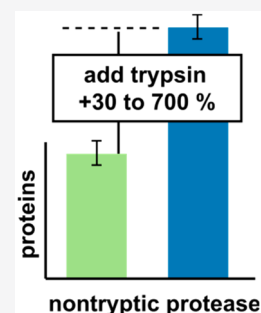


Article Recommendations



Supporting Information

ABSTRACT: Trypsin is the most used enzyme in proteomics. Nevertheless, proteases with complementary cleavage specificity have been applied in special circumstances. In this work, we analyzed the characteristics of five protease alternatives to trypsin for protein identification and sequence coverage when applied to *S. pombe* whole cell lysates. The specificity of the protease heavily impacted the number of proteins identified. Proteases with higher specificity led to the identification of more proteins than proteases with lower specificity. However, AspN, GluC, chymotrypsin, and proteinase K largely benefited from being paired with trypsin in sequential digestion, as had been shown by us for elastase before. In the most extreme case, predigesting with trypsin improves the number of identified proteins for proteinase K by 731%. Trypsin predigestion also improved the protein identifications of other proteases, AspN (+62%), GluC (+80%), and chymotrypsin (+21%). Interestingly, the sequential digest with trypsin and AspN yielded even a higher number of protein identifications than digesting with trypsin alone.



Trypsin is the protease of choice for mass spectrometry (MS)-based proteomics. It cleaves carboxyterminal of Arg and Lys residues, resulting in a positive charge at the peptide C-terminus, which is advantageous for MS analysis.^{1,2} Nevertheless, other proteases are frequently used to obtain complementary data.^{3,4}

Among these, AspN and GluC target acidic amino acid residues (Figure 1a). Both enzymes generate peptide mixtures of comparable complexity to that of trypsin and have been successfully used in many studies.^{4–7} Chymotrypsin, which targets primarily aromatic residues, has also been used.^{7–9} In contrast, broad specificity proteases are much less widely used in proteomics. This is likely due to the high complexity of the peptide mixtures that they generate. To our knowledge, their application has been limited to prefractionated samples. Proteinase K, for example, was used to “shave” surface-exposed loops from proteins in membrane vesicles.^{10,11}

Our group has previously shown that the number of identified peptides, when using alternatives to trypsin, could largely be improved by a sequential combination with trypsin. This includes AspN, GluC, chymotrypsin, and elastase for the detection of cross-link sites^{12–15} and elastase applied to *S. pombe* whole cell lysates.¹⁴ The sequential digestion increased the number of identified cross-links up to 19-fold for the Taf4–12 complex compared to digesting with elastase alone.¹⁴ Introducing positively charged C-termini through trypsin improves the detection of previously nontryptic peptides. Importantly, smaller peptides are protected from the second protease.^{12,14} Thus, use of two proteases does not lead to the very small peptides that *in silico* digestion would predict. As a consequence, using elastase after trypsin does not lead to the same peptide complexity as using elastase alone.

In this study, we analyzed whether the introduction of trypsin in a sequential digest might improve the application of AspN, GluC, chymotrypsin, and proteinase K on unfractionated *S. pombe* lysate.

METHODS

Public Data Sets. The data on trypsin, elastase, trypsin–elastase, and elastase–trypsin were taken from our previous work¹⁴ and retrieved from PRIDE with the data set identifier PXD011459.

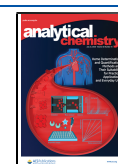
Sample Preparation. One gram of frozen and ground *S. pombe* cells were resuspended in 2 mL of RIPA (Sigma-Aldrich, St. Louis, MO) supplemented with the protease inhibitor cocktail cOMplete according to the manufacturer’s instructions (Roche, Basel). To remove the cell debris, the samples were centrifugated at 1200g for 15 min. The lysates were subjected to gel electrophoresis on a 4%–12% Bis-Tris gel (Life Technologies, Carlsbad, CA) for 5 min and stained using Imperial Protein Stain (Thermo Fisher Scientific, Rockford, IL). After excising the stained gel area as a single fraction, the proteins were first reduced with dithiothreitol and then alkylated with iodoacetamide.

The first protease (trypsin (1:100), elastase (1:100), AspN (1:100), GluC (1:50), chymotrypsin (1:50), and proteinase K (1:50)) was incubated for 16 h at 37 °C (besides chymotrypsin

Received: February 3, 2020

Accepted: June 16, 2020

Published: July 6, 2020



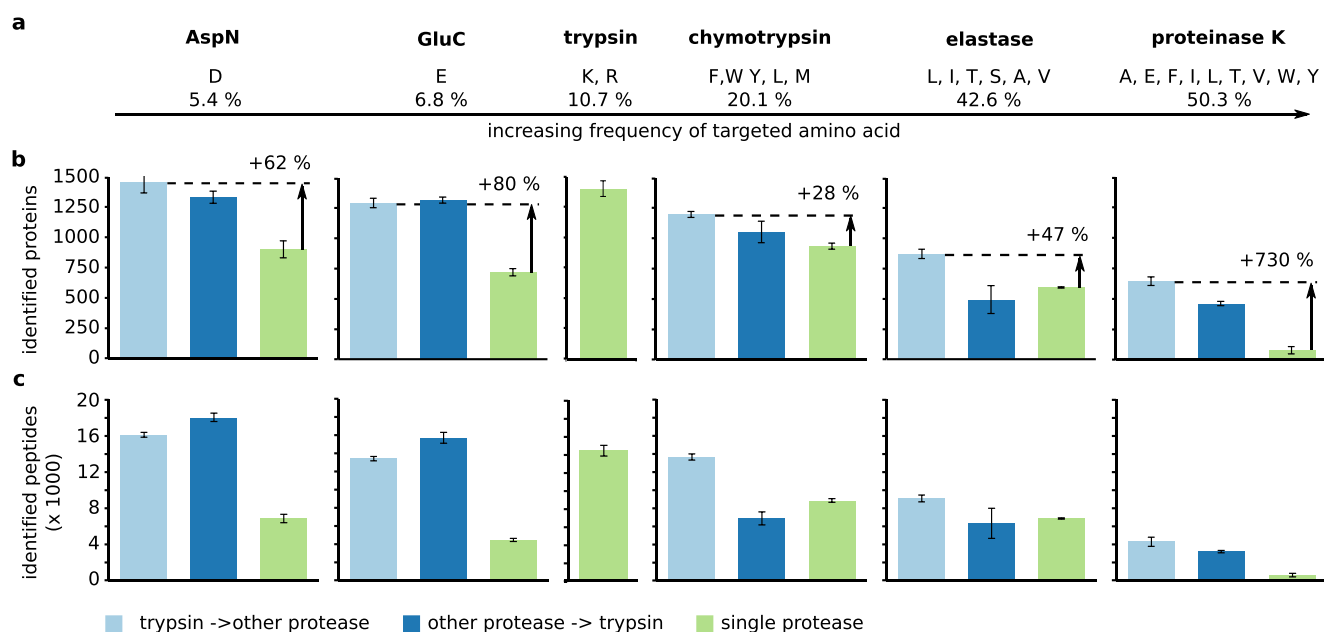


Figure 1. Impact of different proteases and protease combinations on the identification of proteins and peptides. (a) Frequency of the amino acids targeted by AspN, GluC, chymotrypsin, elastase, and proteinase K according to the UniProtKB/TrEMBL release report. Number of (b) proteins and (c) peptides identified with different protease combinations. Trypsin–other protease, light blue; other protease–trypsin, dark blue; single protease, green. Error bars are standard deviation (SD) of at least five independent digestion experiments.

at RT). The second protease was added for 4 h at 37 °C (besides elastase for 30 min).

We used a standardized protocol to desalt and concentrate the peptides on C18 StageTips for subsequent analysis.^{16,17} For each condition, the equivalent of 1 μ g protein starting material was used.

LC-MS/MS. All samples were analyzed on a linear iontrap–orbitrap mass spectrometer (Orbitrap Elite, Thermo Fisher Scientific, Rockford, IL) coupled online to a liquid chromatograph (Ultimate 3000 RSLCnano Systems, Dionex, Thermo Fisher Scientific, UK) with a C18-column (EASY-Spray LC Column, Thermo Fisher Scientific, Rockford, IL). The flow rate was 0.2 μ L/min using 98% mobile phase A (0.1% formic acid) and 2% mobile phase B (80% acetonitrile in 0.1% formic acid). To elute the peptides, the percentage of mobile phase B was first increased to 40% over a time course of 110 min followed by a linear increase to 95% in 11 min. Full MS scans were recorded in the orbitrap at a 120,000 resolution for MS1 with a scan range of 300–1700 m/z . The 20 most intense ions (precursor charge ≥ 2) were selected for fragmentation by collision-induced dissociation, and MS2 spectra were recorded in the ion trap (20,000 ions as a minimal required signal, 35 normalized collision energy, dynamic exclusion for 40 s).

Data Analysis. MaxQuant software¹⁸ (version 1.5.2.8) employing the Andromeda search engine¹⁹ in combination with the PombeBase database²⁰ was used to analyze the samples. The following parameters were used for the search: carbamidomethylation of cysteine as a fixed modification, oxidation of methionine as a variable modification, MS accuracy of 4.5 ppm, and MS/MS tolerance of 0.5 Da. Up to six miscleavages were allowed for digests involving trypsin, AspN, GluC, or chymotrypsin and up to 10 miscleavages for digests containing elastase or proteinase K. Frequencies of amino acids were taken from the statistics of the UniProtKB/

TrEMBL protein database release 2019_11 (<https://www.ebi.ac.uk/uniprot/TrEMBLstats>).

RESULTS AND DISCUSSION

Lysate from *S. pombe* was digested either with trypsin, AspN, GluC, chymotrypsin, elastase, or proteinase K. We also combined each of the proteases other than trypsin in a sequential digest with trypsin as either the first or second protease.

Adding trypsin to the digest with AspN and GluC improved the protein (AspN = 899 \pm 69, trypsin–AspN = 1455 \pm 85, AspN–trypsin = 1331 \pm 50, GluC = 719 \pm 28, trypsin–GluC = 1294 \pm 37, GluC–trypsin = 1319 \pm 25) identification (Figure 1b). Peptide identifications also improved (AspN = 6828 \pm 514, trypsin–AspN = 16087 \pm 327, AspN–trypsin = 17968 \pm 470, GluC = 4467 \pm 182, trypsin–GluC = 13461 \pm 260, GluC–trypsin = 15713 \pm 600) (Figure 1c). The order of proteases had only a minor influence on the identifications.

Using trypsin prior to chymotrypsin or elastase also improved the identification of proteins (chymotrypsin = 938 \pm 27, trypsin–chymotrypsin = 1200 \pm 25, elastase = 593 \pm 7, trypsin–elastase = 874 \pm 40), and peptides (chymotrypsin = 8818 \pm 232, trypsin–chymotrypsin = 13611 \pm 346, elastase = 6821 \pm 84, trypsin–elastase = 9039 \pm 374). Using trypsin as the second protease had only a minimal effect on the protein (chymotrypsin–trypsin = 1056 \pm 91, elastase–trypsin = 492 \pm 115) and peptide identification (chymotrypsin–trypsin = 6869 \pm 744, elastase–trypsin = 6280 \pm 1680).

Interestingly, digesting with trypsin alone did not give the highest number of protein (1403 \pm 65) and peptide (14410 \pm 571) identifications. We identified more proteins (+4%) and peptides (+12%) when trypsin was followed by AspN.

The biggest impact of sequential digestion with trypsin was seen on the performance of proteinase K. Using proteinase K alone led to very few identifications of proteins (proteinase K = 78 \pm 33) and peptides (proteinase K = 527 \pm 179). This

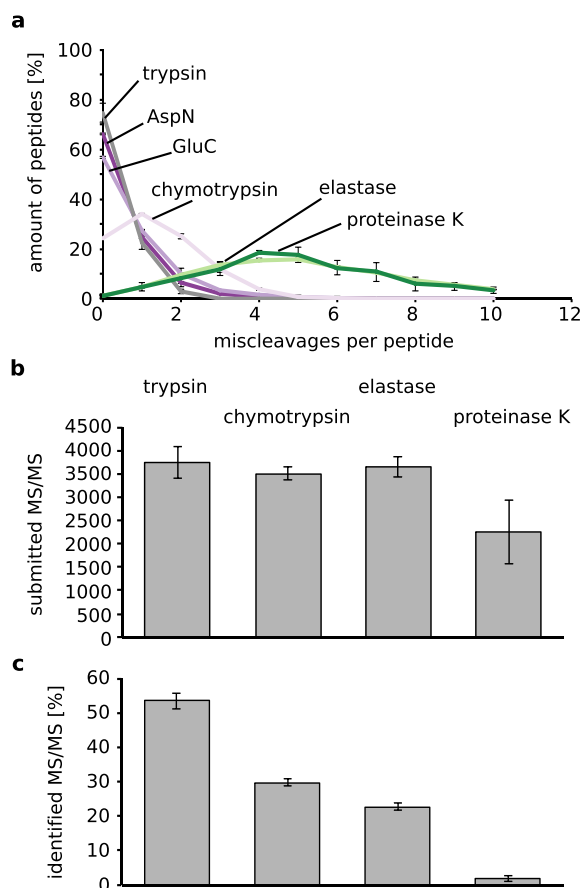


Figure 2. (a) Numbers of miscleavages for each protease AspN, dark violet; GluC, light violet; chymotrypsin, very light violet; elastase, light green; proteinase K dark green. (b) Of the promiscuous proteases, only proteinase K showed a reduced number of submitted MS/MS compared to trypsin, chymotrypsin, and elastase. (c) Identification rate of searched MS/MS decreased with decreasing specificity of the protease.

might be due to very short peptides being generated by proteinase K, which cleaves carboxyterminal of half of all the amino acids. Alternatively, or in addition, the high complexity of the peptide mixture generated by proteinase K might reduce identification rates. Surprisingly, adding trypsin to the proteinase K digest increased the number of identifications for proteins (proteinase K–trypsin = 461 ± 17) and peptides (proteinase K–trypsin = 3169 ± 194). Using trypsin prior to proteinase K further improved on these results as this led to the identification of 8 times more proteins (646 ± 36) and 8

times more peptides (4279 ± 530) compared to proteinase K alone.

In summary, AspN, GluC, and proteinase K profited most of the five tested proteases from the addition of trypsin. The underlying reasons for the observed gains are likely different. AspN and GluC have low amounts of available cleavage sites and therefore generate relatively long peptides. Many of these will be unfavorably long for mass spectrometric detection. In addition, they are missing a terminal positive charge. Adding trypsin introduces such a C-terminal charge and shortens very long peptides, both enhancing peptide detection in MS analysis.

AspN and GluC are highly efficient (Figure 2a), while for chymotrypsin and especially elastase and proteinase K many miscleavages were detected. Although we cannot exclude that undigested protein from the first digest may be the source for the additional identification of peptides and proteins, the high efficiency of GluC and AspN makes it unlikely to be the case for these enzymes. Also, the LC-MS data did not indicate the presence of a large quantity of semidigested proteins, as judged from the absence of a late eluting and highly charged cluster of ions (data not shown).

To analyze possible reasons for the low identification rates of more promiscuous cutters, we looked at the submitted and identified MS/MS spectra (Figure 2b, c). Only proteinase K had a reduced number of submitted MS/MS spectra. This might be due to the complexity of the peptide mixture resulting from proteinase K. However, the main problem was the low identification success of these MS/MS spectra. The same applied to the spectra from other less specific proteases. One of the reasons might be cofragmentation of several peptides as the mixture is more complex than for specific proteases. This is supported by the fact that AspN and GluC showed similar identification rates to trypsin. Another reason might be the increase in the database size and the problems associated with it for identification.

While AspN and GluC are very specific proteases, over 50% of the residues are potential cleavage sites for proteinase K. The problem for proteinase K is therefore not a lack of cleavage sites. Adding trypsin to proteinase K increased identifications and thus ruled out the possibility that peptides generated by proteinase K alone, at least under standard conditions, are generally too short for proteomics. If therefore complexity of a proteinase K digest is the reason for the low identification yields of proteinase K; then, the addition of trypsin must reduce this complexity. Adding trypsin might unify “ragged” proteinase K peptides that share either the N- or C-terminus but have different lengths (Figure 3a). In this way, trypsin leads to a concentration increase of peptides by

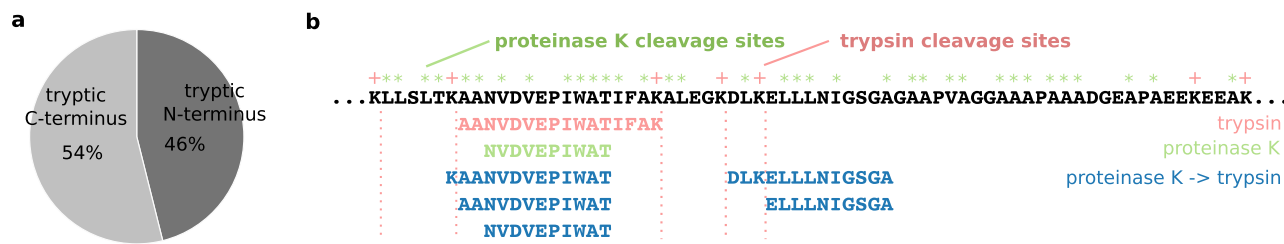


Figure 3. (a) Comparison of semitryptic peptides with a tryptic N- or C-terminus after digesting whole *S. pombe* with proteinase K followed by trypsin. (b) Peptides that have been identified in the N-terminal region of 60S acidic ribosomal protein P1-alpha 1 (26–94) with either trypsin, proteinase K, or proteinase K followed by trypsin. Trypsin, red; proteinase K, green; proteinase K–trypsin, dark blue.

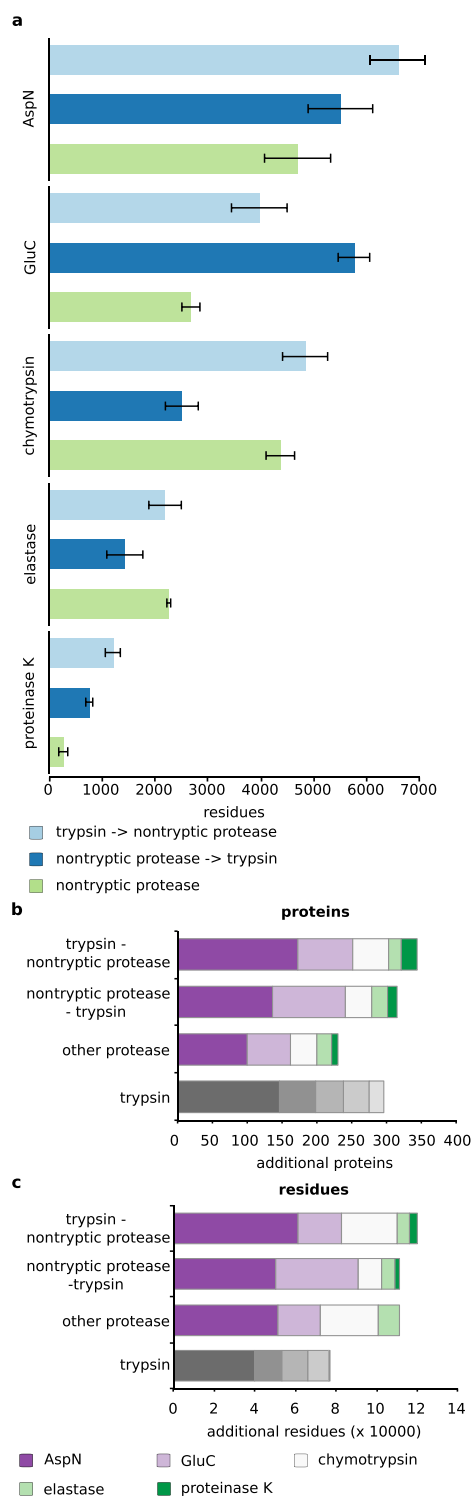


Figure 4. Comparison of residues gained with the three protease combinations for (a) AspN, GluC, chymotrypsin, elastase, and proteinase K on top of trypsin digest. Trypsin–other protease, light blue; other protease–trypsin, dark blue; single protease, green. Comparison of (b) proteins and (c) residues gained on top of a tryptic digest through sequential digestion variants, parallel digestion, and replica of tryptic digestion. AspN, dark violet; GluC, light violet; chymotrypsin, white; elastase, light green; proteinase K, dark green.

reducing sample complexity. At least when trypsin is used first, an additional mechanism must be considered that was previously described for sequential digestion.^{12,14} The second

enzyme does not cleave shorter peptides with high efficiency, effectively leading to short tryptic peptides being protected from proteinase K. In either case, the complexity that is normally introduced through proteinase K is reduced by the tryptic treatment.

End trimming and short peptide protection alone are likely not the sole explanations. We observed previously that among all observed mixed-protease action peptides, i.e., those peptides that were generated by trypsin action on one end and another protease at the other end, there is a misbalance: tryptic C-termini are more prevalent than N-termini generated by trypsin (semityptic peptides with tryptic N-terminus = 652 ± 42 , semityptic peptides with tryptic C-terminus = 763 ± 15) (Figure 3a). This means also the improved observability of peptides with a basic C-terminal residue contributes to the observed effect of sequential digestion on identification rates.

As an example, we analyzed the 60S acidic ribosomal protein P1-alpha 1 (Figure 3b). There are no trypsin cleavages sites between residues 56 and 90, so this region is not covered when trypsin is used alone. Digesting with proteinase K alone did not improve the coverage for this region, although or possibly because every other residue is a potential cleavage site for proteinase K. Peptides from this region could only be identified when proteinase K and trypsin were used in a sequential digest.

We then wondered how far the proteins and peptides that were observed in the different uses of proteases alone or in combination with trypsin covered different sequence space. We measured this in number of unique residues. As one would expect, this followed the same trends seen for protein and peptide identifications. For AspN and GluC, the largest number of residues was covered when trypsin was used following the other protease (Figure S-1a, b). For chymotrypsin, elastase, and proteinase K, the inverse order, i.e., trypsin first, yielded the larger coverage (Figure S-1c–e). Nonetheless, the different conditions yielded substantial nonoverlap. When combining the results of two digestion conditions, one would combine the data obtained by the protease alone with that of a trypsin-first sequential digest. Their overlap is substantially smaller ($4 \pm 2\%$ to $38 \pm 1\%$) than what we observed here for trypsin replicas ($83 \pm 2\%$).

Next, we compared the gain of residues on top of the trypsin digest that was observed for each digestion protocol (Figure 4a). For AspN, GluC, and proteinase K, there was a significant increase of additional identified residues if trypsin was added prior to the digest. Curiously, the highest gain in residues for AspN was achieved with a sequential AspN–trypsin digest. For elastase and chymotrypsin, adding trypsin prior to their usage did not increase the number of identified significantly. Reversing the order in the sequential digest even decreased the gain in residues.

Finally, we analyzed the gain in identified proteins and residues when using different combinations of digestion conditions (Figure 4b, c). We combined the results of either five replicas of trypsin, trypsin followed by either of the five other proteases, or either of the five other proteases followed by trypsin. An initial trypsin digest served as the reference, in which 1484 proteins and 202,556 residues were identified. This followed the rationale that one would always use trypsin for an initial analysis, although trypsin followed by AspN in a sequential digest consistently gave here higher protein and peptide identifications. The highest numbers of complementary proteins (344) and residues (119,763) were identified

when trypsin was used first in a sequential digest in combination with either of the five other proteases, followed by the inverted setup in which trypsin was used last (proteins = 315, residues = 111,126). Digestions with nontryptic proteases alone were outperformed by trypsin replicas in terms of protein identification (230 versus 296) but not in terms of residue coverage (111,069 versus 76,645).

CONCLUSION

In this study, we investigated the impact of adding trypsin to other proteases in proteomics. Sequential digestion has been used before,^{5,6,21} and we here add a systematic evaluation of different protease combinations. Protein and peptide identifications improved when combining any of the tested proteases with trypsin. This is in line with previous studies on cross-linking identification, which benefited from the sequential digest with trypsin.^{12,14} In the most extreme case, the sequential digest with trypsin and AspN outperformed results obtained by trypsin alone. This effect is relatively small, and due to cost considerations, trypsin will remain the protease of first choice in proteomics also after our study. However, situations where alternative proteases are currently used could in the future benefit from adding a sequential digestion step with trypsin. As trypsin is compatible with the buffer conditions of the tested proteases, this requires no other additional step than adding trypsin.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c00478>.

Additional analysis that compares the covered residues between the three digestion protocols for AspN, GluC, chymotrypsin, elastase, and proteinase K (PDF)

AUTHOR INFORMATION

Corresponding Author

Juri Rappsilber – Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany; Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, U.K.; orcid.org/0000-0001-5999-1310; Email: juri.rappsilber@tu-berlin.de

Authors

Therese Dau – Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany; Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, U.K.; orcid.org/0000-0003-0251-9490

Giulia Bartolomucci – Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, U.K.

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.0c00478>

Author Contributions

T.D. and J.R. conceived this study and interpreted data. T.D. and G.B. conducted all experiments. T.D. and J.R. wrote the manuscript with input from G.B. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by a research stipend to T.D. (DA 1861/2-1) from the Deutsche Forschungsgemeinschaft and by the Wellcome Trust through a Senior Research Fellowship to J.R. (103139) and a multiuser equipment grant (108504). The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium²² via the PRIDE partner repository with the data set identifier PXD011459¹⁴ and PXD017321.

REFERENCES

- (1) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77* (18), 5800–5813.
- (2) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118* (35), 8365–8374.
- (3) Swaney, D. L.; Wenger, C. D.; Coon, J. J. *J. Proteome Res.* **2010**, *9*, 1323–1329.
- (4) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. *Nat. Protoc.* **2016**, *11*, 993–1006.
- (5) Wiśniewski, J. R.; Mann, M. *Anal. Chem.* **2012**, *84*, 2631–2637.
- (6) Guo, X.; Trudgian, D. C.; Lemoff, A.; Yadavalli, S.; Mirzaei, H. *Mol. Cell. Proteomics* **2014**, *13*, 1573–1584.
- (7) Giansanti, P.; Aye, T. T.; van den Toorn, H.; Peng, M.; van Breukelen, B.; Heck, A. J. R. *Cell Rep.* **2015**, *11* (11), 1834–1843.
- (8) Wang, X.; Codreanu, S. G.; Wen, B.; Li, K.; Chambers, M. C.; Liebler, D. C.; Zhang, B. *Mol. Cell. Proteomics* **2018**, *17* (3), 422–430.
- (9) Fischer, F.; Poetsch, A. *Proteome Sci.* **2006**, *4*, 1–12.
- (10) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd. *Nat. Biotechnol.* **2003**, *21*, 532–538.
- (11) Fischer, F.; Wolters, D.; Rögner, M.; Poetsch, A. *Mol. Cell. Proteomics* **2006**, *5*, 444.
- (12) Mendes, M. L.; Fischer, L.; Chen, Z. A.; Barbon, M.; O'Reilly, F. J.; Giese, S. H.; Bohlke-Schneider, M.; Belsom, A.; Dau, T.; Combe, C. W.; et al. *Mol. Syst. Biol.* **2019**, *15* (9), 1–13.
- (13) Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Förster, F.; Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11* (3), M111.014126.
- (14) Dau, T.; Gupta, K.; Berger, I.; Rappsilber, J. *Anal. Chem.* **2019**, *91* (7), 4472–4478.
- (15) Stieger, C. E.; Doppler, P.; Mechtler, K. *J. Proteome Res.* **2019**, *18* (3), 1363–1370.
- (16) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2*, 1896–1906.
- (17) Rappsilber, J.; Ishihama, Y.; Mann, M. *Anal. Chem.* **2003**, *75* (3), 663–670.
- (18) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.
- (19) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.
- (20) McDowall, M. D.; Harris, M. A.; Lock, A.; Rutherford, K.; Staines, D. M.; Bähler, J.; Kersey, P. J.; Oliver, S. G.; Wood, V. *Nucleic Acids Res.* **2015**, *43* (Database issue), D656–61.
- (21) Larsen, M. R.; Højrup, P.; Roepstorff, P. *Mol. Cell. Proteomics* **2005**, *4*, 107–119.
- (22) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. *Nat. Biotechnol.* **2014**, *32* (3), 223–226.