Contents lists available at ScienceDirect

Data in Brief

Data Article

TAILS N-terminomic and proteomic datasets of healthy human dental pulp

Ulrich Eckhard [a,b], Giada Marino [a,b], Simon R. Abbey [a,b], Ian Matthew [b], Christopher M. Overall [a,b,c,*]

[a] Centre for Blood Research, The Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada
[b] Department of Oral Biological and Medical Sciences, Faculty of Dentistry, University of British Columbia, Vancouver, BC, Canada
[c] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

ARTICLE INFO

ABSTRACT

The Data described here provide the in depth proteomic assessment of the human dental pulp proteome and N-terminome (Eckhard et al., 2015) [1]. A total of 9 human dental pulps were processed and analyzed by the positional proteomics technique TAILS (Terminal Amine Isotopic Labeling of Substrates) N-terminomics. 38 liquid chromatography tandem mass spectrometry (LC-MS/MS) datasets were collected and analyzed using four database search engines in combination with statistical downstream evaluation, to yield the by far largest proteomic and N-terminomic dataset of any dental tissue to date. The raw mass spectrometry data and the corresponding metadata have been deposited in ProteomeXchange with the PXD identifier <**PXD002264**>; Supplementary Tables described in this article are available via Mendeley Data (http://dx.doi.org/10.17632/555j3kk4sw.1).

© 2015 Published by Elsevier Inc.

* Corresponding author at: Centre for Blood Research, The Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada.
E-mail address: chris.overall@ubc.ca (C.M. Overall).

Specifications Table

Subject area	Biology
More specific subject area	Dental biology, proteolytic processing, protein N-termini, proteomics, N-terminomics, Human Proteome Project (HPP)
Type of data	Mass spectrometry raw-files; search engine output files; processed metadata (.xlsx) reporting identified peptide spectrum matches (1% FDR) and proteins (protein probability ≥ 0.95)
How data was acquired	Liquid chromatography tandem mass spectrometry (LC-MS/MS): Accurate Mass G6550A Quadrupole-time-of-flight (Q-TOF) mass spectrometer coupled on-line to a 1200 Series nanoflow HPLC with a Chip Cube interface (Agilent).
Data format	RAW files:.d-folders and.mzXML-files; peak lists:.mgf; post database search output files from Mascot [2], X!Tandem [3], Comet [4], and MS-GF+ [5]: .pep.xml; metadata: .xlsx-files
Experimental factors	Healthy dental pulps were collected within 10 min of routine extraction of lower third molars (wisdom teeth; teeth 38 and 48); written informed consent was obtained from the patients before surgery. Teeth were partially sectioned and mechanically split, exposing the dental pulp. The pulp was immediately transferred into 250 µl 8 M guanidine hydrochloride, frozen on dry ice, and stored at $-80\,^\circ$C for a maximum period of 30 days. Specimens were separately homogenized on ice, proteins extracted, and samples cleaned-up using chloroform/methanol precipitation. Pellets were redissolved in 0.5 mL of 8 M guanidine hydrochloride and protein concentrations were determined using 1:10 dilutions (in ddH2O) and Bradford protein assay (Bio-Rad) with bovine serum albumin (BSA) for calibration.
Experimental features	TAILS N-terminomics [6–10] was performed on 1 mg of human dental pulp proteome extracts to identify both proteins (preTAILS; pulp proteome) and corresponding protein N-termini (TAILS; pulp N-terminome). Primary amines of protein N-termini and lysine sidechains were blocked and labeled by dimethylation before trypsinization at the intact protein level. Labeled proteomes were then digested using proteomics-grade trypsin overnight, and trypsin-generated internal and C-terminal peptides with new free N-terminal primary amines were covalently bound to an aldehyde-derivatized, dendritic polyglycerol polymer (HPG-ALD; http://flintbox.com/public/project/1948/). Unbound peptides representing original and processed protein N-termini (either naturally blocked by *e.g.* Nα-acetylation or experimentally labeled by dimethylation) were separated from polymer-bound internal and C-terminal peptides using ultra filtration and analyzed by LC-MS/MS (TAILS sample). To increase proteome coverage, approx. 50 µg of each pulp sample were analyzed immediately after tryptic digest, omitting the polymer-mediated N-terminal enrichment step (preTAILS).
Data source location	Overall Laboratory, Centre for Blood Research, Department of Oral Biological and Medical Sciences, Faculty of Dentistry, University of British Columbia, Vancouver, BC, Canada. 49°15'44.5"N 123°14'41.8"W.
Data accessibility	The mass spectrometry raw data and metadata have been deposited to ProteomeXchange with the PXD identifier < PXD002264 >. Supplementary tables described in this article are available via Mendeley Data, http://dx.doi.org/10.17632/555j3kk4sw/1

Value of the data

- Largest compendium human dental pulp proteome datasets. Generated by liquid chromatography tandem mass spectrometry (LC-MS/MS), reporting > 4000 proteins and > 9000 protein N-termini.
- Comprehensive assessment of global *in vivo* proteolytic processing by TAILS N-terminomics in a healthy human tissue.
- Proteolytic processing represents an eminent post-translational modification which frequently alters protein function and localization. Thus large scale positional proteomics datasets such as this one are required to decipher pervasive proteolytic networks in health and disease.

1. Data

The data in the ProteomeXchange archive (PXD002264) provide a novel description of the human pulp proteome and N-terminome as seen by TAILS N-terminomics [1]. These data were analyzed using four different database search engines, namely Mascot [2], X! Tandem [3], Comet [4], and MS-GF+ [5] in combination with PeptideProphet [11], iProphet [12], and ProteinProphet [13], all compiled within the Trans Proteomic Pipeline (v4.8.0, PHILAE) [14]. A variety of other free or commercial database search engines such as MyriMatch [15], MS Amanda [16], or OMSSA [17], and data analysis pipelines (*e.g.* MaxQuant [18], Peptide Shaker [19], or Scaffold[20]) could be used for additional analysis, *e.g.* to explore additional N-terminal modifications [21,22].

We provide a comprehensive set of > 380000 Peptide Spectrum Matches (PSMs), corresponding to > 21000 unique peptides which mapped to > 4000 proteins. More than 9000 protein N-termini were identified, including 5292 neo-N-termini, indicative of pervasive proteolytic processing even in healthy human tissues [1,8,23]. Furthermore, 316 and 762 protein N-termini were identified with-Met1 intact or removed (N-terminal Methionine Excision), and 125, 21, and 58 representing natural N-termini after signal-, transit-, and pro-peptide removal, respectively. Together with 24 UniProt curated internal processing sites (representing *e.g.* the proteolytic release of cryptic peptides or known proteolytic activation events as in the case of complement component C3 and the release of C3a anaphylatoxin), a total of 1306 UniProt [24] and TopFIND [25] curated protein N-termini were identified.

For an up-to-date TAILS N-terminomics protocol, please refer to our webpage (www.clip.ubc.ca; CLIP stands for Canadian Laboratories for Innovations in Proteomics), which hosts many highly valuable, proteolytic processing centered, proteomic resources, such as (i) the knowledgebase and analysis resource for protein termini and protease processing, TopFIND [26,25,27], (ii) WebPICS [28] for the streamlined analysis of active site specificity profiling experiments by PICS (Proteomic identification of protease Cleavage Sites) [29–33], (iii) resources such as CLIPPER or TAILS annotator aimed to facilitate TAILS analysis [34], and information on (iv) the protease and inhibitor centric CLIP-CHIP™ microarray [35], and (v) the new proteomic tool LysargiNase, a prote**ase** cleaving **N**-terminal to **lys**ine and **arg**inine residues and thus perfectly complementing trypsin in any proteomic approach [36], and making protein C-terminal peptides more amenable to mass spectrometry due to the LysargiNase-generated N-terminal basic residue [36,37].

The following materials and methods section will enable other investigators and laboratories to design similar experimental procedures to study human dental pulp or any other human or animal tissue by TAILS N-terminomics, or a comparable proteomic technique. Importantly, we used only trypsin for pulp proteome digestion. However, by using additional digestion approaches (*e.g.* by using GluC, LysargiNase, chymotrypsin, or multiple enzymes) [37–39], an even more comprehensive picture of the dental pulp proteome and N-terminome may be feasible.

2. Materials and methods

2.1. Dental pulp collection and proteome extraction

Nine healthy dental pulps were collected within 5–10 min of prophylactic extraction of healthy teeth numbered 38 and 48. A written informed consent was obtained from the patients before surgery in accordance with a protocol approved by the University of British Columbia Clinical Research Ethics Board (UBC CREB). After extensive washing, extracted third molars (wisdom teeth) were sectioned without traumatizing the pulp tissue, and immediately transferred into 250 μl 8 M guanidine hydrochloride, snap frozen on dry ice, and stored at −80 °C. Specimens were separately homogenized on ice using a tissue homogenizer (Ultra-Turrax; IKA Works, Inc.), and proteins were precipitated by chloroform/methanol [40]. Pellets were redissolved in 0.5 ml of 8 M guanidine hydrochloride and protein concentrations were determined by Bradford protein assay (Bio-Rad) using 1:10 dilutions of the samples (in ddH2O).

2.2. TAILS N-terminomics

Protein N-termini were enriched from human dental pulp proteomes by TAILS N-terminomics as described in great detail previously [6–9]. 1.0 mg of non-fractionated sample was diluted in 4 M guanidine hydrochloride, reduced with 5 mM dithiothreitol (DTT; 30 min, 65 °C) and cysteines were carbamidomethylated using 10 mM iodoacetamide (45 min, room temperature in the dark). After quenching with 10 mM DTT (30 min, room temperature), pH was adjusted to 6.5 for reductive dimethylation of primary amines (*i.e.* α-amines of protein N-termini and ε-amines of lysine sidechains) with 40 mM isotopically heavy formaldehyde ($^{13}CD2$ in D_2O; Cambridge Isotopes) and 20 mM sodium cyanoborohydride (overnight, 37 °C). To ensure completion of amine blocking, 20 mM heavy formaldehyde and 20 mM cyanoborohydride were added (2 h, 37 °C) after overnight incubation. After quenching with 100 mM Tris–HCl, pH 6.8 (30 min, 37 °C), samples were precipitated with chloroform/methanol for reaction clean-up [40]. Protein pellets were resolubilized in a small volume (25–50 μl of 50 mM NaOH, pH–neutralized using 100 mM HEPES (pH 7.5) to 250 μl, diluted 1:1 with HPLC-grade water, and digested with mass spectrometry-grade trypsin (Trypsin Gold, Promega) at a proteome:enzyme ratio of 100:1 (w/w) overnight (37 °C). Digestion efficiency was confirmed by SDS-PAGE; in case of incomplete digestion more trypsin was added (2 h, 37 °C). An aliquot of 50 μg of tryptic digest was saved for shotgun-like analysis (preTAILS). The TAILS samples were incubated with a 5-fold excess (w/w) of water soluble HPG-ALD polymer (http://flintbox.com/public/project/1948/) and 20 mM sodium cyanoborohydride, (overnight, 37 °C, pH 6.8) to remove trypsin-generated internal and C-terminal peptides. After quenching the reaction with Tris–HCl (100 mM, pH 6.8, 30 min, 37 °C), unbound peptides representing naturally blocked or experimentally labeled N-terminal peptides were recovered in the filtrate following ultra-filtration (Amicon Ultra-0.5, MWCO 10 kDa). All preTAILS and TAILS samples were desalted using C18 StageTips [41], snap frozen in liquid nitrogen, and stored at −80 °C until LC-MS/MS analysis.

2.3. Mass spectrometry

Purified peptide samples were analyzed using a quadrupole time-of-flight mass spectrometer (Accurate Mass G6550A Q-TOF) coupled online to an 1200 Series nanoflow HPLC (160 nl enrichment column; 0.075 mm × 150 mm analytical column packed with Zorbax 300SB-C18 5 μm stationary phase) with a Chip Cube ionization interface (all Agilent Technologies) with temperature set at 6 °C. Each sample was automatically loaded on the enrichment column at flow rate 4 μl/min of Buffer A (0.1% formic acid in HPLC-grade water) and at 4 μl injection flush volume. After that, a 110.2 min gradient was established with the nano-pump at 300 nl/min from 0% to 5% Buffer B (99.9% acetonitrile, 0.1% formic acid) over 2 min, then from 5% to 45% Buffer B in the next 78 min, then increased to 60% over 10 min period, further increased to 95% Buffer B over 0.1 min, held at 95% for 20 min, and then reduced to 3% Buffer B for 0.1 min to recondition the column for the next analysis. Peptides were ionized by electrospray ionization (ESI; 1.8 kV), and mass spectrometry analysis was performed in

positive polarity with precursor ions detected from 300 to 2000 *m/z*. The top three ions per scan were selected for collision-induced dissociation (CID) using a narrow exclusion window of 1.3 amu (atomic mass unit) and at a MS/MS scan rate of two spectra per second. Collision energy was calculated automatically depending on the charge state of the parent ions, and precursor ions were then excluded from further CID for 30 s. The entire LC-MS/MS system was run by Mass Hunter version B.02.01 (Agilent Technologies).

2.4. Data analysis

Conversion of acquired MS/MS raw data into mgf and mzXML files was performed using MSConvert [42]. Four different search engines, namely Mascot v2.4 [2], X! TANDEM CYCLONE TPP 2011.12.01.1 [3,14], MS-GF+ v10072 [5], and Comet 2015.01 rev 0 [4], were used for peptide spectral matching in the human UniProt protein database (release October 2013). The following database search criteria were applied: semi-ArgC cleavage pattern allowing for 2 missed cleavages; 20 ppm tolerance for MS1 and 50 ppm for MS2 (0.25 Da in case of Mascot [2]); cysteine carbamidomethylation ($+57.0215$ Da) and lysine dimethylation ($+34.0631$ Da) were set as fixed modifications; N-terminal acetylation ($+42.0106$ Da) and dimethylation ($+34.0631$ Da) were set as variable. Further variable modifications included: cyclization of N-terminal glutamine (Gln- > pyro-Glu; -17.0266 Da), glutamate (Glu- > pyro-Glu; -18.0106 Da), and carbamidomethylated cysteine (pyro-cmC; -17.0266 Da), and Met oxidation ($+15.9949$ Da). PeptideProphet [11] and iProphet [12] were used within the Trans Proteomic Pipeline (v4.8.0 PHILAE) [14] to statistically evaluate and combine all identified Peptide Spectrum Matches (PSMs) using a 1% false discovery rate (FDR) cut-off. Protein-Prophet [13] was used for peptide grouping and only proteins with a probability ≥ 0.95 were reported, corresponding to a protein FDR of approx. 0.7%. Note, if peptides match multiple proteins, ProteinProphet [13] determines protein groups (*i.e.* including all proteins identified by the same set of peptides) and names one representative entry. Proteins and protein N-termini were annotated using neXtProt (release 2014-09-19) [43], UniProtKB (April 2015) [24], and TopFIND [26,25,27]; *e.g.* chromosomal location, protein evidence status, sequence position and curation status of identified protein N-termini, and sequence distances to known protein maturation sites such as signal-, transit-, and pro-peptide removal sites, were added to the respective entries.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2015.10.003.

References

[1] U. Eckhard, G. Marino, S.R. Abbey, G. Tharmarajah, I. Matthew, C.M. Overall, The human dental pulp proteome and N-Terminome: levering the unexplored potential of semitryptic peptides enriched by TAILS to identify missing proteins in the human proteome project in underexplored tissues, J. Proteome Res. 14 (2015) 3568–3582. http://dx.doi.org/10.1021/acs.jproteome.5b00579.

[2] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis 20 (1999) 3551–3567. http://dx.doi.org/10.1002/(SICI)1522-2683(19991201)20:18 < 3551::AID-ELPS3551 > 3.0.CO;2-2.

[3] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, Bioinformatics 20 (2004) 1466–1467. http://dx.doi.org/10.1093/bioinformatics/bth092.

[4] J.K. Eng, T.A. Jahan, M.R. Hoopmann, Comet: an open-source MS/MS sequence database search tool, Proteomics 13 (2013) 22–24. http://dx.doi.org/10.1002/pmic.201200439.

[5] S. Kim, P.A. Pevzner, MS-GF+ makes progress towards a universal database search tool for proteomics, Nat. Commun. 5 (2014) 5277. http://dx.doi.org/10.1038/ncomms6277.

[6] O. Kleifeld, A. Doucet, A. Prudova, U. auf dem Keller, M. Gioia, J.N. Kizhakkedathu, et al., Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates, Nat. Protoc. 6 (2011) 1578–1611. http://dx.doi.org/10.1038/nprot.2011.382.

[7] O. Kleifeld, A. Doucet, U. auf dem Keller, A. Prudova, O. Schilling, R.K. Kainthan, et al., Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products, Nat. Biotechnol. 28 (2010) 281–288. http://dx.doi.org/10.1038/nbt.1611.

[8] A. Prudova, K. Serrano, U. Eckhard, N. Fortelny, D.V. Devine, C.M. Overall, TAILS N-terminomics of human platelets reveals pervasive metalloproteinase dependent proteolytic processing in storage, Blood (2014), http://dx.doi.org/10.1182/blood-2014-04-569640, blood–2014–04–569640.

[9] U. Auf dem Keller, A. Prudova, U. Eckhard, B. Fingleton, C.M. Overall, Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation, Sci. Signal. (2013), http://dx.doi.org/10.1126/scisignal.2003512, rs2–rs2.

[10] A. Prudova, U. auf dem Keller, G.S. Butler, C.M. Overall, Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics, Mol. Cell. Proteomics 9 (2010) 894–911. http://dx.doi.org/10.1074/mcp.M000050-MCP201.

[11] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, Anal. Chem. 74 (2002) 5383–5392.

[12] D. Shteynberg, E.W. Deutsch, H. Lam, J.K. Eng, Z. Sun, N. Tasman, et al., iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates, Mol. Cell. Proteomics (2011), http://dx.doi.org/10.1074/mcp.M111.007690, M111.007690.

[13] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, Anal. Chem. 75 (2003) 4646–4658.

[14] E.W. Deutsch, L. Mendoza, D. Shteynberg, J. Slagel, Z. Sun, R.L. Moritz, Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics, Proteomics Clin. Appl. (2015), http://dx.doi.org/10.1002/prca.201400164.

[15] D.L. Tabb, C.G. Fernando, M.C. Chambers, MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis, J. Proteome Res. 6 (2007) 654–661. http://dx.doi.org/10.1021/pr0604054.

[16] V. Dorfer, P. Pichler, T. Stranzl, J. Stadlmann, T. Taus, S. Winkler, et al., MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra, J. Proteome Res. 13 (2014) 3679–3684. http://dx.doi.org/10.1021/pr500202e.

[17] L.Y. Geer, S.P. Markey, J.A. Kowalak, L. Wagner, M. Xu, D.M. Maynard, et al., Open mass spectrometry search algorithm, J. Proteome Res. 3 (2004) 958–964. http://dx.doi.org/10.1021/pr0499491.

[18] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372. http://dx.doi.org/10.1038/nbt.1511.

[19] M. Vaudel, J.M. Burkhart, R.P. Zahedi, E. Oveland, F.S. Berven, A. Sickmann, et al., PeptideShaker enables reanalysis of MS-derived proteomics data sets, Nat. Biotechnol. 33 (2015) 22–24. http://dx.doi.org/10.1038/nbt.3109.

[20] B.C. Searle, Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies, Proteomics 10 (2010) 1265–1269. http://dx.doi.org/10.1002/pmic.200900437.

[21] G. Marino, U. Eckhard, C.M. Overall, Protein Termini and Their Modifications Revealed by Positional Proteomics, ACS Chem. Biol. 10 (2015) 1754–1764. http://dx.doi.org/10.1021/acschembio.5b00189.

[22] Z.W. Lai, A. Petrera, O. Schilling, Protein amino-terminal modifications and proteomic approaches for N-terminal profiling, Curr. Opin. Chem. Biol. 24 (2015) 71–79. http://dx.doi.org/10.1016/j.cbpa.2014.10.026.

[23] P.F. Lange, P.F. Huesgen, K. Nguyen, C.M. Overall, Annotating N termini for the human proteome project: N termini and Nα-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome, J. Proteome Res. 13 (2014) 2028–2044. http://dx.doi.org/10.1021/pr401191w.

[24] UniProt Consortium, UniProt: a hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212. http://dx.doi.org/10.1093/nar/gku989.

[25] N. Fortelny, S. Yang, P. Pavlidis, P.F. Lange, C.M. Overall, Proteome TopFIND 3.0 with TopFINDer and PathFINDer: database and analysis tools for the association of protein termini to pre- and post-translational events, Nucleic Acids Res. 43 (2015) D290–D297. http://dx.doi.org/10.1093/nar/gku1012.

[26] P.F. Lange, C.M. Overall, TopFIND, a knowledgebase linking protein termini with function, Nat. Methods 8 (2011) 703–704. http://dx.doi.org/10.1038/nmeth.1669.

[27] P.F. Lange, P.F. Huesgen, C.M. Overall, TopFIND 2.0–linking protein termini with proteolytic processing and modifications altering protein function, Nucleic Acids Res. 40 (2012) D351–D361. http://dx.doi.org/10.1093/nar/gkr1025.

[28] O. Schilling, U. auf dem Keller, C.M. Overall, Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites, Biol. Chem. 392 (2011) 1031–1037. http://dx.doi.org/10.1515/BC.2011.158.

[29] O. Schilling, P.F. Huesgen, O. Barré, U. Auf dem Keller, C.M. Overall, Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry, Nat. Protoc. 6 (2011) 111–120. http://dx.doi.org/10.1038/nprot.2010.178.

[30] O. Schilling, C.M. Overall, Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites, Nat. Biotechnol. 26 (2008) 685–694. http://dx.doi.org/10.1038/nbt1408.

[31] G. Marino, P.F. Huesgen, U. Eckhard, C.M. Overall, W.P. Schröder, C. Funk, Family-wide characterization of matrix metal-loproteinases from Arabidopsis thaliana reveals their distinct proteolytic activity and cleavage site specificity, Biochem. J. 457 (2014) 335–346. http://dx.doi.org/10.1042/BJ20130196.

[32] U. Eckhard, P.F. Huesgen, H. Brandstetter, C.M. Overall, Proteomic protease specificity profiling of clostridial collagenases reveals their intrinsic nature as dedicated degraders of collagen, J. Proteomics 100 (2014) 102–114. http://dx.doi.org/10.1016/j.jprot.2013.10.004.

[33] O. Barré, A. Dufour, U. Eckhard, R. Kappelhoff, F. Béliveau, R. Leduc, et al., Cleavage specificity analysis of six type II transmembrane serine proteases (TTSPs) using PICS with proteome-derived peptide libraries, PLoS One 9 (2014) e105984. http://dx.doi.org/10.1371/journal.pone.0105984.

[34] U. auf dem Keller, C.M. Overall, CLIPPER: an add-on to the Trans-Proteomic Pipeline for the automated analysis of TAILS N-terminomics data, Biol. Chem. 393 (2012) 1477–1483. http://dx.doi.org/10.1515/hsz-2012-0269.

[35] R. Kappelhoff, U. Auf dem Keller, C.M. Overall, Analysis of the degradome with the CLIP-CHIP microarray, Methods Mol. Biol. 622 (2010) 175–193. http://dx.doi.org/10.1007/978-1-60327-299-5_10.

[36] P.F. Huesgen, P.F. Lange, L.D. Rogers, N. Solis, U. Eckhard, O. Kleifeld, et al., LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification, Nat. Methods 12 (2015) 55–58. http://dx.doi.org/10.1038/nmeth.3177.

[37] L. Tsiatsiani, A.J.R. Heck, Proteomics beyond trypsin, FEBS J. 282 (2015) 2612–2626. http://dx.doi.org/10.1111/febs.13287.

[38] X. Guo, D.C. Trudgian, A. Lemoff, S. Yadavalli, H. Mirzaei, Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics, Mol. Cell. Proteomics 13 (2014) 1573–1584. http://dx.doi.org/10.1074/mcp.M113.035170.

[39] T. Glatter, C. Ludwig, E. Ahrné, R. Aebersold, A.J.R. Heck, A. Schmidt, Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion, J. Proteome Res. 11 (2012) 5145–5156. http://dx.doi.org/10.1021/pr300273g.

[40] D. Wessel, U.I. Flügge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, Anal. Biochem. 138 (1984) 141–143.

[41] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2 (2007) 1896–1906. http://dx.doi.org/10.1038/nprot.2007.261.

[42] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, et al., A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol. 30 (2012) 918–920. http://dx.doi.org/10.1038/nbt.2377.

[43] P. Gaudet, G. Argoud-Puy, I. Cusin, P. Duek, O. Evalet, A. Gateau, et al., neXtProt: organizing protein knowledge in the context of human proteome projects, J. Proteome Res. 12 (2013) 293–298. http://dx.doi.org/10.1021/pr300830v.