



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

## Active site specificity profiling datasets of matrix metalloproteinases (MMPs) 1, 2, 3, 7, 8, 9, 12, 13 and 14



Ulrich Eckhard<sup>a,10</sup>, Pitter F. Huesgen<sup>a,1,10</sup>, Oliver Schilling<sup>a,2,10</sup>,  
 Caroline L. Bellac<sup>a,3,11</sup>, Georgina S. Butler<sup>a,11</sup>,  
 Jennifer H. Cox<sup>a,4,11</sup>, Antoine Dufour<sup>a,11</sup>, Verena Goebeler<sup>a,11</sup>,  
 Reinhild Kappelhoff<sup>a,11</sup>, Ulrich auf dem Keller<sup>a,5,11</sup>,  
 Theo Klein<sup>a,11</sup>, Philipp F. Lange<sup>a,6,11</sup>, Giada Marino<sup>a,11</sup>,  
 Charlotte J. Morrison<sup>a,7,11</sup>, Anna Prudova<sup>a,11</sup>,  
 David Rodriguez<sup>a,8,11</sup>, Amanda E. Starr<sup>a,9,11</sup>, Yili Wang<sup>a,11</sup>,  
 Christopher M. Overall<sup>a,b,\*</sup>

<sup>a</sup> Centre for Blood Research, Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, BC, Canada

<sup>b</sup> Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada

## ARTICLE INFO

## Article history:

Received 24 October 2015

Received in revised form

12 February 2016

Accepted 15 February 2016

## ABSTRACT

The data described provide a comprehensive resource for the family-wide active site specificity portrayal of the human matrix metalloproteinase family. We used the high-throughput proteomic technique PICS (Proteomic Identification of protease Cleavage Sites) to comprehensively assay 9 different MMPs. We identified

\* Corresponding author at: Centre for Blood Research, Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, BC, Canada.

E-mail address: [chris.overall@ubc.ca](mailto:chris.overall@ubc.ca) (C.M. Overall).

<sup>1</sup> Present address: Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, Germany.

<sup>2</sup> Present address: Institute of Molecular Medicine and Cell Research, University of Freiburg, Freiburg, Germany.

<sup>3</sup> Present address: Swissmedic, Swiss Agency for Therapeutic Products, Bern, Switzerland.

<sup>4</sup> Present address: Inception Sciences, Vancouver, BC, Canada.

<sup>5</sup> Present address: Institute of Molecular Health Sciences, ETH Zürich, Zürich, Switzerland.

<sup>6</sup> Present address: Department of Pathology, University of British Columbia, Vancouver, BC, Canada.

<sup>7</sup> Present address: Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

<sup>8</sup> Present address: Department of Bioquímica y Biología Molecular, Universidad de Oviedo, Oviedo, Spain.

<sup>9</sup> Present address: Ottawa Institute of Systems Biology, University of Ottawa, Canada.

<sup>10</sup> Contributed equally.

<sup>11</sup> In alphabetical order.

<http://dx.doi.org/10.1016/j.dib.2016.02.036>

2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Available online 22 February 2016

**Keywords:**  
Matrix metalloproteinases  
MMPs  
PICS  
Proteomics  
Quenched fluorescence  
Specificity profiling  
Cleavage sites

more than 4300 peptide cleavage sites, spanning both the prime and non-prime sides of the scissile peptide bond allowing detailed subsite cooperativity analysis. The proteomic cleavage data were expanded by kinetic analysis using a set of 6 quenched-fluorescent peptide substrates designed using these results. These datasets represent one of the largest specificity profiling efforts with subsequent structural follow up for any protease family and put the spotlight on the specificity similarities and differences of the MMP family. A detailed analysis of this data may be found in Eckhard et al. (2015) [1]. The raw mass spectrometry data and the corresponding metadata have been deposited in PRIDE/ProteomeXchange with the accession number **PXD002265**.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific sub- ject area	Proteolytic enzymes, metalloproteinases, substrate specificity profiling, inhibitor design, drug discovery, matrix biology, extra cellular matrix (ECM).
Type of data	Mass spectrometry raw-files; search engine output files; metadata; quenched fluorescent peptide cleavage data; specificity profiling analysis (.xlsx).
How data was acquired	Liquid chromatography tandem mass spectrometry (LC-MS/MS): either QSTAR XL or QSTAR Pulsar I (Applied Biosystems) mass spectrometer coupled on-line to LC Packings capillary LC system (Dionex).
Data format	RAW files: .wiff and .mzXML-files; .tandem and .pepxml post-database search output files from X! Tandem [2].
Experimental factors	(A) Human MMPs 1, 2, 3, 8, 9 and 13 were expressed and purified from CHO or <i>Timp2</i> <sup>-/-</sup> MEF-conditioned medium [3–7]. Soluble MMP14 (A21-R513) was purified from <i>Pichia pastoris</i> [3]. MMP7 was purchased from Enzo Life Sciences and MMP12 was a kind gift from Novartis. ProMMP3 was activated using chymotrypsin; all other proMMPs were activated using APMA. (B) Proteome-derived peptide libraries were prepared from K562 cells in the presence of protease inhibitors [8,9]. Proteins were denatured and cysteine side-chains alkylated. After reaction clean-up, proteomes were digested with trypsin or GluC giving orthogonal peptide libraries. Primary amines were dimethylated, peptide libraries purified, and stored as 200 µg aliquots at – 80 °C.
Experimental features	PICS cleavage assays [8,9] were performed by incubating peptide libraries with MMP. Cleaved peptides with neo-N-termini were biotinylated and affinity purified. Eluates were desalted and analyzed by LC-MS/MS. Spectra were matched to peptides using X! Tandem [2] and statistically evaluated with PeptideProphet [10,11]. Identified peptides represent prime-side cleavage products and complete cleavage sites were reconstructed using Web-PICS [12].
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 have been deposited in PRIDE/ProteomeXchange with the accession number <b>PXD002265</b> .

## Value of the data

- Comprehensive specificity profiling data of nine matrix metalloproteinases using PICS proteomics.
- Largest compendium of matrix metalloproteinase P6–P6' cleavage sites, reporting more than 4300 cleaved peptides.
- Identified cleavage sites allow in-depth cooperativity analysis of specificity subsites.
- Quenched fluorescent peptide cleavage data corroborating the specificity profiles.
- These data can guide small molecule drug development and help in discrimination between direct and indirect MMP targets.

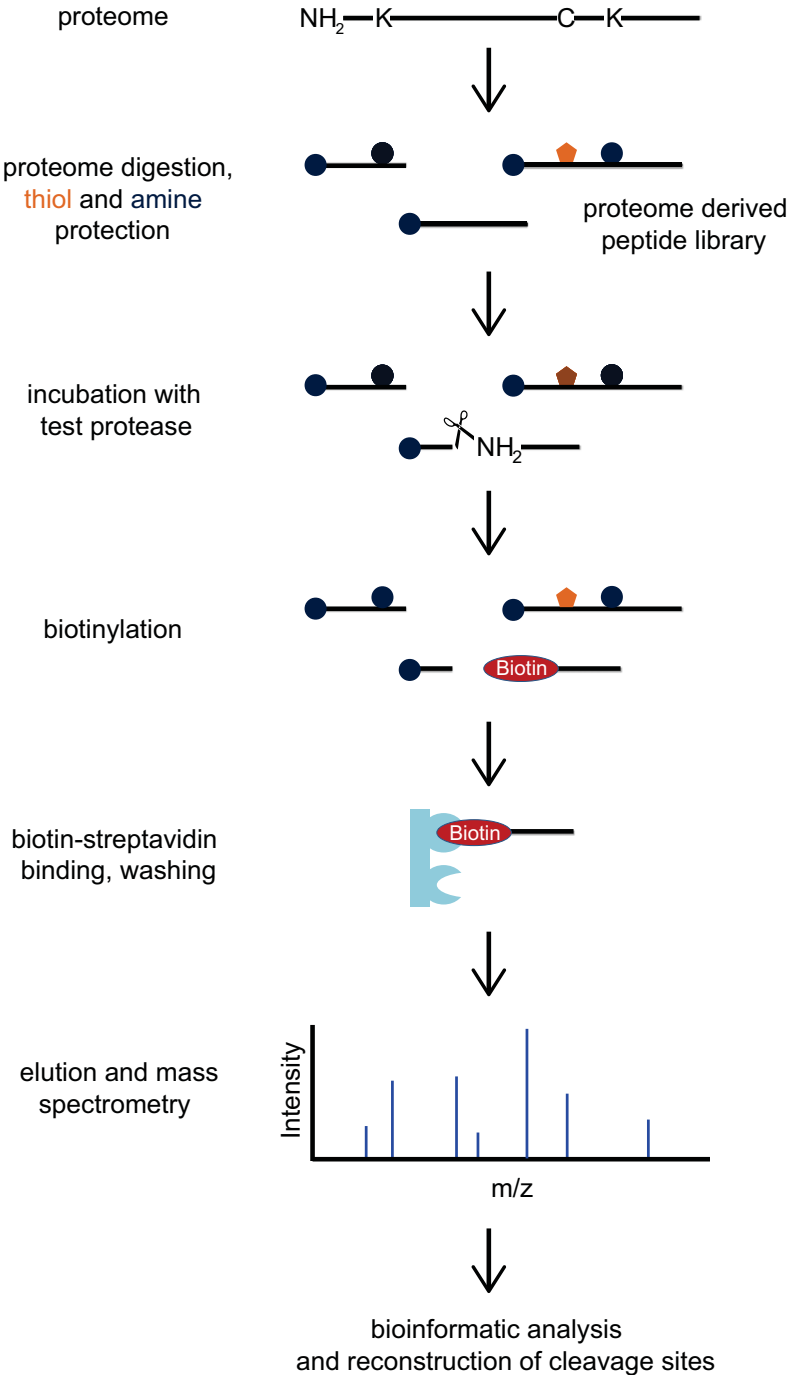
## 1. Data

Matrix metalloproteinases (MMPs) regulate the structural matrix environment and extracellular signaling by precise proteolytic cleavage. Unraveling complex *in vivo* proteolytic networks is challenging. Thus comprehensive specificity profiles of all proteases involved are needed to guide interpretation.

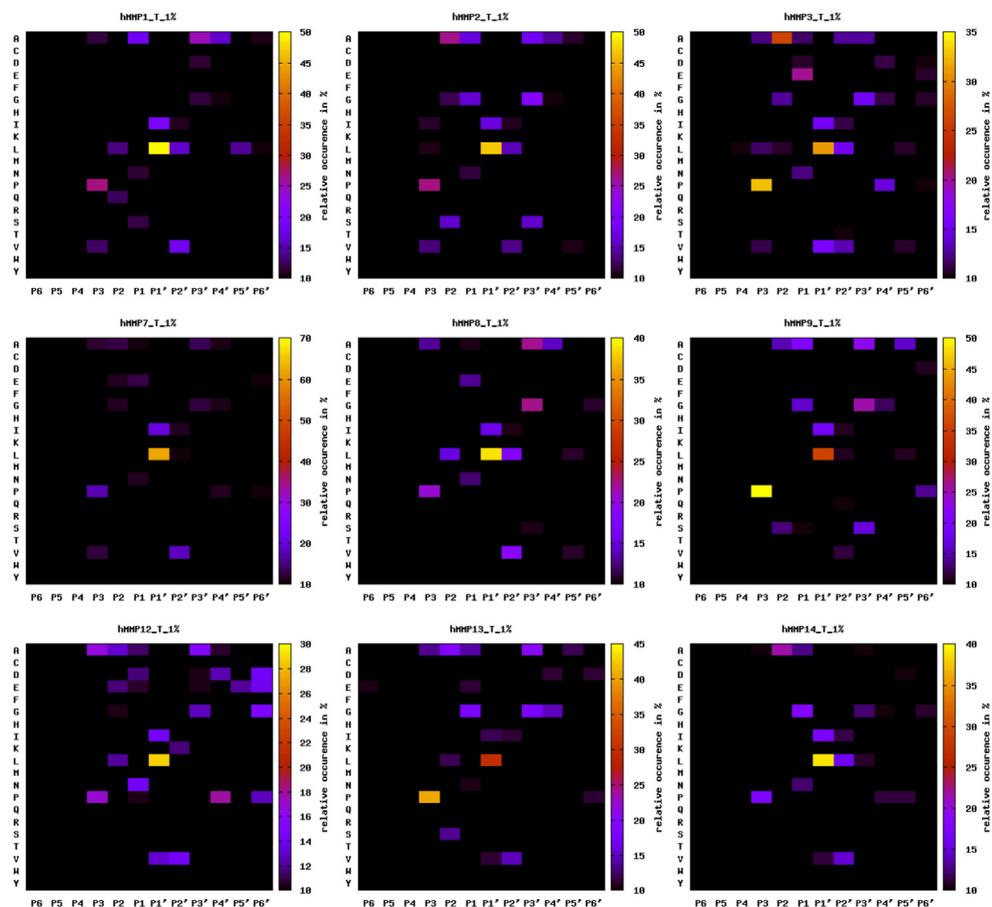
The data in the ProteomeXchange archive (PXD002265) and accompanying data of the present article provide a comprehensive resource for the individual assessment of the active site specificity of nine representative members of the human matrix metalloproteinase (MMP) family. The in-depth specificity comparison based on these proteomic data corroborated with kinetic analysis using a set of 6 quenched fluorescent peptides and *in silico* peptide docking was presented recently [1].

MMPs 1, 2, 3, 7, 8, 9, 12, 13 and 14 were all assayed by the high-throughput Proteomic Identification of protease Cleavage Sites (PICS; or PICS proteomics) method [8,9] and using two orthogonal human whole-proteome peptide libraries (generated with trypsin or GluC). These data were analyzed using X! Tandem [2] for peptide-spectrum matching, and PeptideProphet [11] for statistical evaluation. However, other search engines such as Mascot [13], MS-GF+ [14], Comet [15], or MS Amanda [16] can be used to extend the number of matched spectra by combining the results using e.g. iProphet [17] within the Trans Proteomic Pipeline [10] or PeptideShaker [18]. Here we provide additional data to enable other researchers to (i) reinvestigate our analysis to identify additional subsite cooperativity effects and so far unexplored specificity preferences, and (ii) reanalyze our raw-data with entirely new concepts or ideas in mind, such as specificity-oriented protease evolution [19,20], functional phylogeny [21], or substrate-driven mapping of the degradome [22].

A representative PICS workflow is depicted in Fig. 1, and a graphical representation of the various MMP domain architectures is shown in Supplementary Fig S1. Specificity profiling results were summarized as heat maps in Figs. 2 and 3 (trypsin generated libraries) and Supplementary Figs 2 and 3 (GluC-generated libraries). Supplementary Table S1 provides a general resource for the MMPs analyzed, giving their domain boundaries and links to the major databases in protein and protease research: UniProt [23], Pfam [24], MEROPS [25], and TopFIND [26–28]. Supplementary Table S2 depicts key residues and all secondary structure elements characterizing the different MMP catalytic domains, and shows a structural overlay of all analyzed MMPs. Supplementary Tables 3 and 4 contain the individual MMP specificity profiling results obtained from trypsin- and GluC-generated peptide libraries, respectively. Identified thioacylated prime-side cleavage products (column A), the WebPICS [12] results (columns C–I) including HeatMaps ([www.gnuplot.info](http://www.gnuplot.info)) and iceLogos [29], and the individual subsite analysis (columns K–W) together with the amino acid distribution in the human proteome (UniProt release 2013\_10) used for calculating normalized abundances (columns Y–AA) are shown. Additionally, a high resolution PICS workflow is shown on both index pages, and all original WebPICS results are provided in the supplement as a combined zip-file. Supplementary Table S5 provides the data matrices underlying our subsite cooperativity analysis, in which we fixed certain subsites (e.g. P3-Pro or P1'-Leu) and analyzed the changes of the individual amino acid occurrences in the other subsites (P6–P6'). Supplementary Table S6 provides the raw data of our quenched fluorescent cleavage assays, including a graphical representation of two of the peptides (PLG↓L and PAN↓L) and a high resolution table summarizing the results normalized to the standard MMP substrate PLG↓L (originally referred to as QF-24) [30]. Detailed database search settings for PICS data analysis are given



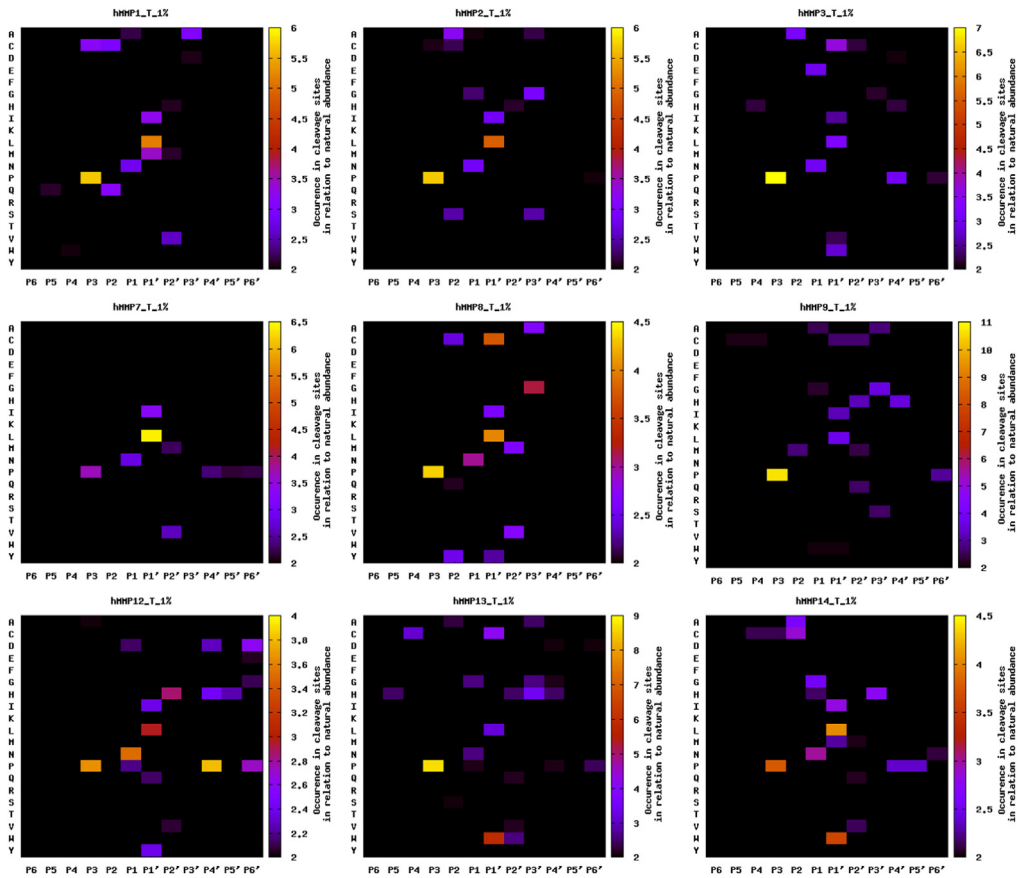
**Fig. 1.** Proteomic Identification of protease Cleavage Sites (PICS) workflow. PICS libraries are generated from cellular proteomes, e.g. human cell cultures, and thus represent biological sequence diversity. Specific endoproteases such as trypsin, GluC, or chymotrypsin are then used to digest proteomes into peptides amenable for mass spectrometry. Primary amines and thiols are chemically protected and the peptide library is purified. Next, the proteome-derived peptide library is incubated with the endoprotease of interest (e.g. MMP1). Prime side cleavage products (i.e. peptide fragments C-terminal of the cleavage site depicted between P1 and P1') possess free and thus reactive N-termini that are subsequently tagged with cleavable biotin allowing specific isolation with immobilized streptavidin. Following elution, prime-side cleavage products are identified by LC-MS/MS, and the corresponding non-prime sequences are reconstructed bioinformatically, e.g. by using the free webservice WebPICS (<http://clipserve.clip.ubc.ca/pics/>) [12].



**Fig. 2.** Sequence specificity profiles of MMPs 1, 2, 3, 7, 8, 9, 12, 13, and 14 using trypsin-generated human peptide libraries. Identified cleavage sites are summarized as heat maps showing relative amino acid occurrence. P6–P6' subsite positions are shown on the x axes with the identified cleavage site between P1 and P1'. Plotted amino acids are indicated on the y axes with single-letter codes. Please refer to [Supplementary Fig S2](#) for corresponding results using GluC-generated peptide libraries.

in [Supplementary Table S7](#). Additionally, a combined zip-file of all WebPICS results can be found in the supplement. All mass spectrometry raw data and corresponding metadata have been deposited in the ProteomeXchange Consortium database (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [51] with the PXD identifier <PXD002265>.

We previously used PICS to characterize a wide selection of different proteases, such as clostridial collagenases [31], plant metalloproteinases [32], human type II transmembrane serine proteases (TTSPs) [33], the *archaeal* protease LysargiNase [34], snake venom serine proteinases [35], human coagulation factor Xa [12], and caspases 3 and 7 [8], proving both its versatility and robustness. Importantly, PICS is designed for active site specificity profiling and not for the identification of native substrates. For the latter task, the Overall lab developed TAILS (Terminal Amine Isotopic Labeling of Substrates) [36,37], which we have successfully used e.g. for the identification of natural substrates of dipeptidyl peptidases 8 and 9 [38], the human gelatinases MMP2 and 9 [39], membrane-type 6 matrix metalloproteinase MMP25 [40] and the meprins [41]. Over the last few years we have adapted TAILS for the study of complex *in vivo* biological systems [42] and identification of N-terminal modifications such as proteolytic processing that alter protein stability or function [43,44]. We have assessed the N-terminome of various tissues and cells, such as skin [42], erythrocytes [43], platelets [45], and dental



**Fig. 3.** MMP sequence specificity profiles identified in trypsin-generated human peptide libraries and summarized as heat maps showing fold-change over natural abundance. P6–P6′ subsite positions are shown on the x axes with the identified cleavage site between P1 and P1′. Plotted amino acids are indicated on the y axes with single-letter codes. Please refer to [Supplementary Fig S3](#) for corresponding results using GluC-generated peptide libraries.

pulp [46]. In combination, TAILS N-terminomics and PICS proteomics allow an in depth characterization of any biological system and protease in an unbiased, proteomics-centered manner.

The following materials and methods section will enable other investigators and laboratories to design similar experimental procedures to study matrix metalloproteinases or any other protease by PICS proteomics. Please refer to our recent dental pulp proteomics and N-terminomics Data in Brief article for more information on TAILS [47].

## 2. Experimental design, materials and methods

### 2.1. Expression and purification of human MMPs

#### 2.1.1. Summary

MMPs 1, 2, 3, 8, 9 and 13 were expressed as zymogens using the pGW1HG vector (kindly provided by British Biotech Pharmaceuticals, Oxford, UK), and purified from serum-free conditioned medium from (i) Chinese hamster ovary (CHO) cells or (ii) murine embryonic *Timp2*<sup>-/-</sup> fibroblasts (MEFs) [3–7]. Soluble MMP14 lacking the C-terminal transmembrane and cytoplasmic domain (A21–R513) was

purified from *Pichia pastoris* [3]. Purified proMMPs were aliquoted into single use aliquots of 10 µg, flash-frozen with liquid nitrogen, and stored at  $-70^{\circ}\text{C}$  until use. For more expression and purification details, see below. Active MMP7 was purchased from Enzo Life Sciences, and MMP12 was a kind gift from Novartis Pharma AG (Basel, CH). ProMMP3 was activated using chymotrypsin. All others proMMPs were activated using APMA.

### 2.1.2. ProMMP1

Chinese hamster ovary (CHO-K1) cells (American Type Culture Collection) were maintained in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% cosmic calf serum (HyClone Laboratories, Inc.) and non-essential amino acids (Invitrogen). Cells were transfected with pGW1HG-MMP1 and selected with 25 µg/ml mycophenolic acid (Invitrogen). MMP1-expressing clones were expanded to confluence in roller bottles (850 cm<sup>2</sup>, BD Biosciences), washed with phosphate-buffered saline (PBS; 138 mM NaCl, 2.7 mM KCl, 20 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4), and incubated in 100 ml of serum-free CHO-S-SFM II medium (Invitrogen). Conditioned serum-free medium was collected every 1–2 days for up to 8 days. ProMMP1 was purified from collected culture supernatants using an (i) Orange-Sepharose column equilibrated in MES buffer (50 mM MES, pH 6.0, 5 mM CaCl<sub>2</sub>, 0.1 M NaCl, 0.025% sodium azide), and eluted with 1 M NaCl (in MES buffer). Elution fractions were subsequently loaded on (ii) Zn<sup>2+</sup>-chelating Sepharose Fast Flow resin (Amersham Biosciences), and chromatographed with a linear imidazole gradient (0 to 0.5 M). Fractions containing proMMP1 were pooled and dialyzed into HEPES buffer (50 mM HEPES pH 7.2, 5 mM CaCl<sub>2</sub>, 0.1 M NaCl).

### 2.1.3. ProMMP2

TIMP-2-free human proMMP2 was expressed in ras/myc-transformed *Timp2*<sup>-/-</sup> fibroblasts. Cells were grown in Dulbecco's modified Eagle's medium with 10% cosmic calf serum (HyClone Laboratories Inc), transfected with MMP2-pGW1HG, and selected by using 25 µg/ml mycophenolic acid. Serum-free conditioned medium was harvested from roller bottles and proMMP2 was purified at 4 °C in MES buffer by (i) gelatin-Sepharose (Amersham Biosciences) chromatography. After elution with 10% dimethyl sulfoxide in HEPES buffer, samples were dialyzed into MES buffer, and loaded in tandem onto (ii) lentil lectin-Sepharose (Sigma) to remove MMP9 and fibronectin, and (iii) a 1 ml gelatin-Sepharose column to capture proMMP2. After elution using 10% dimethyl sulfoxide (gelatin-Sepharose column only), fractions containing proMMP2 were pooled and dialyzed into MES buffer.

### 2.1.4. ProMMP3

Recombinant C-terminally FLAG-tagged human proMMP3 was expressed from pGW1HG in CHO-K1 cells and purified from supernatants in MES buffer using a (i) Green-agarose (Sigma) column. After elution with 1 M NaCl, eluates were loaded on a (ii) Zn<sup>2+</sup>-chelating column (Amersham Biosciences) and chromatographed with an imidazole gradient. Fractions containing proMMP3 were pooled, dialyzed into Tris-buffered saline (TBS; 50 mM Tris, 150 mM NaCl, pH 7.4), and subsequently loaded onto an (iii) anti-FLAG-agarose column (Sigma). After elution with 100 mM glycine, pH 3.5, fractions were immediately adjusted to pH 7–8 using 1 M Tris pH 8.0, and fractions containing proMMP3 were pooled and dialyzed into HEPES buffer.

### 2.1.5. ProMMP8

CHO-K1 cells were transfected with pGW1HG-MMP8 and selected using 25 µg/ml mycophenolic acid (Invitrogen); conditioned medium was collected from roller bottles. To remove gelatinases (MMP2 and MMP9), culture supernatants were chromatographed over (i) gelatin-Sepharose 4B resin (Amersham Biosciences) connected in tandem with (ii) a Red-Sepharose CL-6B column (Amersham Biosciences). MMP8 was eluted from Red-Sepharose with 1 M NaCl in TBS, and (iii) loaded on a column of Zn<sup>2+</sup>-chelating Sepharose resin (Amersham Biosciences). Fractions containing MMP-8 were pooled and chromatographed over (iv) lentil lectin-agarose-Sepharose 4B (Sigma-Aldrich) and eluted with 100 mM α-D-methylmannopyranoside (Sigma) in TBS. Purified proMMP8 was buffer exchanged into collagenase assay buffer (50 mM Tris, 200 mM NaCl, 5 mM CaCl<sub>2</sub>, 0.05% Brij-35, pH 7.4) using a PD-10 Sephadex G-25 column (Amersham Biosciences).



### 2.1.6. **ProMMP9**

Human MMP9 was expressed from pGW1HG in CHO-K1 cells with 25 µg/ml mycophenolic acid for selection. ProMMP9 was captured from conditioned medium on a gelatin-Sepharose column (Amersham Biosciences) in MES buffer, and eluted after extensive washing in MES buffer supplemented with 10% dimethyl sulfoxide. ProMMP-9 was dialyzed into HEPES buffer.

### 2.1.7. **ProMMP13**

Recombinant C-terminally FLAG-tagged human MMP13 was expressed in CHO-K1 cells from pGW1HG and purified from culture supernatants using a (i) green-agarose column (Sigma). After extensive washing with MES buffer, bound protein was eluted using 1 M NaCl, and fractions containing MMP-13 were dialyzed into TBS. MMP13 was then purified to homogeneity using an (ii) anti-FLAG-agarose column (Sigma). After elution with 100 mM glycine (pH 3.5), fractions were immediately adjusted to pH 7–8 using 1 M Tris pH 8.0. ProMMP13 was dialyzed into HEPES buffer.

### 2.1.8. **ProMMP14**

Soluble MT1-MMP with a FLAG tag in place of the transmembrane and cytoplasmic domains was cloned into pPIC9 (Invitrogen) and expressed in Pichia GS115 cells (Invitrogen). Cells were grown in 500 ml baffled flasks. After 24 h, 0.5% methanol was added to induce recombinant protein expression. Culture medium was diluted in MES buffer and MT1-MMP was purified using a red agarose column (Sigma). After extensive washing with MES buffer, protein was eluted with 1 M NaCl, and fractions containing proMMP14 were pooled and dialyzed into HEPES buffer.

## 3. **PICS peptide library preparation.**

Human whole proteome-derived peptide libraries for MMP specificity profiling were prepared as described in great detail in Nature Protocols [9]: in brief, cell pellets were collected from human lymphoblast cell K562 cultures and lysed in 20 mM HEPES (pH 7.5) supplemented with 0.1% (w/v) SDS and protease inhibitors to prevent unwanted proteolysis (1 × Roche cOmplete plus 1 mM PMSF and 10 mM EDTA). Cell debris was removed by centrifugation (26,000g, 1 h, 4 °C); soluble proteins were denatured using guanidine hydrochloride (4 M), and cysteine side-chains were reduced with 20 mM DTT (1 h, 37 °C). Free sulfhydryl groups were protected with 40 mM iodoacetamide (3 h, 20 °C) to avoid peptide crosslinking and reactions were stopped by adding more DTT (5 mM, 15 min, 20 °C). Reaction clean-up was performed using chloroform/methanol precipitation as described elsewhere [48], pellets were air-dried and re-suspended in 100 mM HEPES, 5 mM CaCl<sub>2</sub>, pH 7.5, and digested with TPCK-treated trypsin or GluC (*Staphylococcus aureus* protease V8, Worthington) at a protease to proteome ratio of 1:100 (w/w) overnight at 37 °C. Note, another protease often used for PICS library preparation is chymotrypsin. After inactivation of trypsin/GluC with 1 mM PMSF (30 min, 20 °C), undigested protein aggregates were removed by centrifugation (20,000g, 10 min, 4 °C). Primary amines of peptide N-termini (α-amines) and lysine side chains (ε-amines) were blocked by reductive dimethylation with 30 mM formaldehyde (CH<sub>2</sub>O) and 15 mM sodium cyanoborohydride (NaCNBH<sub>3</sub>, Sterogene) at 20 °C for 16 h overnight (pH 6–7). To ensure completeness of amine blocking, another 15 mM formaldehyde and 15 mM sodium cyanoborohydride were added and incubated for additional two hours. Samples were desalted by size exclusion chromatography using Sephadex G-10 columns (10 mM potassium phosphate buffer, pH 2.7, 10% (v/v) methanol), and after methanol removal by vacuum concentration (SpeedVac, Thermo), peptides were purified by reversed-phase chromatography on an ÄKTA™ high-performance liquid chromatography system (Äkta Explorer, GE Healthcare) using a RESOURCE RPC column (GE Healthcare); wash buffer contained 0.3% (v/v) formic acid, and samples were eluted in 80% (v/v) acetonitrile, both in HPLC-grade H<sub>2</sub>O. These PICS peptide libraries were concentrated by rotary evaporation under vacuum, re-suspended in water, and stored in 200–400 µg aliquots of 5–15 mg/ml at –80 °C until use. Peptide concentration was estimated using the bicinchoninic assay (BCA, Pierce). All reagents were purchased from Sigma-Aldrich unless otherwise specified.



#### 4. PICS cleavage site specificity assay.

MMP cleavage assays were performed by incubation of 200–400 µg human whole-proteome peptide library with active recombinant MMP at a protease to peptide library ratio of 1:100 (w/w) in 50 mM HEPES, 150 mM NaCl, 5 mM CaCl<sub>2</sub> at pH 7.4, overnight, and stopped by heat inactivation at 70 °C for 30 min. Prime-side cleavage products generated by MMP cleavage were subsequently isolated by positive enrichment using the biotin handle. In short, cleaved peptides with a free primary amine at the N-terminus generated by MMP activity were biotinylated by incubation with 0.5 mM Sulfo-NHS-SS-Biotin, an amine-reactive biotin with a redox-sensitive and thus cleavable disulfide linker (Thermo Scientific) for 2 h at 20 °C. Biotinylated prime-side cleavage products were separated from uncleaved peptides by affinity purification, incubating with 300 µl Streptavidin Sepharose slurry (GE Healthcare) for 2 h with mild agitation. After extensive washing (50 mM HEPES, pH 7.2), biotinylated peptides were eluted with 20 mM DTT (2 h, 20 °C), desalted using reversed-phase solid phase extraction (Sep-Pak C18, Waters) with binding and washing in 0.1% (v/v) formic acid and elution in 80% (v/v) acetonitrile, both in HPLC-grade H<sub>2</sub>O. Eluates were vacuum dried to near dryness using a SpeedVac concentrator (Thermo), brought to 10 µl with 0.1% formic acid, and stored at –80 °C until LC–MS/MS analysis.

#### 5. LC–MS/MS, peptide spectrum matching, and data analysis

LC–MS/MS analysis was performed using an LC Packings capillary LC system (Dionex) coupled online to a quadrupole time-of-flight mass spectrometer operated either by the UBC Center for Blood Research Mass Spectrometry Suite (QSTAR XL; Applied Biosystems), or by the UBC Proteomics Core Facility (QSTAR Pulsar I, Applied Biosystems). Samples were diluted in 0.3% (v/v) formic acid and loaded onto a column packed with Magic C18 resin (Michrom Bioresources). Peptides were eluted using a 2–80% (v/v) acetonitrile gradient in 0.1% (v/v) formic acid over 95 min. MS/MS data were acquired automatically, using Analyst QS software, v1.1 (Applied Biosystems) for data-dependent acquisition based on a 1 s MS survey scan from 350 m/z to 1500 m/z, followed by up to 3 MS/MS scans of 2 s each. Single charged ions were excluded because in ESI mode, peptides typically carry multiple charges. Centroids were calculated for the acquired data that was converted to mzXML format using msConvert [49]. Peptides were identified from the human UniProtKB/SwissProt database containing canonical and isoform protein sequences (downloaded October 2013) using the search engine X! Tandem [2] in conjunction with PeptideProphet [11], both implemented in the Trans Proteomic Pipeline v4.3 [10], at an estimated false discovery rate (FDR) of 1%. Search parameters included a mass tolerance of 200 ppm for parental ions and 0.2 Da (Da) for fragment ions, allowing up to two missed cleavages. The following fixed peptide modifications were set: carbamidomethylation of cysteine side chains (+57.02 Da) and dimethylation of lysine ε-amines (+28.03 Da); methionine oxidation (+15.99 Da). N-terminal dimethylation (+28.03 Da) and thioacylation (+88.00 Da) were set as variables. Note, N-terminally thioacylated peptides identified by LC–MS/MS represent prime-side cleavage products of the proteases of interest. The complete cleavage sites were reconstructed bioinformatically using the open web-based program WebPICS [12], available at <http://clipserve.clip.ubc.ca/pics/>, which generates a non-redundant list of identified cleavage sites by matching each prime side peptide sequence to the human IPI database (v3.69, 174784 entries; EMBL-EBI, UK) and extracting the non-prime cleavage side sequence up to the next cleavage site of the enzyme used for library generation, *i.e.* to the next N-terminal Asp or Glu for GluC-libraries, or the next N-terminal Arg or Lys in the case of trypsin-generated libraries. Subsite positions with ambiguous information coming *e.g.* from different protein isoforms are omitted and replaced by X for further analysis. Identified cleavage sites can be summarized as heat maps, by using *e.g.* Gnuplot ([www.gnuplot.info](http://www.gnuplot.info)), or iceLogos (<http://iomics.ugent.be/iceLogoserver/index.html>) [29]

## 6. Quenched fluorescence protease activity assay

Synthetic quenched fluorescent (QF) peptides were purchased from ChinaPeptides Co. Ltd. (Shanghai, China), dissolved in DMSO and protected from light. Working stocks (100  $\mu$ M) were prepared in DMSO using the molar extinction coefficient of the conjugated quencher (DNP; (2,4)-dinitrophenyl) of  $6.985 \text{ cm}^{-1} \text{ mM}^{-1}$  at 400 nm [50]. MMP zymogens (MMP1, MMP2, MMP8, MMP9, MMP14) were activated in 100 mM Tris, pH 7.5, 100 mM NaCl, 10 mM  $\text{CaCl}_2$ , and 0.05% Brij-35, using 1 mM APMA (para-aminophenylmercuric acetate) at 37 °C for 30 min. Chymotrypsin was used to activate MMP3 at a ratio of 1:100 (w/w) for 30 min at 37 °C, and was subsequently inactivated using 1 mM PMSF. MMP7, MMP12, and MMP13 were typically auto-activated during purification. Quenched fluorescent peptide assays were performed immediately after MMP activation in the presence of protease inhibitor cocktail (HALT™, Life Technologies, no EDTA added) using a multi-wavelength fluorescence scanner (POLARstar OPTIMA, BMG Labtech). Each MMP (1–10 nM) was incubated with 1  $\mu$ M QF-peptide in 100  $\mu$ L of 100 mM Tris, pH 7.5, 100 mM NaCl, 10 mM  $\text{CaCl}_2$ , and 0.05% Brij-35, and the increase in fluorescence was measured at 45 s intervals for 1 h at 37 °C. The excitation and emission wavelengths were set to 320 and 405 nm, respectively, and all measurements were performed in duplicate. Experiments were repeated three times with independent substrate and MMP preparations on consecutive days.

## Acknowledgments

C.M.O. holds a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology. J.H.C. and A.E.S. were supported by graduate fellowships from Natural Sciences and Engineering Research Council of Canada (NSERC), Canadian Institutes of Health Research (CIHR), and Michael Smith Foundation for Health Research (MSFHR). A.P. and G.M. were co-funded by post-doctoral fellowships from the UBC Center for Blood Research. P.F.L. was supported by a Feodor Lynen Research Fellowship of the Alexander von Humboldt Foundation, and O.S. and U.a.d.K. were supported by fellowships from the German Research Foundation (DFG). The German Academic Exchange Service (DAAD) and the MSFHR supported P.F.H. U.E. and A.D. were supported by a post-doctoral fellowship from MSFHR, and C.L.B. was supported by postdoctoral fellowships of the Swiss National Science Foundation and the Novartis Jubilee Foundation. This work was supported by project grants from CIHR (MOP-11433, MOP-37937, and MOP-111055), and infrastructure grants from both the MSFHR and the Canada Foundations for Innovation (CFI). We thank the Pride Team (<http://www.ebi.ac.uk/services/teams/pride>), especially Tobias Ternent and Attila Csordas, for excellent assistance with MS data deposition, Jason Rogalski, Wei Chen, and Suzanne Perry from the University of British Columbia Proteomics Core Facility (PCF) and the UBC Centre for Blood Research Mass Spectrometry Suite for LC-MS/MS measurements, and all current and former members of the Overall Lab for their continuous support and inspiring discussions.

## 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.2016.02.036>.

## References

- [1] U. Eckhard, P.F. Huesgen, O. Schilling, C.L. Bellac, G.S. Butler, J.H. Cox, et al., Active site specificity of the matrix metalloproteinase family: proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses, *Matrix Biol.: J. Int. Soc. Matrix Biol.* (2015), <http://dx.doi.org/10.1016/j.matbio.2015.09.003>.
- [2] 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>.

- [3] C.J. Morrison, C.M. Overall, TIMP independence of matrix metalloproteinase (MMP)-2 activation by membrane type 2 (MT2)-MMP is determined by contributions of both the MT2-MMP catalytic and hemopexin C domains, *J. Biol. Chem.* 281 (2006) 26528–26539. <http://dx.doi.org/10.1074/jbc.M603331200>.
- [4] G.S. Butler, E.M. Tam, C.M. Overall, The canonical methionine 392 of matrix metalloproteinase 2 (gelatinase A) is not required for catalytic efficiency or structural integrity: probing the role of the methionine-turn in the metzincin metalloprotease superfamily, *J. Biol. Chem.* 279 (2004) 15615–15620. <http://dx.doi.org/10.1074/jbc.M312727200>.
- [5] G.R. Pelman, C.J. Morrison, C.M. Overall, Pivotal molecular determinants of peptidic and collagen triple helicase activities reside in the S3' subsite of matrix metalloproteinase 8 (MMP-8): the role of hydrogen bonding potential of ASN188 and TYR189 and the connecting cis bond, *J. Biol. Chem.* 280 (2005) 2370–2377. <http://dx.doi.org/10.1074/jbc.M409603200>.
- [6] C. Morrison, S. Mancini, J. Cipollone, R. Kappelhoff, C. Roskelley, C. Overall, Microarray and proteomic analysis of breast cancer cell and osteoblast co-cultures: role of osteoblast matrix metalloproteinase (MMP)-13 in bone metastasis, *J. Biol. Chem.* 286 (2011) 34271–34285. <http://dx.doi.org/10.1074/jbc.M111.222513>.
- [7] Z. Wang, R. Juttermann, P.D. Soloway, TIMP-2 is required for efficient activation of proMMP-2 in vivo, *J. Biol. Chem.* 275 (2000) 26411–26415. <http://dx.doi.org/10.1074/jbc.M001270200>.
- [8] 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>.
- [9] 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>.
- [10] 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, *Proteom. Clin. Appl.* (2015). <http://dx.doi.org/10.1002/prca.201400164>.
- [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] 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>.
- [13] 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](http://dx.doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2).
- [14] 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>.
- [15] 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>.
- [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] 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. Proteom.* (2011). <http://dx.doi.org/10.1074/mcp.M111.007690>.
- [18] M. Vaudel, J.M. Burkhardt, 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>.
- [19] J.E. Fuchs, S. von Grafenstein, R.G. Huber, M.A. Margreiter, G.M. Spitzer, H.G. Wallnoefer, et al., Cleavage entropy as quantitative measure of protease specificity, *Plos Comput. Biol.* 9 (2013) e1003007. <http://dx.doi.org/10.1371/journal.pcbi.1003007>.
- [20] J.J. Perona, C.S. Craik, Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold, *J. Biol. Chem.* 272 (1997) 29987–29990.
- [21] B.I. Ratnikov, P. Cieplak, K. Gramatikoff, J. Pierce, A. Eroshkin, Y. Igarashi, et al., Basis for substrate recognition and distinction by matrix metalloproteinases, *Proc. Natl. Acad. Sci. USA* 111 (2014) E4148–E4155. <http://dx.doi.org/10.1073/pnas.1406134111>.
- [22] J.E. Fuchs, S. von Grafenstein, R.G. Huber, C. Kramer, K.R. Liedl, Substrate-driven mapping of the degradome by comparison of sequence logos, *Plos Comput. Biol.* 9 (2013) e1003353. <http://dx.doi.org/10.1371/journal.pcbi.1003353>.
- [23] UniProt Consortium, UniProt: a hub for protein information, *Nucleic Acids Res.* 43 (2015) D204–D212. <http://dx.doi.org/10.1093/nar/gku989>.
- [24] R.D. Finn, A. Bateman, J. Clements, P. Coghill, R.Y. Eberhardt, S.R. Eddy, et al., Pfam: the protein families database, *Nucleic Acids Res.* 42 (2014) D222–D230. <http://dx.doi.org/10.1093/nar/gkt1223>.
- [25] N.D. Rawlings, M. Waller, A.J. Barrett, A. Bateman, MEROPS: the database of proteolytic enzymes, their substrates and inhibitors, *Nucleic Acids Res.* 42 (2014) D503–D509. <http://dx.doi.org/10.1093/nar/gkt953>.
- [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] 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>.
- [29] N. Colaert, K. Helens, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo, *Nat. Methods* 6 (2009) 786–787. <http://dx.doi.org/10.1038/nmeth.1109-786>.
- [30] C.G. Knight, F. Willenbrock, G. Murphy, A novel coumarin-labelled peptide for sensitive continuous assays of the matrix metalloproteinases, *FEBS Lett.* 296 (1992) 263–266.

- [31] 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. Proteom.* 100 (2014) 102–114. <http://dx.doi.org/10.1016/j.jprot.2013.10.004>.
- [32] G. Marino, P.F. Huesgen, U. Eckhard, C.M. Overall, W.P. Schröder, C. Funk, Family-wide characterization of matrix metalloproteinases 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>.
- [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] 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>.
- [35] A. Zelanis, P.F. Huesgen, A.K. Oliveira, A.K. Tashima, S.M.T. Serrano, C.M. Overall, Snake venom serine proteinases specificity mapping by proteomic identification of cleavage sites, *J. Proteom.* 113 (2015) 260–267. <http://dx.doi.org/10.1016/j.jprot.2014.10.002>.
- [36] 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>.
- [37] 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>.
- [38] C.H. Wilson, D. Indarto, A. Doucet, L.D. Pogson, M.R. Pitman, K. McNicholas, et al., Identifying natural substrates for dipeptidyl peptidases 8 and 9 using terminal amine isotopic labeling of substrates (TAILS) reveals *in vivo* roles in cellular homeostasis and energy metabolism, *J. Biol. Chem.* 288 (2013) 13936–13949. <http://dx.doi.org/10.1074/jbc.M112.445841>.
- [39] 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. Proteom.* 9 (2010) 894–911. <http://dx.doi.org/10.1074/mcp.M000050-MCP201>.
- [40] A.E. Starr, C.L. Bellac, A. Dufour, V. Goebeler, C.M. Overall, Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities, *J. Biol. Chem.* 287 (2012) 13382–13395. <http://dx.doi.org/10.1074/jbc.M111.314179>.
- [41] C. Becker-Pauly, O. Barré, O. Schilling, U. auf dem Keller, A. Ohler, C. Broder, et al., Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates, *Mol. Cell. Proteom.* (2011). <http://dx.doi.org/10.1074/mcp.M111.009233>.
- [42] 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.* 6 (2013) rs2.. <http://dx.doi.org/10.1126/scisignal.2003512>.
- [43] P.F. Lange, P.F. Huesgen, K. Nguyen, C.M. Overall, Annotating N termini for the human proteome project: N termini and N $\alpha$ -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>.
- [44] 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>.
- [45] 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* 124 (2014) e49–e60. <http://dx.doi.org/10.1182/blood-2014-04-569640>.
- [46] U. Eckhard, G. Marino, S.R. Abbey, G. Tharmarajah, I. Matthew, C.M. Overall, The human dental pulp proteome and N-terminome: leveraging 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>.
- [47] U. Eckhard, G. Marino, S.R. Abbey, I. Matthew, C.M. Overall, TAILS N-terminomic and proteomic datasets of healthy human dental pulp, *Data Brief.* 5 (2015) 542–548. <http://dx.doi.org/10.1016/j.dib.2015.10.003>.
- [48] 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.
- [49] 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>.
- [50] M. Abel, K. Iversen, A. Planas, U. Christensen, Pre-steady-state kinetics of *Bacillus licheniformis* 1,3-1,4-beta-glucanase: evidence for a regulatory binding site, *Biochem. J.* 371 (2003) 997–1003. <http://dx.doi.org/10.1042/BJ20021504>.
- [51] J.A. Vizcaino, R.G. Côté, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, et al., The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013, *Nucleic Acids Res.* 41 (2013) D1063–1069. <http://dx.doi.org/10.1093/nar/gks1262>.