Active site specificity profiling of the matrix metalloproteinase family: Proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses

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

*1 - **Centre for Blood Research,** Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, BC, Canada*
*2 - **Present address: Central Institute for Engineering,** Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, Germany*
*3 - **Present address: Institute of Molecular Medicine and Cell Research,** University of Freiburg, Freiburg, Germany*
*4 - **Present address: Swissmedic,** Swiss Agency for Therapeutic Products, Bern, Switzerland*
*5 - **Present address: Inception Sciences,** Vancouver, BC, Canada*
*6 - **Present address: Institute of Molecular Health Sciences,** ETH Zürich, Zürich, Switzerland*
*7 - **Present address: Department of Pathology,** University of British Columbia, Vancouver, BC, Canada*
*8 - **Present address: Department of Medical Genetics,** University of British Columbia, Vancouver, BC, Canada*
*9 - **Present address: Department of Bioquímica y Biología Molecular,** Universidad de Oviedo, Oviedo, Spain*
*10 - **Present address: Ottawa Institute of Systems Biology,** University of Ottawa, Canada*
*11 - **Department of Biochemistry and Molecular Biology,** University of British Columbia, Vancouver, BC, Canada*

***Correspondence to Christopher M. Overall:** Centre for Blood Research, University of British Columbia, 4.401 Life Sciences Centre, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada. chris.overall@ubc.ca*
http://dx.doi.org/10.1016/j.matbio.2015.09.003
Edited by R. Iozzo

Abstract

Secreted and membrane tethered matrix metalloproteinases (MMPs) are key homeostatic proteases regulating the extracellular signaling and structural matrix environment of cells and tissues. For drug targeting of proteases, selectivity for individual molecules is highly desired and can be met by high yield active site specificity profiling. Using the high throughput Proteomic Identification of protease Cleavage Sites (PICS) method to simultaneously profile both the prime and non-prime sides of the cleavage sites of nine human MMPs, we identified more than 4300 cleavages from P6 to P6′ in biologically diverse human peptide libraries. MMP specificity and kinetic efficiency were mainly guided by aliphatic and aromatic residues in P1′ (with a ~32–93% preference for leucine depending on the MMP), and basic and small residues in P2′ and P3′, respectively. A wide differential preference for the hallmark P3 proline was found between MMPs ranging from 15 to 46%, yet when combined in the same peptide with the universally preferred P1′ leucine, an unexpected negative cooperativity emerged. This was not observed in previous studies, probably due to the paucity of approaches that profile both the prime and non-prime sides together, and the masking of subsite cooperativity effects by global heat maps and iceLogos. These caveats make it critical to check for these biologically highly important effects by fixing all 20 amino acids one-by-one in the respective subsites and thorough assessing of the inferred specificity logo changes. Indeed an analysis of bona fide MEROPS physiological substrate cleavage data revealed that of the 37 natural substrates with either a P3-Pro or a P1′-Leu only 5 shared both features, confirming the PICS data. Upon probing with several new quenched-fluorescent peptides, rationally designed on our specificity data, the negative cooperativity was explained by reduced non-prime side flexibility constraining accommodation of the rigidifying P3 proline with leucine locked in S1′. Similar negative cooperativity between P3 proline and the novel preference for asparagine in P1 cements our conclusion that non-prime side flexibility greatly impacts MMP binding affinity and cleavage efficiency. Thus, unexpected sequence cooperativity consequences were revealed by PICS that uniquely

encompasses both the non-prime and prime sides flanking the proteomic-pinpointed scissile bond.

Introduction

Matrix metalloproteinases (MMPs) are a 23-member family of zinc dependent and calcium containing secreted endopeptidases in humans belonging to the M(A) clan of metallopeptidases and the metzincin subclan of proteases [1–5] (Fig. 1A). Initially described as degraders of all ~300 extracellular matrix proteins [6,7], in numerous physiological and pathological processes [8–11], but without experimental evidence for cleavage of most of these proteins, discovery of indirect signaling roles was delayed as a consequence of their name. MMPs liberate matrix protein-bound cytokines and growth factors (e.g. VEGF and TGFβ) from the extracellular matrix [12] and release bioactive cryptic peptides upon cleavage of ECM proteins such as laminin-5 and collagen IV [13–15], or matricellular proteins such as osteopontin or SPARC [16,17]. Excellent reviews putting the spotlight on these bioactive peptides called matrikines and/or matricryptins have been published recently [18–20]. Following deeper analysis of the MMP substrate repertoire, the substrate degradome, widespread direct signaling roles by precision proteolytic cleavage of diverse bioactive molecules in innate immunity, tissue homeostasis, and pathology, is now recognized [21–23]. Thus, proteolytic processing by MMPs, either directly via activating or inactivating cleavage or indirectly due to cleavage of interacting or inhibitory proteins [24,25] also regulates multiple other proteases in the protease web. Prominent examples include: (i) upregulation of the complement cascade and interferon-α by MMP2 [26] and MMP12 [111], respectively, (ii) downregulation of complement by MMP12 [27], (iii) shedding of membrane anchored HB-EGF by MMP3 and TNF-α by MMP7 [28,29], (iv) precise processing of stromal cell-derived factor-α by MMP2, switching the CXCR4 ligand into a highly neurotoxic CXCR3 binding protein [30], and (v) pervasive regulation of leukocytes by chemokine processing [31,32]. Thus, through precise proteolytic processing, MMPs have pivotal roles in the regulation of a myriad of biological pathways, both in homeostasis and disease. Excessive MMP activation is associated with multiple diseases yet elevated MMP activity is not necessarily detrimental, as MMP-mediated degradation of the ECM can result in reduced fibrosis [33], and MMPs 3, 8, 9, and 12 are protective in murine cancer models [34,35]. Further examples highlighting the importance of ECM remodeling and MMPs in development and disease can be found in recent reviews [36–39], yet the higher order control of cells by remodeling of signaling networks is suggested as the more dominant role *in vivo* [21–23].

Making up the MMP family are three collagenases (MMPs 1, 8, and 13), two gelatinases (MMPs 2 and 9), three stromelysins (MMPs 3, 10, and 11), two matrilysins (MMPs 7 and 26), six membrane-type MMPs, subdivided into GPI-anchored (MMPs 17 and 25) and transmembrane-type MMPs (MMPs 14, 15, 16, 24), with MMPs 12, 19, 20, 21, 23, 27, and 28 grouped together as 'other' MMPs [2,3,40]. Exclusive to the two gelatinases are three fibronectin type-II-like repeats, inserted in tandem just before the active site (Fig. 1B), facilitating gelatin and collagen binding [41–43]. Common to all MMPs is an extended zincin motif HExGHxxGxxH (Supplementary Fig. 1) and the "hydrophobic basement"-forming and name-giving Met-turn ("Metzincin"), which is positioned underneath the catalytic zinc. Three histidine residues and a water molecule coordinate the catalytic zinc in the active site crevice, which divides the catalytic domain when viewed in standard orientation into an N-terminal upper and a C-terminal lower subdomain [44]. Upon substrate binding, the carbonyl group of the scissile peptide bond coordinates with the zinc atom thus becoming strongly polarized. The glutamate of the active site motif acts as a general base, pulling a proton from the zinc-activated nucleophilic water, turning it into OH^-, which subsequently attacks and directly hydrolyzes the amide bond of the bound substrate [1,2,4].

Knowledge of protease cleavage site specificity is not only fundamental to the understanding of the dynamic interactions between proteolytic enzymes and their substrates, but is also a prerequisite for the design of sensitive and specific activity assays and the development of inhibitory drugs [45–47]. With liquid chromatography–tandem mass-spectrometry (LC–MS/MS) being among the most sensitive methods to resolve and detect complex peptide mixtures, the combination of peptide libraries with mass spectrometry was the rational next step for accurate active site protease specificity profiling. This approach suffered from the low success of *de novo* sequencing of highly redundant synthetic peptide sequences; however, this problem was solved using peptide libraries generated from fully sequenced species that enabled use of the conventional proteomic search engines for peptide spectral matching [48,49]. This improvement underpins the high throughput Proteomic

Identification of protease Cleavage Sites (PICS) method, which uniquely determines both the prime and non-prime cleavage site preferences in a single experiment: prime side sequences are identified by LC–MS/MS and the corresponding non-prime side sequences are derived bioinformatically by database searches. PICS uses diverse, proteome-derived, and database-searchable peptide libraries generated by highly specific endoproteinases, typically trypsin, chymotrypsin, or *Staphylococcus aureus* protease V8 (GluC), allowing highly sensitive and unbiased active site specificity profiling using these denatured peptides [48,49]. For in depth profiling, complementary libraries differing in their C-terminal amino acid residues are used, as for example GluC-generated peptide libraries lack internal acidic residues, and tryptic PICS libraries are devoid of basic residues. Thus, PICS allows active site specificity profiling without any prior knowledge of the physiological role, structure, or sequence preference of the tested protease. Alignment of hundreds of identified individual cleavage sites subsequently provides an unbiased and detailed picture of the amino acid sequences flanking the scissile peptide bonds, thus allowing the generation of a robust proteolytic signature of numerous proteases [50–54]. Unlike other approaches profiling just the prime or non-prime side, or using positional scanning, spatial information is retained in the cleaved peptides of a PICS experiment enabling detailed cooperativity analyses to be performed. Like all denatured peptide library-based approaches, PICS data cannot be reliably used for native substrate identification and ignores exosite interactions on distant or proximal domains due to constraints of peptide length used in the assays. Rather its utility resides in active site profiling in a unique manner from P6 to P6′ that can be used for assay development and drug targeting.

Matrix metalloproteinases specificity is guided by positional preferences, both N-terminal (non--prime side) and C-terminal (prime side) to the scissile peptide bond. Hence PICS is well suited for an unbiased comparison of specificity determinants within the MMP family. In this study, we present a family-wide specificity portrait of nine human matrix metalloproteinases determined by PICS representing the largest specificity profile for any protease family to date spanning P6 to P6′. We assayed all secreted collagenases (MMPs 1, 8, and 13), both gelatinases (MMPs 2 and 9), one representative member of the stromelysins (MMP3) and matrilysins (MMP7), membrane-type MMPs (MMP14), and macrophage MMP12, to obtain a comprehensive picture of their active site preferences and animosities, and corroborated the results with in-depth structural and synthetic peptide cleavage analysis.

Results

Substrate specificity profiling using PICS

We identified more than 2800 cleavage sites using tryptic libraries and over 1500 using GluC-generated peptide libraries, giving a median of 439 identified cleavages for each MMP analyzed (Fig. 1C; Supplementary Tables 1 and 2). A minimum of 128 (MMP8; GluC-generated peptide library) and a maximum of 699 (MMP2; trypsin-generated peptide library) cleavages were identified. Importantly, trypsin-generated peptide libraries typically yield 2–3 times higher cleavage numbers since semi-tryptic peptides with their C-terminal basic residue are more amenable to mass spectrometry than GluC-generated peptides. Note, the number of identified cleavages does not necessarily correlate with enzyme activity but rather reflects instrument performance on the day of MS-analysis. Global specificity features shared by most MMPs were determined as well as distinguishing features characterizing individual MMPs or MMP sub-groups. We summarized their individual subsite specificities as heat maps and iceLogos in Figs. 2 and 3 (trypsin-generated PICS libraries) and Supplementary Figs. 2 and 3 (GluC libraries). Overall, all MMPs showed comparably similar specificity profiles with some subsites being more important or discriminating than others. Notably, specificity was mainly guided by aliphatic and aromatic residues in P1′ followed by a pronounced preference for proline in P3, the two well recognized specificity features of MMPs [2,55]. Although all tested MMPs adhered to this general specificity profile (Figs. 4 and 5), there are numerous features that distinguish individual MMPs, which were further elaborated using a panel of quenched fluorescence peptides (Fig. 6). Below, we discuss both shared and distinguishing MMP specificity determinants for the positions P3 to P3′. With the exception of acidic and basic residues, which can be only investigated in tryptic or GluC-generated libraries, respectively, we focused on features present within both libraries that were thus individually validated by separate library classes. Of note, as can be seen by the agreement for all MMPs of the individual specificity determination heat maps using tryptic (Figs. 2 and 3, Supplementary Table 1) and GluC (Supplementary Figs. 2 and 3, Supplementary Table 2) libraries, this data concordance highlights the robustness and reproducibility of PICS.

Subsite P3

Although the P3 preference for proline is a hallmark feature of the MMP family, we found a wide differential preference for proline. MMPs 9 and 13 displayed the highest prevalence of P3 proline (46 and 35%, respectively). On the other hand, MMPs 7 (15%), 12 (17%) and 14 (15%) displayed a comparably low

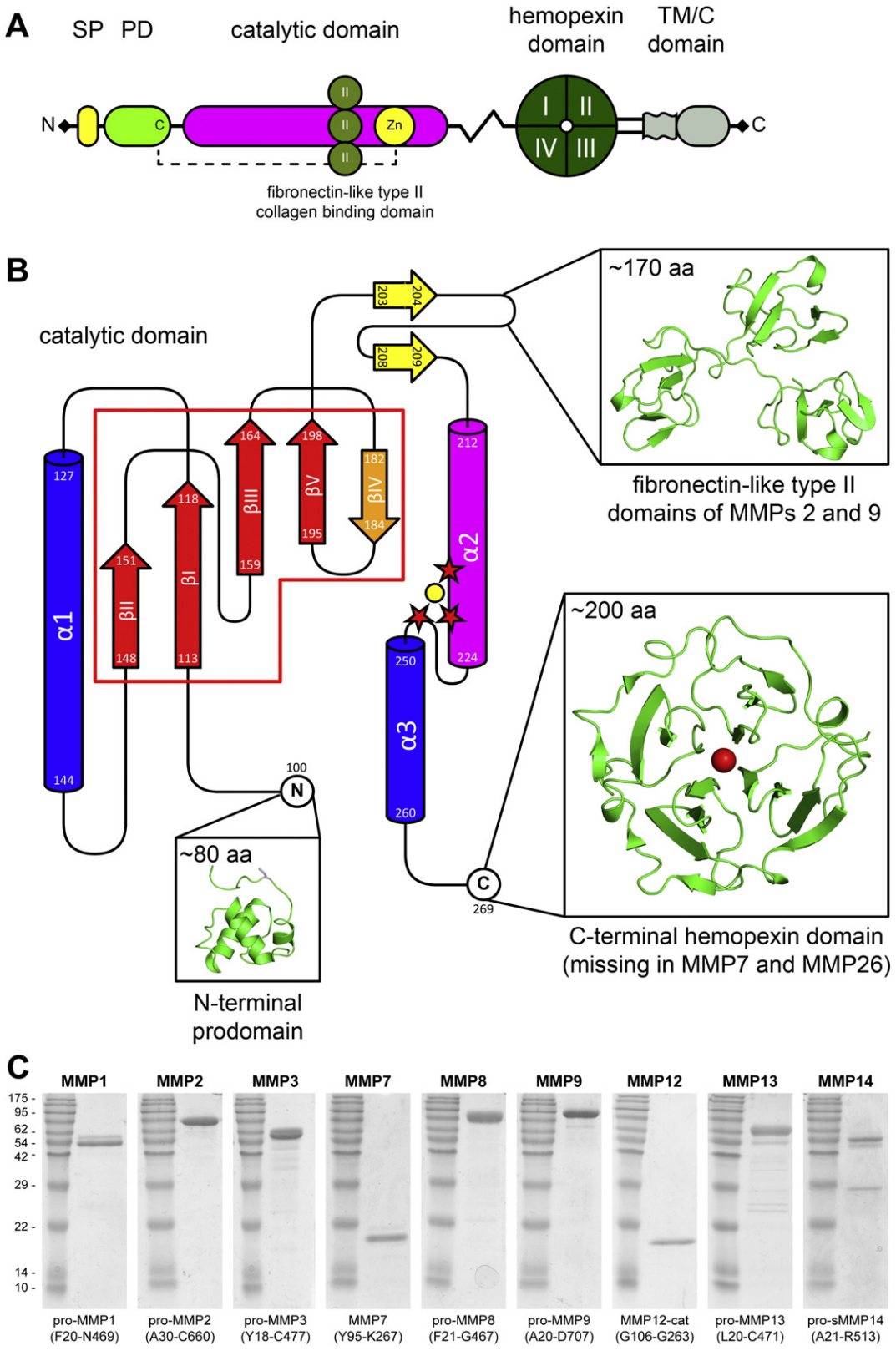


Fig. 1 (*legend on next page*)

proline preference as assayed by both tryptic and GluC PICS libraries. Most MMPs also favored small to large aliphatic residues in P3, such as alanine, valine, isoleucine, and leucine, with a combined frequency ranging from 32% (MMP9) to 46% (MMP3).

Subsite P2

Specificity profiles that we obtained were fairly diverse at P2 within the MMP family. For most MMPs, the presence of acidic residues in P2 co-occurred with an increased presence of proline in P3, an effect that was also observed for basic residues in MMPs 8 and 14. Most, but not all, MMPs displayed a certain preference for small residues such as alanine or glycine, most prominent for MMPs 2 and 3 (>34%), whereas MMP1 (14%) and MMP8 (12%) selected larger residues, such as leucine, phenylalanine, tyrosine and glutamine (>36%). For 7 out of the 9 MMPs we detected noticeable P2 selectivity for charged residues: Whereas MMPs 1, 7, 8, 12 and 14 possessed partial preference for acidic but not basic residues, MMPs 9 and 13 selected positively charged residues as noted before [56–59]. Interestingly, preferences towards acidic and basic residues in P2 seemed to be not necessarily mutually exclusive since several MMPs, most prominently MMP12, readily accommodated both amino acid classes. However, MMPs with a strong preference for small residues (MMPs 2 and 3) did not, or could not, show a concurrent preference for charged residues; these two features appeared to be mutually exclusive, but supposedly more due to the size than the physicochemical properties of charged amino acids. The selectivity of MMP9 for basic residues in P2 is linked to an aspartic residue (Asp410) in proximity to S2 [56], likely forming a salt bridge; its gelatinase counterpart MMP2 possesses a bulkier glutamic residue at the structurally equivalent position, narrowing the subsite and inducing the preference for small residues in MMP2 [56]. A similar effect was seen for MMP3, where the structurally equivalent residue Phe227 fills up nearly the entire subsite, and thus restricts it to small aliphatic residues (see Figs. 7 and 8).

Subsite P1

We identified a two to four-fold enrichment for asparagine in P1 for all tested MMPs, a preference previously only described for MMP2 [48]. Additional preferences included small amino acids such as glycine, alanine or serine, which was particularly strong in MMPs 2, 9, 13 and 14, whereas aromatic and larger aliphatic residues were barely observed in P1. A modest preference for acidic residues was identified for several MMPs, most frequently in MMPs 3, 7, 8 and 12, so logically reducing their preference for small amino acids here. The opposite was found for MMPs 2, 9, and 14, which showed a strong preference for small, but not for acidic amino acids. The strong preference for acidic residues in P1 by MMP3, but not MMPs 1 and 9, is corroborated by previous specificity studies [60,61], and is thought to mediate efficient MMP3 cleavage of substrates not effectively targeted by other MMPs [62]. Interestingly, preferences for basic and acidic residues are once again not mutually exclusive, but occur alongside each other, most apparent in the profiles obtained for MMP7, implying hydrogen bonding rather than salt bridging.

Subsite P1′

P1′ leucine clearly constituted the primary specificity determinant for all MMPs probed in this study. Together with isoleucine, and to a lesser extent valine and methionine, a combined occurrence of 60% (MMP13) to 93% (MMP7) was observed. In good agreement with previous studies [61], MMP7 was differentiated from the other MMPs by displaying the highest leucine preference (62%), whereas MMPs 3, 9, 12 and 13 showed the lowest frequency (32–33%). Some MMPs also displayed a modest preference for large aromatic residues, prominently seen for MMP8 and to a slightly lesser extent for MMPs 3, 12, 13, and 14. This feature is confirmed by a study using mixture-based oriented peptide libraries [55] showing a moderate preference for aromatic residues, that, though not exceeding the natural abundance by more than 2.5 fold, still indicates a comparably deeper S1′ pocket. In PICS, MMPs 8 and

Fig. 1. (A) Schematic representation of the protein structures and arrangements of the MMP family. SP signal peptide, PD prodomain, TM/C transmembrane/cytoplasmic domain. The minimal MMPs 7 and 26 have an N-terminal signal peptide, prodomain, whereas archetypical MMPs (e.g. MMP1) also carry a four-bladed hemopexin C domain attached to the catalytic domain by a linker, and membrane-type MMPs also have a stalk domain and either a transmembrane and cytoplasmic domain (e.g. MMP14) or a GPI-anchoring sequence (e.g. MMP17) at their C-terminus. The gelatinases MMP2 and 9 harbor three supplementary collagen binding fibronectin-like type II domains inserted in the catalytic domain. (B) Topology diagram of an MMP catalytic domain with sequence numbering according to MMP1 (UniProt ID: P03956), complemented by ribbon representations of the prodomain, the fibronectin-like type II domains, and the hemopexin C domain. (C) Coomassie brilliant blue-stained SDS-PAGE analysis (15% gels) of all MMPs used in this study. 1–2 μg of each MMP were loaded together with PiNK plus prestained protein ladder (GeneDireX). Doublets visible around 50 to 60 kDa for MMP1, MMP13 and MMP14 result from heterogeneous glycosylation, and for soluble MMP14 (sMMP14) the Hpx-domain truncated isoform was also detected (band at approx. 28 kDa). Of note, to convert the UniProtKB sequence numbering used throughout the manuscript (starting at Met1) to traditional numbering of secreted proMMPs, please use the following off-sets to account for loss of the respective signal peptide: MMP1: -21, MMP2: -29, MMP3: -17, MMP7: -17, MMP8: -20, MMP9: -19, MMP12: -15, MMP13: -19, and MMP14: -28.

13 showed more than 2.3-fold enrichment for tyrosine and tryptophan, respectively, whereas MMPs 12 and 14 enriched for both phenylalanine and tryptophan to a similar extent. This limited preference is also supported by other specificity studies using synthetic peptides [63,64]. Notably, our study also included two MMPs annotated as having small S1′ pockets, namely MMPs 1 and 7 [65], and indeed in both cases all aromatic residues remained with a combined occurrence of 2.9 and 3.5%, clearly below their natural abundance of 7.4%. Unexpectedly, a weak P1′ preference for glutamine was observed for MMPs 9, 12 and 13, a feature entirely absent in MMPs 1 and 7, whereas MMPs 3, 9 and 13 displayed a noticeable preference for carbamidomethylated cysteine introduced during PICS library preparation, most likely due to its aliphatic property. Of note, a recently established differentiating feature between collagenases and gelatinases [66], namely a low P1′ frequency of leucine in gelatinases, could not be verified in our study.

Since P1′ leucine and P3 proline constitute the major MMP specificity determinants, we systematically probed subsite cooperativity for these two features (Supplementary Table 3): To our surprise, all MMPs displayed negative subsite cooperativity for P1′ leucine and P3 proline. For example, in the case of MMP13, the occurrence of P3 proline decreases from 34.8% (all cleaved peptides) to 20.2% (cleaved peptides with P1′ leucine) while the occurrence of P1′ leucine decreases from 33.2% (all cleaved peptides) to 19.3% (cleaved peptides with P3 proline). This effect was least pronounced in MMPs 1 and 2. Importantly, subsite cooperativity is easily detected by the PICS assay but cannot be assessed by global heat maps or iceLogos — a frequently ignored or underestimated caveat, that needs thorough individual assessment of all subsites and amino acids.

Subsite P2′

In P2′, MMP specificity was dominated by a preference for isoleucine and valine (average occurrence of 23%) together with a preference for positively charged amino acids (19%). The latter feature was detected mostly in GluC libraries since tryptic libraries have lysine and arginine only at the C-terminus of the peptide. Corroborated by a previous study [55], a distinct preference for basic residues was identified for all MMPs; whereas it was most prominent for MMP12, it appeared to be of lesser importance for MMPs 1, 2, and 8. For MMP9, a comparably low occurrence of basic residues in the GluC profile was counterbalanced by a notable overrepresentation of histidine in the tryptic profile. Of note, the arginine preference we identified in P2′ for MMP12 perfectly correlates with its unusual processing and inactivation of the ELR^+ neutrophil chemokines within the E↓LR site, which is not observed for other MMPs [67]. The preference for medium-sized aliphatic residues such as valine and isoleucine was more evident with tryptic PICS libraries, most likely as this feature is superseded by the pronounced MMP preference for basic residues in GluC libraries, illustrating the benefit of using more than one peptide library type for comprehensive active site specificity profiling.

Subsite P3′

P3′ selectivity appeared to be as inconsistent as that of P2 across the MMP family and thus could be of similar importance as P2 for the specific targeting of individual family members as previously suggested [64]. Most MMPs showed a marked preference for small amino acids such as glycine and alanine, with a combined average occurrence of 31%. This preference was most pronounced for MMPs 8 and 9 (>44%), and least for MMP14 (20%). Notably, MMPs 1 and 8 showed preferences for acidic residues, but not basic residues, whereas MMPs 3, 7, and 12, and to a lesser extent MMPs 13 and 14 selected both. As seen for P2 and P1, the preference for basic and acidic residues was not mutually exclusive. A minor preference for valine or isoleucine was identified for MMPs 1 and 14, and most MMPs readily accommodated hydrophilic residues, even though no individual MMP emerged as having a particularly strong hydrophilic preference. Finally, the sequence diversity identified in this study is in strong agreement with the structural description of the S3′ site as being of both hydrophobic and polar in nature [65].

Shared features in the MMP specificity profile

To obtain a global view of the common features of the MMP family active site specificity profile, we created heat maps comparing the 9 MMPs at the individual P3 to P3′ positions (Fig. 4), and a composite heat map and iceLogo using a total of 1800 cleavage sites, randomly selected, 100 per MMP for each peptide library (Fig. 5A). Both representations were dominated by

Fig. 2. PICS sequence specificity profiles of MMPs 1, 2, 3, 7, and 8 using trypsin-generated human peptide libraries. Identified cleavage sites (n values shown) are summarized as heat maps showing relative occurrence (left panels) and the more incisive fold-change over the natural abundance of amino acids (middle panels) and as iceLogos, showing percent difference compared to natural amino acid abundance (right panels). P6 to P6′ subsite positions are shown on the x axes; plotted amino acids are indicated on the y axes with one-letter codes. In iceLogos, significantly over-represented amino acids are shown above the x axis, under-represented residues below the x axis, and amino acids that have not been identified are depicted in pink. For easier comparison, the same scales have been applied to all MMPs.

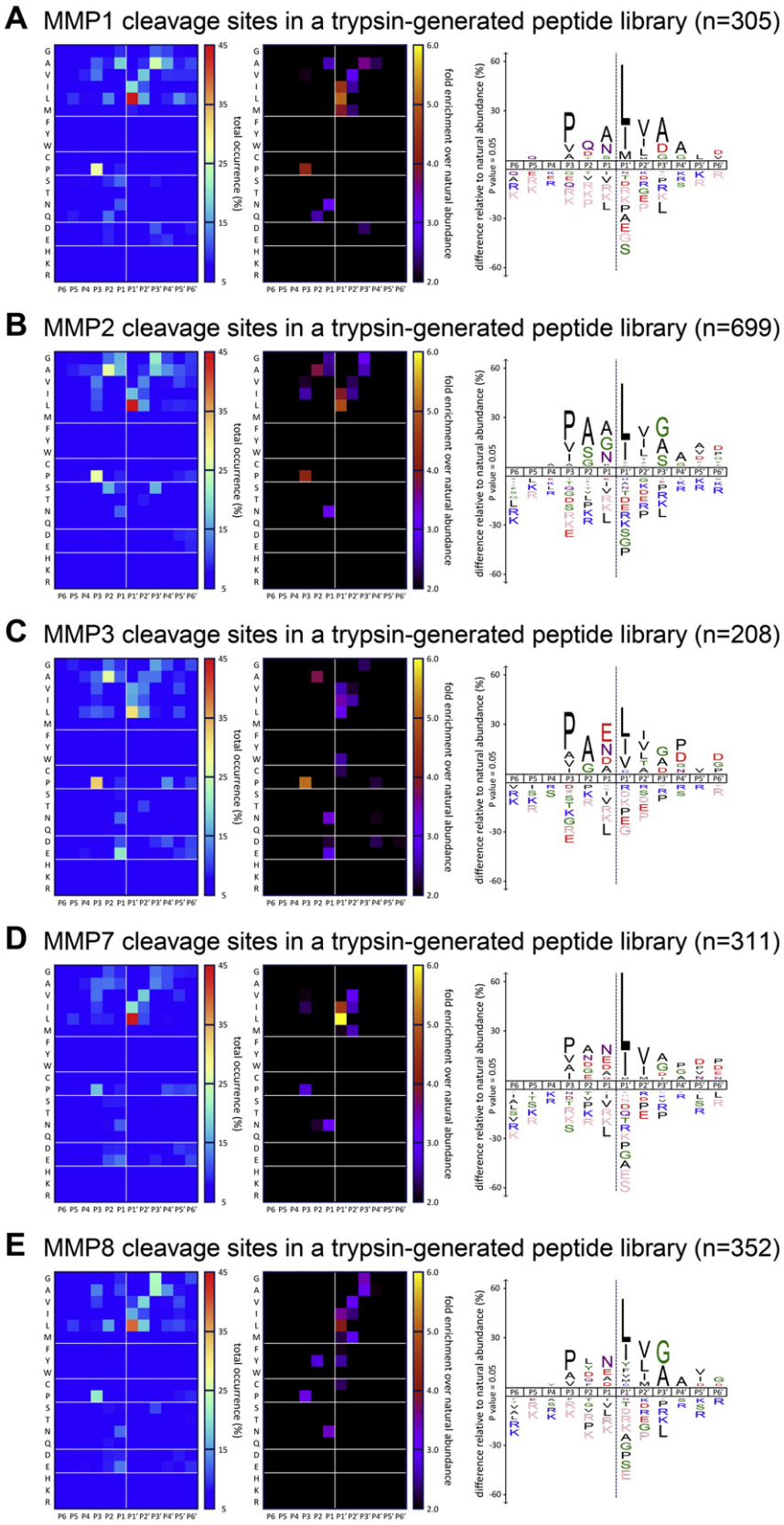


Fig. 2 (*legend on previous page*)

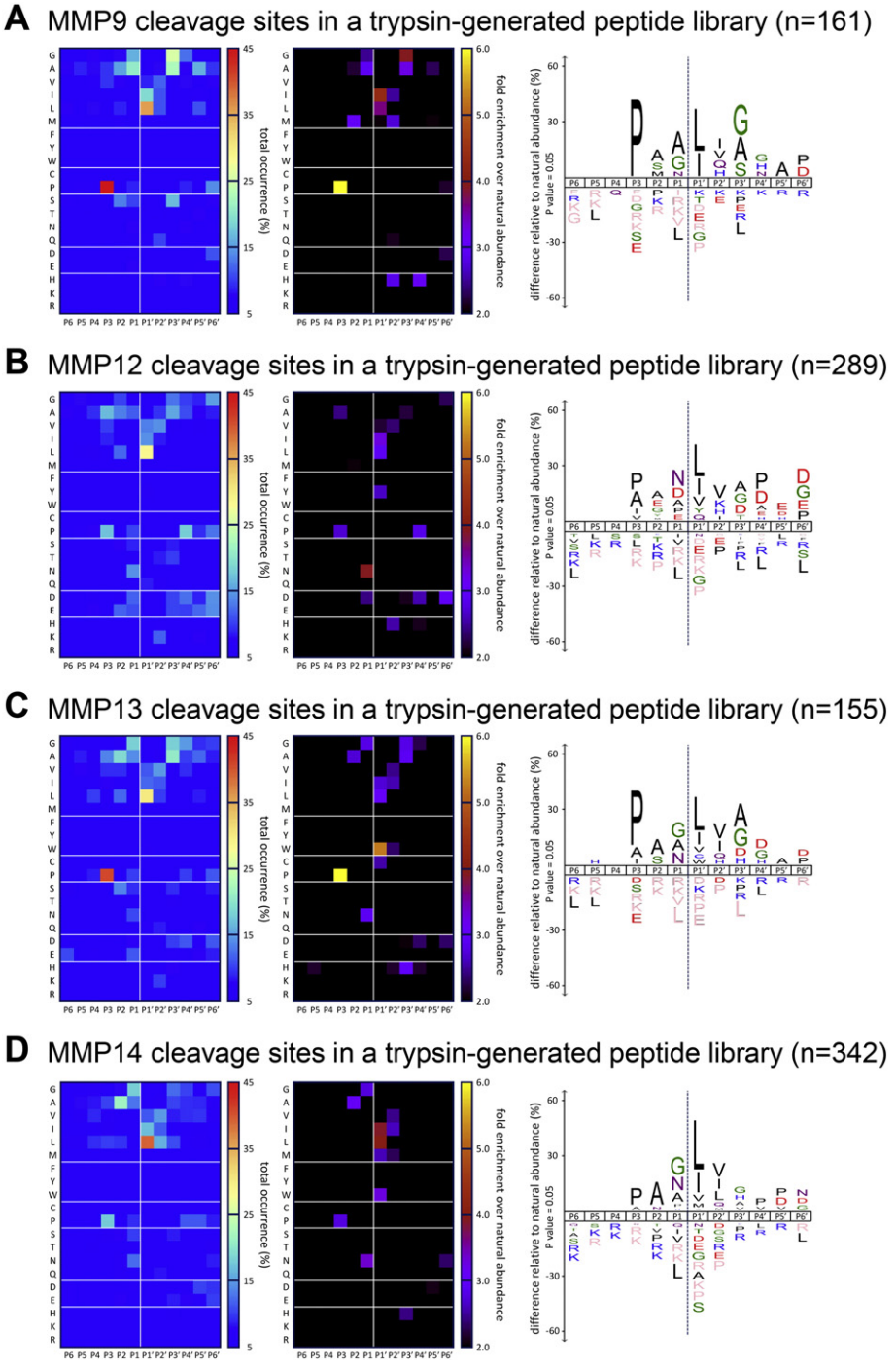


Fig. 3. PICS sequence specificity profiles of MMPs 9, 12, 13, and 14 using trypsin-generated human peptide libraries. Identified cleavage sites (n values shown) are summarized as heat maps showing relative occurrence (left panels) and the more incisive fold-change over the natural abundance of amino acids (middle panels) and as iceLogos, showing percent difference compared to natural amino acid abundance (right panels). P6 to P6′ subsite positions are shown on the x axes; plotted amino acids are indicated on the y axes with one-letter codes. In iceLogos, significantly over-represented amino acids are shown above the x axis, under-represented residues below the x axis, and amino acids that have not been identified are depicted in pink. For easier comparison, the same scales have been applied to all MMPs.

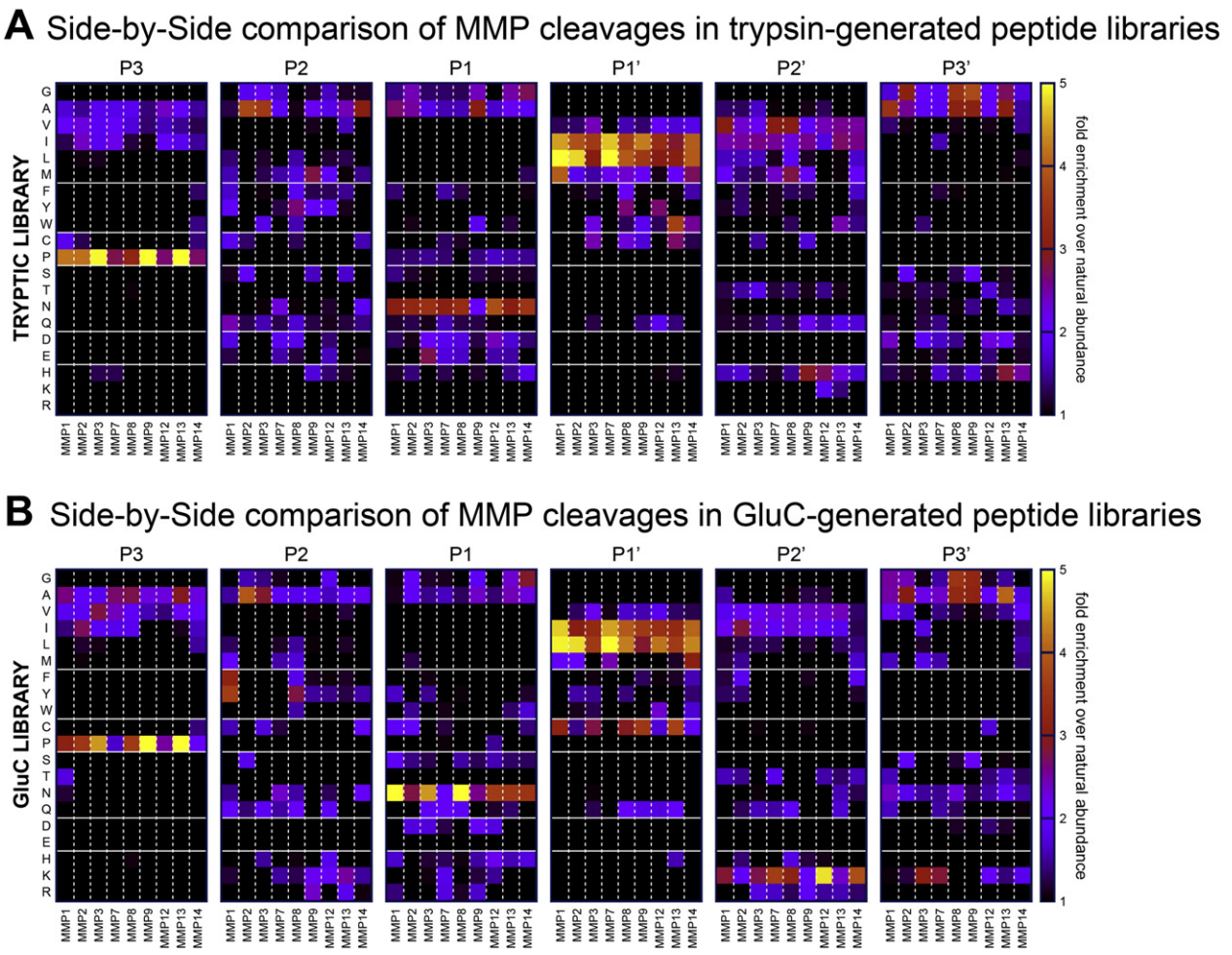
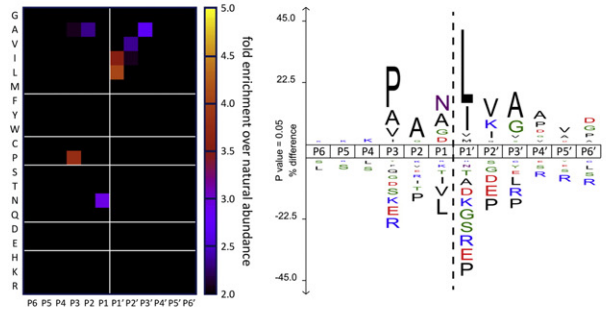


Fig. 4. Side-by-side comparison of MMP cleavages in human (A) trypsin- and (B) GluC-generated peptide libraries. Identified cleavage sites were summarized as heat maps showing fold-change over the natural abundance of amino acids for each P3 to P3′ position. The same scales have been applied to all MMPs.

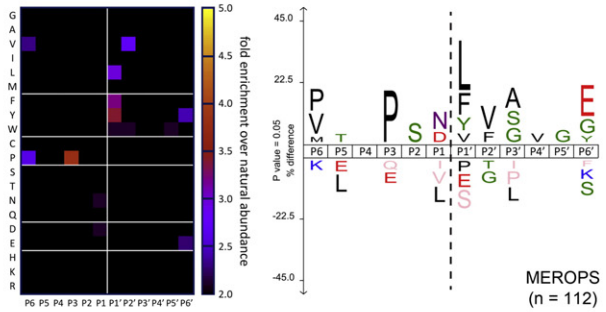
the two hallmark specificity features of MMPs, proline in P3 and Leu/Ile in P1′. Other than proline, only alanine, valine, and isoleucine were readily accommodated in P3 and showed enrichment above their natural abundance. In contrast, P2 displayed a much higher diversity, but as these individual preferences cancel each other out in the combined analysis, only alanine remained as the common denominator. Notably, Pro was nearly entirely missing in P2. Surprisingly, in P1, asparagine stood out in both libraries with an overall 3-fold enrichment, whereas valine, isoleucine, and leucine were all diminished >2-fold. All tested MMPs showed at least a 2-fold enrichment for asparagine, whereas only 4 reached this threshold for the next most prevalent P1 features. Due to the high prevalence of leucine and isoleucine in P1′, only 6 more amino acids, cysteine, glutamine, methionine, phenylalanine, tryptophan, and valine stayed above their natural abundance in the combined analysis, yet not reaching 2-fold enrichment. In position P2′, three residues emerged as common preferences: valine, isoleucine and lysine were all enriched >2-fold in 8 out of the 9 MMPs in tryptic and GluC libraries, whereas aspartate, glutamate, and proline were more than 3-fold underrepresented. Similar to P2, P3′ displayed a rather divergent picture with small hydrophobic residues such as glycine, alanine, and valine being the only mutuality with an average 2-fold enrichment.

Intrigued by the negative cooperativity between P3-Pro and P1′-Leu, we downloaded all high confidence MEROPS curated [68] native substrates of MMPs 1, 2, 3, 7, 8, 9, 12, 13 and 14; i.e. substrates with cleavage sites verified by N-terminal sequencing. These corresponded to 112 redundant (i.e. more than one MMP cleaved at the same site) and 79 non-redundant P6–P6′ cleavages (Supplementary Table 4). Analysis of the cleavage specificity profiles of the native substrates from MEROPS (Fig. 5B) was remarkably similar to the PICS proteomic analyses (Fig. 5A). Interestingly, even though half of the native substrates had either proline in P3 or leucine in P1′ (Supplementary Table 4 and Fig. 5E), only 5 (6.3%)

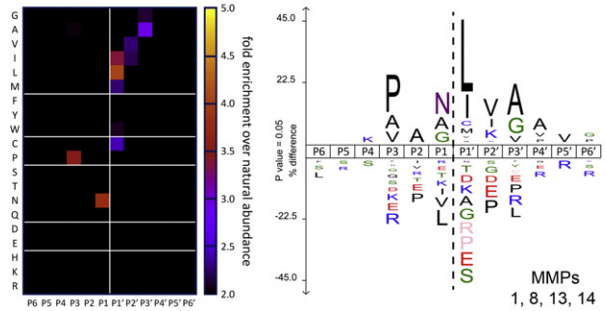
A PICS derived Pan-MMP specificity profile



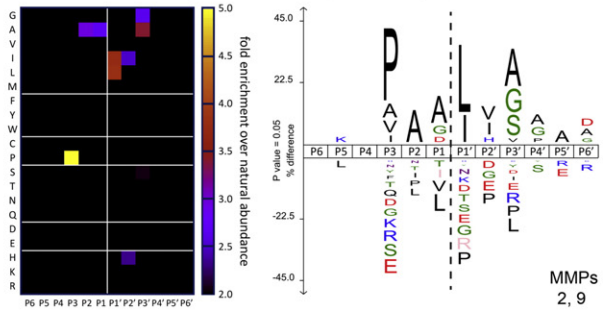
B Pan-MMP profile derived from native substrates



C PICS derived collagenase specificity profile



D PICS derived gelatinase specificity profile



E MEROPS curated physiological MMP substrates with NT-evidence and P3-Pro or P1'-Leu

MMP(s)	Substrate	Cleavage Site	P3	P2	P1	P1'	P2'	P3'
14	O35235: Tnfsf11	M145↓M	P	A	M	M	E	G
9	O35744: Chitinase-3-like protein 3	A327↓Y	P	Y	A	Y	Q	G
14	P02679: Fibrinogen gamma chain	M104↓I	P	N	M	I	D	A
8	P02751: Fibronectin	A44↓V	P	V	A	V	S	Q
2	P02751: Fibronectin	Q629↓W	P	I	Q	W	N	A
7	P17931: Galectin-3	A62↓Y	P	G	A	Y	P	G
7	P17931: Galectin-3	P76↓G	P	A	P	G	V	Y
8	P20366: Substance P	Q63↓F	P	Q	Q	F	F	G
1, 2, 8, 13	P50228: C-X-C motif chemokine 5	S44↓V	P	S	S	V	I	A
2	P63089: Pleiotrophin	N90↓W	P	C	N	W	K	K
2	P63089: Pleiotrophin	P149↓Q	P	K	P	Q	A	E
7	Q08380: Galectin-3-binding protein	D452↓Y	P	S	D	Y	R	Y
1, 2, 3, 7, 8, 9	Q29011: Aggrecan core protein	N344↓F	P	E	N	F	F	G
9	Q9WVJ5: Crystallin, beta B1	A47↓K	P	A	A	K	V	G
2	P01034: Cystatin C	R34↓L	P	P	R	L	V	G
2	P01584: Interleukin-1 beta	E141↓L	P	Y	E	L	K	A
2, 7, 9	P18827: Syndecan-1	G82↓L	P	T	G	L	E	A
14	P31431: Syndecan-4	D31↓L	P	Q	D	L	L	E
2, 7, 9	P31431: Syndecan-4	K105↓L	P	K	K	L	E	E

MMP(s)	Substrate	Cleavage Site	P3	P2	P1	P1'	P2'	P3'
7	O75596: C-type lectin domain family 3 member A	A57↓L	V	N	A	L	K	E
7	O75596: C-type lectin domain family 3 member A	A63↓L	I	Q	A	L	Q	T
7	O75596: C-type lectin domain family 3 member A	F151↓L	I	S	F	L	N	W
8	P01011: Alpha-1-antichymotrypsin	A385↓L	L	S	A	L	V	E
7	P02671: Fibrinogen alpha chain precursor	E260↓L	R	M	E	L	E	R
7	P02671: Fibrinogen alpha chain precursor	E441↓L	D	K	E	L	R	T
3	P02671: Fibrinogen alpha chain precursor	K432↓L	T	E	K	L	V	T
7	P02671: Fibrinogen alpha chain precursor	F27↓L	G	D	F	L	A	E
7, 14	P02675: Fibrinogen beta chain	D153↓L	L	K	D	L	W	Q
3, 7	P02675: Fibrinogen beta chain	L150↓L	M	Y	L	L	K	D
3, 7, 14	P02679: Fibrinogen gamma chain	T109↓L	A	A	T	L	K	S
2, 8	P02751: Fibronectin	P720↓L	F	S	P	L	V	A
7	P17931: Galectin-3	P113↓L	A	G	P	L	I	V
7	P18827: Syndecan-1	G245↓L	S	Q	G	L	L	D
8	P20366: Substance P	G66↓L	F	F	G	L	M	G
7	P37889-1: Fibulin-2	S579↓L	A	L	S	L	G	T
1, 7, 8	Q29011: Aggrecan core protein	D447↓L	S	E	D	L	V	V
7	Q6FH10: DCN protein	E273↓L	L	R	E	L	H	L

Fig. 5 (*legend on next page*)

possessed both MMP hallmark features, strongly corroborating the PICS negative cooperativity results. MMP7 cuts two of these substrates possessing both P3-Pro and P1′-Leu, both syndecans. Just as other proteases have substrate-binding exosite domains [41,69] that enhance chances of substrate cleavage, MMP7, which lacks a C-terminal hemopexin domain, binds to glycosaminoglycan side chains which likely facilitates cleavage of these otherwise less preferred sites [70].

Additionally, when we compared the collagenase specificity profiles (Fig. 5C) with the gelatinase specificity profiles (Fig. 5D), only small divergences emerged. Whereas the occurrence of P3-Pro was significantly lower in collagenases compared to in gelatinases (21.9% vs. 36.1%), the opposite was the case for P1-Asn (13.6 vs. 5.2%), but this was partly compensated by higher occurrences of aspartate, glutamate, and glutamine in the gelatinases (14.7 vs. 19.7%). Of note, even though the iceLogo representation indicated a higher P1′ promiscuity for collagenases, the combined occurrence of carbamidomethylated cysteine, methionine, valine, and glutamine (10.5 vs. 10.4%) and the occurrences of the dominating leucine (39.7 vs. 37.4%) and isoleucine (14.7 vs. 16.4%) were highly similar in both sub-groups (Fig. 5C and D).

Refining MMP specificity profiles by quenched fluorescent probes and in silico peptide docking

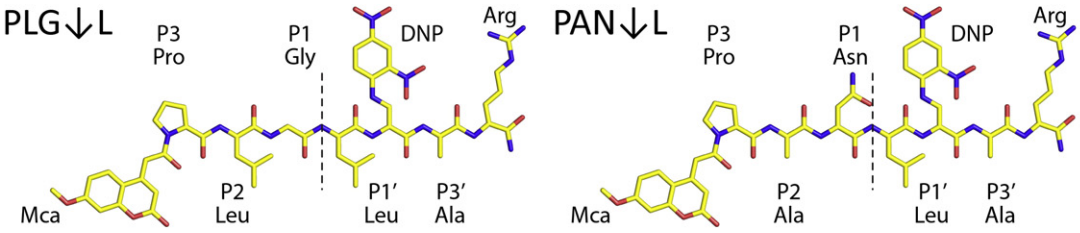
The quenched fluorescent (QF) peptide QF-24 is the classical and long-used MMP substrate, originally designed based on the loose consensus of the collagen type I cleavage site by MMPs, PLG↓L [71]. QF-24 Mca-PLG↓L-Dpa-AR-NH_2 is C-terminally amidated and carries the fluorophore (Mca) and quencher (Dpa) with a penultimate arginine for improved solubility. Based on shared features of the 9 MMPs in their specificity profiles (Figs. 4 and 5) we designed a new quenched fluorescent peptide: Mca-PAN↓L-Dpa-AR-NH_2 (Fig. 6A). To probe for the individual subsite preferences of each MMP we interchanged the respective P2 and P1 positions in PLG↓L and PAN↓L to give PLN↓L and PAG↓L in two more QF-peptides. We also tested the importance of Pro in P3 by its replacement with valine: VLG↓L and VAN↓L. All QF-peptides were C-terminally amidated, carried the same fluorophore and quencher, and were obtained from the same commercial vendor at the same purity (>95%). This consistency allowed for side-by-side comparison in quenched fluorescence cleavage assays (Fig. 6B and C). To aid data interpretation, we performed in silico docking [72] with the two heptapeptides APLG↓LLA and APAN↓LLA for all analyzed MMPs (Figs. 7 and 8). These MMPs showed similar active site geometries but with subtle differences, pronounced in the subsite S2, which is relatively shallow in the case of MMPs 3 and 13 and thus strongly selects small amino acids such as glycine, alanine, or serine, but represents a deep pocket in MMP1 and 8 allowing the accommodation of large amino acids such as tyrosine, in perfect accordance with our PICS data. Importantly, each peptide showed different cleavage rates for two or more MMPs that guided our structural interpretation of their cleavage properties as related to the specificity profile for each MMP.

Discussion

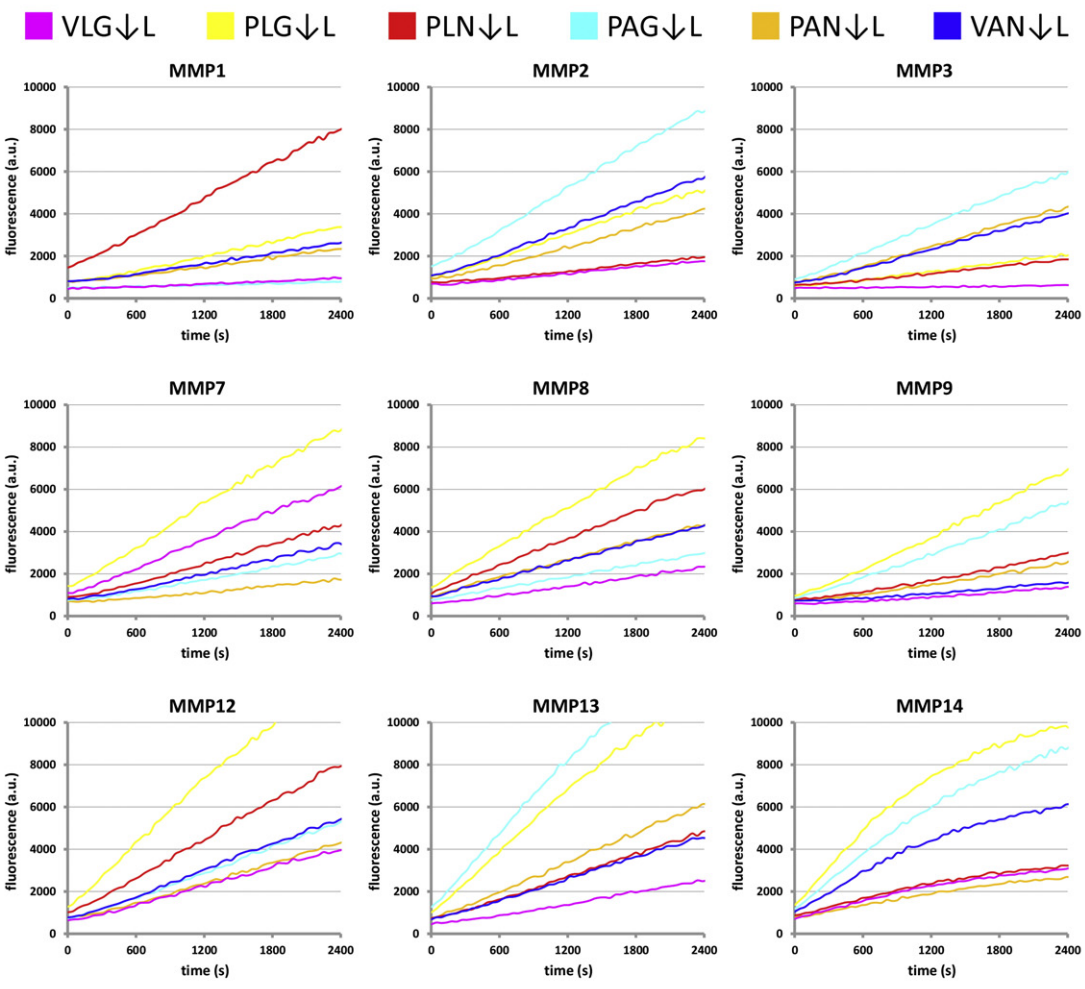
By the simultaneous identification of both the prime and non-prime active site specificity determinants that PICS uniquely profiles, our global proteomic analysis of over 4300 cleaved, biologically diverse peptides by nine important representative members of the MMP family is the most robust family-wide specificity profile reported to date for any protease family. Analysis of shared versus selective specificity features by PICS revealed unexpected negative cooperativity between the P3 proline and the P1′ leucine or isoleucine, which are archetypical hallmark preferences for most, if not all MMPs. Of note, global specificity analysis using heat maps and iceLogos does not identify subsite cooperativity effects which instead need individual assessment by fixing all 20 amino acids one-by-one in the respective subsites and thorough monitoring of the specificity logo changes. We confirmed the negative cooperativity between the P3 proline and P1′ leucine identified here with denatured biologically relevant

Fig. 5. (A) Consensus MMP cleavage site in human trypsin and GluC peptide libraries. 100 unique cleavage events were randomly selected for each MMP and library (total of 1800 sequences) and represented 1123 unique P6-to-P6′ cleavage sites. (B) Consensus MMP specificity profile identified by analyzing all native substrates with N-terminal sequencing evidence documented in MEROPS for MMPs 1, 2, 3, 7, 8, 9, 12, 13, and 14. A total of 112 cleavages were downloaded (accessed on 03/09/2015), corresponding to 79 non-redundant P6–P6′ cleavage sites. (C) Collagenase (MMPs 1, 8, 13, and 14 left) and (D) gelatinase (MMPs 2 and 9) active site specificity profiles obtained by PICS proteomics using trypsin- and GluC-generated peptide libraries. The corresponding data from (A) were used for analysis. All identified cleavage sites were summarized as a heat map showing fold-change over natural abundance and by a corresponding iceLogo displaying the percentage difference of amino acid frequencies between the experimental set and the human proteome reference set. (E) MEROPS curated physiological MMP substrates with N-terminal sequencing evidence and P3-Pro or P1′-Leu. 112 native cleavages corresponding to 79 non-redundant P6–P6′ cleavage sites were downloaded and further analyzed; of these, 23 had leucine in P1′ (29.1%) and 19 proline in P3 (24.1%), but only 5 exhibited both MMP hallmark features (6.3%), in perfect agreement with the negative cooperativity effect of P3-Pro and P1′-Leu identified by PICS. For more details, please refer to Supplementary Table 4.

A Synthetic quenched fluorescent MMP substrates



B



C

relative activity (%)	MMP1	MMP2	MMP3	MMP7	MMP8	MMP9	MMP12	MMP13	MMP14
VLG↓L	7.0 (0.5)	13.0 (2.9)	2.1 (0.1)	66.2 (12.7)	26.5 (6.5)	12.9 (0.6)	30.6 (6.1)	12.0 (0.7)	25.1 (4.8)
PLG↓L	40.2 (1.0)	54.9 (0.6)	29.2 (0.8)	100.0	100.0	100.0	100.0	80.7 (1.6)	100.0
PLN↓L	100.0	15.7 (0.4)	23.7 (0.5)	43.4 (1.9)	69.6 (6.2)	38.4 (2.7)	57.2 (1.4)	26.5 (1.1)	24.1 (2.2)
PAG↓L	4.5 (0.3)	100.0	100.0	26.0 (1.9)	30.9 (2.7)	75.1 (3.9)	38.1 (2.6)	100.0	78.1 (3.7)
PAN↓L	24.9 (2.1)	43.1 (2.6)	68.8 (3.2)	13.6 (0.5)	48.8 (4.1)	32.0 (4.2)	28.7 (0.5)	36.1 (1.2)	17.2 (0.7)
VAN↓L	28.2 (0.1)	64.6 (8.7)	64.9 (1.2)	31.9 (4.2)	53.3 (6.3)	18.6 (6.0)	41.5 (2.5)	26.6 (1.9)	44.6 (7.9)

Fig. 6 (*legend on next page*)

peptide libraries by comparison with all MEROPS-curated native MMP substrates with N-terminal sequencing evidence possessing either one or both a proline at P3 and a leucine at P1′. However, this was not noted in a very recent profiling paper [73].

Also emerging from this analysis was the novel high preference for asparagine in P1 by all analyzed MMPs. Small amino acids in P2 were common as were acidic or small aliphatic residues in P1 that contrasted with the basic and aliphatic residue preferences in P2′. Predominantly small residues in P3′ were favored despite the distance of S3′ from the scissile bond in a more open section of the active site cleft. From these analyses we rationally designed two new QF-peptide substrates that showed improved specificity for three MMPs (MMPs 1, 2 and 3) over the conventional QF-24 peptide substrate: PLN↓L is cleaved 2.5 times faster by MMP1, whereas PAG↓L is cleaved ~2-fold faster by MMP2 and 25% faster by MMP3. Thus, our proteomic analysis strongly upholds the concept that for MMPs, like other metzincins, P1′ is the defining subsite in substrates unlike P1 for most serine and cysteine proteases [63,65].

Proline and aliphatic residues that we identified as P3 specificity determinants are supported by data from mixture-based oriented peptide libraries [55] and structural studies [65] emphasizing the predominantly hydrophobic S3 pocket. A protein-centric substrate screen for MMP9 [74] and phage display profiling of MMP13 [59] also support the P3 proline preference, but a low P3 proline preference is credited as a unique MT-MMP feature [66]. Unexpectedly we found that a proline in P3 markedly reduced the occurrence of the other major specificity residue, leucine in P1′. Thus, 979 cleaved peptides with a P3 proline and 1714 with a P1′ leucine were identified, but in combination as Pxx↓L, only 295 sequences were found. Although a loss of flexibility in the scissile bond may reduce the dual preference for Pxx↓L, this is not the main reason. Indeed, of the Pxx↓L peptides cleaved, only 49 were PxG↓L with other small residues also contributing to a similar degree, e.g. 58 PxA↓L and 39 PxS↓L. However, glycine renders the non-prime side backbone more flexible to accommodate a rigidifying proline at P3 in combination with a locked-in leucine at P1′, and this dominates over the positive selective pressure for other residues that are on average more preferred at P1 than glycine. Thus, although eight of nine MMPs prefer asparagine at P1, seven showed improved cleavage kinetics for peptides with a P1 glycine and P1′ leucine. Indeed, only 138 peptides with an Asn↓Leu scissile bond were cleaved by MMPs other than MMP1. The exception, MMP1, was most unexpected in displaying a strong 2.5-fold preference for the new QF-peptide PLN↓L compared to the common QF24 (PLG↓L), which was actually designed from the MMP1 Gly↓Leu α2 (I) collagen cleavage site [71], yet is highly consistent with its PICS profile (Fig. 2 and Supplementary Fig. 2).

As expected from past studies of the general MMP substrate QF-24, PLG↓L outperformed the other new QF-peptides for 5 out of the 9 MMPs (MMPs 7, 8, 9, 12, and 14). As predicted, we found our new rationally designed peptide, PAG↓L, was kinetically preferred by the other MMPs 2, 3, and 13. Only MMP3 preferred the PAN↓L sequence to the classical PLG↓L sequence, but this is to be expected as PAN↓L matches the MMP3 preference at all four residues from the iceLogo of the pooled PICS cleavage sites. In contrast, the remaining MMPs (7, 8, 9, 12, 13 and 14) have only two or three exactly matching residues in their iceLogos to the PAN↓L residues. Thus, PICS-generated specificity logos represent potential cleavage sites rather than necessarily preferred ones. For example, none of the identified cleavage sites in PICS had the sequence PAN↓L, and even the general substrate sequence PLG↓L was only found twice in 18 experiments in total. Moreover, key examples in natural substrates show that non-preferred residues, even those requiring active site remodeling to accommodate large hydrophobic side chains, can be selected across species. An overlooked example of this disproves the universally stated notion that the shallow S1′ pocket MMPs 1 and 7 can only accommodate small residues, yet the P1′ residue of their own propeptides that are autocatalytically cleaved are phenylalanine and tyrosine, respectively. Hence, designing a universal general MMP peptide substrate from averaged specificity features of 9 MMPs is not feasible as individual features of each protease dominate and must be considered in the design of optimal cleavage assays for individual MMPs and likely too for other proteases.

Whereas PLG↓L generally outperformed PAN↓L, the corresponding peptide pair with valine in P3 showed the opposite; VAN↓L exceeded VLG↓L in all MMPs, but not MMP7, and outperformed PAN↓L in four MMPs or had no significant difference for three MMPs. Compared to 156 different peptides cleaved at N↓L, 211 were cleaved at G↓L. Although this might indicate that a more flexible scissile bond is preferred

Fig. 6. (A) Schematic representation of two quenched fluorescent (QF) peptides used in this study. Left: Mca-Pro-Leu-Gly ↓ Leu-Dpa-Ala-Arg-NH2 (PLG ↓ L), and right: Mca-Pro-Ala-Asn ↓ Leu-Dpa-Ala-Arg-NH2 (PAN ↓ L). All QF-peptides carried N-terminal Mca ([7-methoxycoumarin-4-yl]acetyl) as the fluorophore, had Dpa ([N-3-[2, 4-dinitrophenyl]-L-2,3-diaminopropionyl]) in P2′ with the quencher dinitrophenyl (DNP), and were C-terminally amidated. (B) Overall comparison of MMP activity against a set of six different QF-peptides. (C) Comparison of MMP activities normalized to the preferred peptidic MMP substrate. Experiments were performed in duplicates and repeated three times. Standard deviations are given in parenthesis and were within 10% (except for MMP7 and VLG ↓ L).

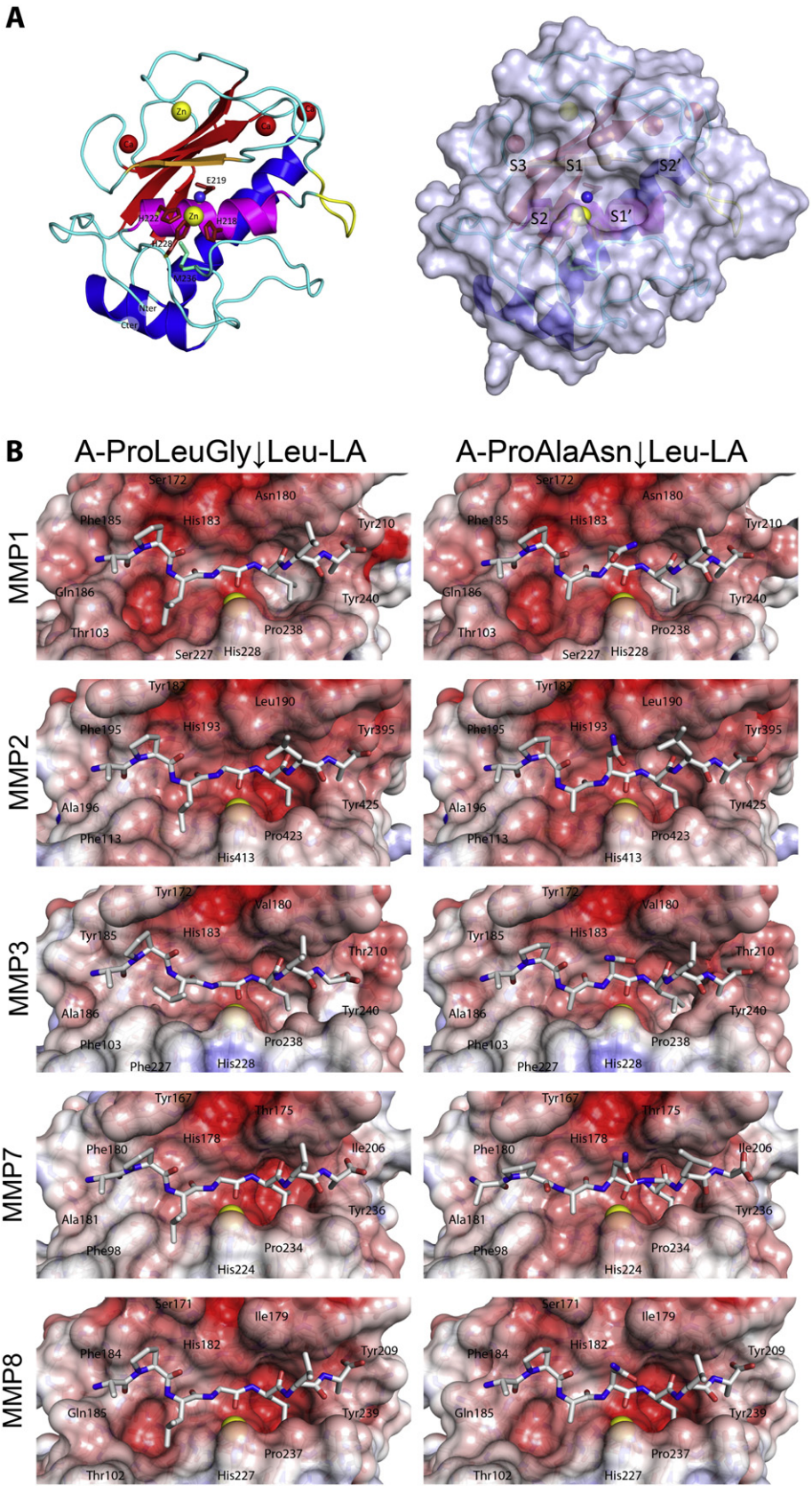


Fig. 7 (*legend on next page*)

for cleavage, we analyzed cooperativity effects between proline in P3 and asparagine in P1 in the PICS data. When asparagine was in P1, co-occurrence with alanine, valine, isoleucine, and leucine in P3 was similar to that of proline (10–15% each), whereas with glycine in P1, even though the combined co-occurrence of alanine, valine, isoleucine, and leucine in P3 stayed high (44%), proline increased more to 31%. Thus, there is cooperativity between having a rigid proline in P3 and a flexible glycine at P1 — without a P3 proline, asparagine preferences can resume. Indeed, if proline is in P3 (PxN↓x) cleavage drops to 56 out of the 4373 cleaved peptides in total and with PxN↓L, only 7 peptides were cleaved by all MMPs. However, when proline in P3 is in combination with glycine in P1, this negative cooperativity is completely lost, supporting our conclusion that non-prime side flexibility impacts binding affinity and cleavage efficiency.

This surprising negative cooperativity between two averaged preferred residues at P3 and P1 can be partly explained by a tyrosine (MMPs 2, 3, 7, 9, and 13) or histidine (MMP12) residue in the upper rim of the catalytic cleft that normally builds up the S3 subsite roof. However, as seen in the MMP13 crystal structures 2d1n and 2d1o [75], this residue can alternatively cap the S1 subsite, making the co-occurrence of proline in P3 and asparagine in P1 exclusive, as beneficial binding interactions can be only provided for one of the two sites. Even though glycine cannot provide any specificity-guiding side-chain interactions, its flexibility seems to be beneficial to accommodate the non-prime side backbone that is rigidified by proline in P3, especially when the substrate latches with large hydrophobic residues such as leucine in S2 and S1′. Furthermore, the glycine-conferred higher flexibility of the scissile bond may facilitate the transition to the tetrahedral cleavage intermediate, and thus may indirectly increase substrate turnover rates.

Even though the collagenase MMPs 1, 8, 13 and 14 preferred PLN↓L, PLG↓L, PAG↓L, and PLG↓L, respectively, no MMP tested, not even the collagenases and gelatinases, preferred leucine in P2 in the PICS cleavage data, yet this is what is present in the classic QF-24 PLG↓L substrate. Besides the similar negative selection of both gelatinases against leucine in P2, only MMP2 favored the exchange to alanine, indicating a larger hydrophobic pocket in MMP9. A comparable preference for alanine over leucine in P2 was additionally only seen in MMPs 3 and 13, most likely due to their rather shallow S2 subsites (Figs. 7 and 8). Cleavage of these QF-peptides also revealed differentiating features within the collagenases. Contrary to MMP13, MMPs 1, 8 and 14 preferred leucine to alanine in P2 in the QF-peptide substrates. However, in both tested scenarios only MMP1 preferred asparagine to glycine in P1 — most likely due to beneficial polar contacts arising from the upper rim residue Asn180. This feature is not present in any of the other tested MMPs (Supplementary Fig. 1), which possess large hydrophobic residues such as leucine, isoleucine, or phenylalanine at the equivalent position. However, in MMP7, the equivalent polar threonine nonetheless is inadequate in length to stabilize asparagine in P1, which is therefore not preferred. Finally, the non-prime side of MMPs is structurally open, especially compared to the more narrow prime side. Thus it seems essential for cleavage efficiency that peptides fully occupy the S2 site so as to bring up the preceding residue to the S3 subsite and to achieve antiparallel alignment of the substrate to the active site facing edge strand of the five-stranded β-pleated sheet of the MMP catalytic domain (Figs. 1B and 7A). In several MMPs this is only possible with large hydrophobic residues, typically built up by two aromatic rings (e.g. Y182 and F195 in MMP2).

It is also important to keep in mind that PICS and QF-peptide cleavage assay results are independent and only conditionally interchangeable. This reflects a strength of PICS, as by allowing low affinity substrates to be turned over with time, a biologically more relevant picture is obtained such as for those natural protein substrates that depend on exosite interactions to promote cleavage of less preferred scissile bonds. This was shown conclusively in protein engineering studies of preferred versus less preferred cleavage sites engineered on protein backbones that either bind a substrate binding hemopexin exosite versus a homologous substrate nonbinding hemopexin domain [69]. This means that by PICS, which analyzes peptide cleavage in the absence of exosite interactions with protein substrates, these features do not mutually select for

Fig. 7. (A) Ribbon and surface representation of the MMP1 catalytic domain in standard orientation showing the S3 to S2′ subsites. The catalytic zinc located in the catalytic groove and the structural zinc located on top of the five-pleated beta sheet are shown as yellow spheres, calcium ions are depicted in red. Proteinaceous zinc ligands and the methionine of the Met-turn are shown as sticks, the active site water is shown as a blue sphere. The central five-stranded β-sheet is colored in red, with the active-site-facing edge strand in orange. The active-site helix is shown in magenta whereas all other helices are colored in blue. The insertion site of the three fibronectin-type II domains in MMPs 2 and 9 is indicated in yellow. (B) In-silico docking of MMPs 1, 2, 3, 7 and 8 with two model heptapeptides, A-PLG↓L-LA (left) and A-PAN↓L-LA (right). The catalytic zinc is depicted as a yellow sphere. Docked peptides and MMPs are shown in stick mode and surface representation with overlaid electrostatic surface potentials (red: negative; blue: positive), respectively.

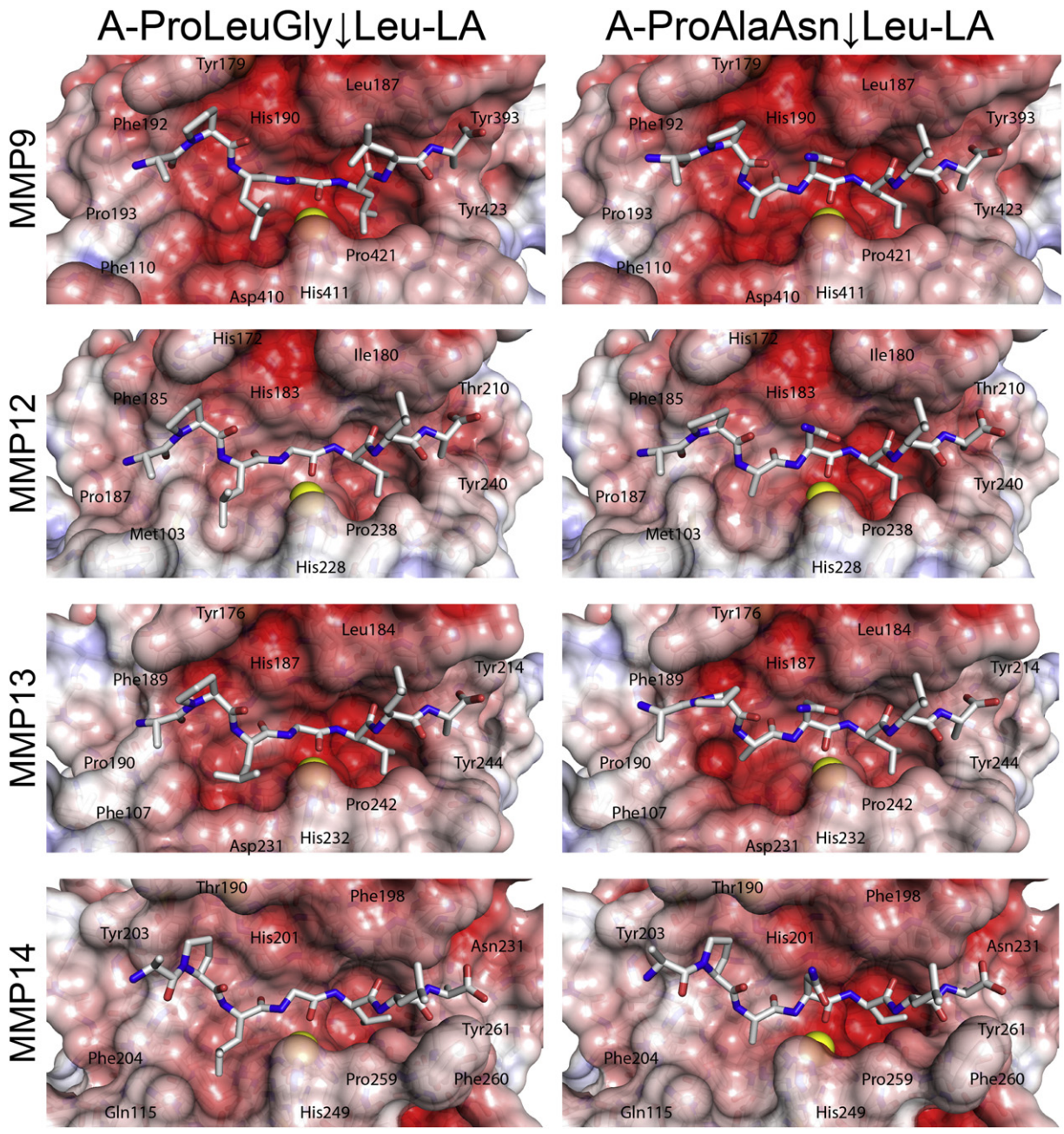


Fig. 8. In-silico docking of MMPs 9, 12, 13 and 14, with two model heptapeptides, A-PLG↓L-LA (left) and A-PAN↓L-LA (right). The catalytic zinc is depicted as a yellow sphere. Docked peptides and MMPs are shown in stick mode and surface representation with overlaid electrostatic surface potentials (red: negative; blue: positive), respectively.

each other. Rather, they allow otherwise less preferred cleaved sequences to be identified, thus generating large numbers of cleavage sites on which robust statistical analyses can be performed. For instance, the two hallmark MMP specificity features P3 proline and P1′ leucine emerged with unexpected negative cooperativity. Yet the presence of both features can still result in a peptide exhibiting favorable cleavage kinetics. Moreover, in a living cell, cleavage does not just rely on a perfect cleavage sequence, but rather on substrate recognition and accessibility of the respective site due to structural considerations, which denatured peptide and synthetic peptide libraries do not

replicate. Thus, even though a certain amount of complementarity of the sequence surrounding the scissile peptide bond and the active site is a prerequisite for cleavage, it is still only one piece in the complex interplay between proteases and their substrates. Indeed, less preferred cleavage sites can be selected for that lead to slow cleavage events which can be a centerpiece of biological control, to dampen sudden oscillations.

The present analysis represents the largest robust specificity profile of a protease family to date by PICS and with 4300 cleavage sites, greatly exceeds the 2405 cleavage site identified by PICS for six type II transmembrane serine proteases previously [50]. The present PICS data can guide small molecule drug development and the interpretation of proteomic results or other proteome-wide natural substrate identification screens such as Terminal Amino Isotopic Labeling of Substrates (TAILS) [76,77], which complements PICS and other peptide approaches that do not identify native substrates [48]. Thus, our compendium of more than 4300 cleavage sites, supported by in-depth peptide cleavage and docking assays, will further assist in the design of improved diagnostic assays and new therapeutics guided by the identified differences in specificity preferences.

PICS provides crucial *in vitro* information that can be exploited to further understand biological roles of MMPs in cleaving components of the ECM and the signaling mediators regulating synthesis or remodeling of the matrix: Specificity profiles can be used to discriminate between direct and indirect effects of a particular protease in the complex dynamic interplay of proteases and inhibitors in the protease web [24] and to distinguish between direct and indirect targets of MMPs in N-terminomics studies *in vivo*. However, despite their obvious strengths, peptide-based approaches such as PICS cannot be used for native substrate discovery nor for the design of exosite-targeting inhibitors to selectively impede the cleavage of a specific substrate or a substrate subset [41,78,79]. The complexity of the higher-order structures of matrix proteins generated by intramolecular and intermolecular interactions forming the complex interconnected ECM renders many techniques of limited use in discovering native substrates and hence roles in ECM biology. For these more complex *in situ* analyses, TAILS N-terminomics is the method of choice as it is proven to successfully identify a myriad of native substrates in complex connective tissues and even in whole arthritic joints of MMP knockout vs. wild-type mice [27] by quantitative proteomic profiling [76,77]. Thus, these high throughput analyses have revealed that ECM components make up just 18% of MMP substrate repertoires, with signaling molecules, receptors and other bioactive mediators in the matrix being more highly represented substrates [6,20–23,26], a big departure from when MMPs were considered mere matrix remodelers [3,12].

Experimental procedures

Expression and purification of human MMPs

MMPs 1, 2, 3, 8, 9 and 13 were expressed from vector pGW1HG and purified from Chinese hamster ovary or *Timp2*$^{-/-}$ murine embryonic fibroblast cell-conditioned medium [31,80,81], and soluble, secretory MMP14 (pro-sMMP14) was purified from *Pichia pastoris* as described elsewhere [82]. The use of different expression and purification schemes was necessary to accommodate the individual needs of the respective MMP variants. Active MMP7 was purchased from Enzo Life Sciences (BML-SE181) and MMP12 was a kind gift from Novartis Pharma AG (Basel, CH). Purity of MMPs was assessed by Coomassie brilliant blue stained SDS-PAGE (15% gels) analysis of 1–2 μg of each MMP under reducing conditions, and using GeneDireX's PiNK Plus Prestained Protein Ladder as molecular weight marker (Fig. 1C). Of note: the mammalian expression vector used in this study (pGW1HG) was at times inaccurately named pGW1GH or pGWIGH in previous publications from our lab. ProMMPs were activated in 100 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, and 0.05% Brij-35, using 1 mM para-aminophenylmercuric acetate at 37 °C for 30 min. Chymotrypsin was used for MMP3 activation (1:100 (*w*/*w*), 30 min, 37 °C) and was subsequently inactivated by adding 1 mM PMSF.

PICS proteome-derived peptide library preparation

Human proteome-derived peptide libraries for MMP assays were prepared as described [48,49]. Cell pellets collected from a human lymphoblast cell line K-562 (ATCC # CCL-243) were lysed in 20 mM Hepes, pH 7.5, supplemented with 0.1% (*w*/*v*) SDS and protease inhibitors (1× Roche cOmplete protease inhibitor cocktail supplemented with 10 mM EDTA and 1 mM PMSF), and centrifuged to remove cell debris (26,000 × *g*, 1 h, 4 °C). Proteins in the supernatant were denatured with 4 M guanidine hydrochloride and cysteine side chains were reduced with 20 mM DTT (1 h, 37 °C) and then protected with 40 mM iodoacetamide (3 h, 20 °C). It is notable that whereas most synthetic peptide libraries do not incorporate cysteine, in PICS libraries cysteine is present as a carbamidomethylated form. Cell lysate proteins were subsequently purified by chloroform/methanol precipitation [110], air-dried, re-suspended in 100 mM Hepes, 5 mM $CaCl_2$, pH 7.5, and digested with trypsin (TPCK treated, Sigma-Aldrich) or GluC (*S. aureus* protease V8, Worthington) at a protease to proteome ratio of 1:100 (*w*/*w*) for 16 h, 37 °C. After

deactivation of trypsin or GluC with 1.0 mM PMSF (30 min, 20 °C), N-terminal and lysine primary amines were blocked by reductive dimethylation with 30 mM formaldehyde and 15 mM sodium cyanoborohydride ($NaCNBH_3$, Sterogene) for 16 h at 20 °C. The resulting peptides were desalted by size exclusion chromatography (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), further purified by reversed-phase chromatography (RESOURCE RPC column, GE Healthcare; wash step with 0.3% (*v*/*v*) formic acid; elution in 80% (*v*/*v*) acetonitrile). Eluates were concentrated to a minimal volume by vacuum concentration, re-suspended in water, and stored at −80 °C as peptide libraries of 200 to 400 µg in concentrations of 5 to 15 mg/ml. All reagents were purchased from Sigma-Aldrich unless otherwise specified.

PICS MMP assays and enrichment of prime side cleavage products

PICS cleavage assays were performed by incubation of 200 to 400 µg peptide library with active recombinant MMP (Fig. 1C) at a protease to peptide library ratio of 1:100 (*w*/*w*) in 50 mM Hepes, 150 mM NaCl and 5 mM $CaCl_2$, pH 7.4. Importantly, the proteome source has negligible influence on the PICS result in our experience [76]. Cleaved peptides presenting neo N-termini generated by MMP cleavage were biotinylated by incubation with 0.5 mM Sulfo-NHS-SS-Biotin (amine-reactive biotinylation reagent with a cleavable disulfide linker; Life Technologies) for 2 h at ambient temperature. Biotinylated prime side cleavage products were affinity purified from uncleaved peptides by incubation with 300 µl streptavidin Sepharose slurry (GE Healthcare) for 2 h at ambient temperature with mild agitation. After extensive washing, biotinylated peptides were eluted with 20 mM DTT (2 h, 22 °C) and 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 50% (*v*/*v*) acetonitrile. The eluates were vacuum dried to near dryness and stored at −80 °C prior to LC–MS/MS analysis.

LC–MS/MS analysis and spectrum to sequence assignment

Samples were analyzed by LC–MS/MS using a LC Packing's capillary LC system (Dionex) coupled online to a quadrupole time-of-flight mass spectrometer (QSTAR XL, Applied Biosystems, operated by the UBC Centre for Blood Research Mass Spectrometry Suite, or QSTAR Pulsar I, Applied Biosystems, operated by the UBC Proteomics Core Facility). 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 to 80% (*v*/*v*) gradient of organic phase over 95 min. Buffer A was 2% (*v*/*v*) acetonitrile with 0.1% (*v*/*v*) formic acid and buffer B was 80% (*v*/*v*) acetonitrile with 0.1% (*v*/*v*) formic acid. MS/MS data were acquired automatically, using the Analyst QS software, v1.1 (Applied Biosystems) for information-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. Peak lists were converted to the mzXML format and peptides identified from the human UniProtKB/Swiss-Prot database [83] containing canonical and isoform protein sequences (downloaded October 2013), using X!TANDEM [84] in conjunction with PeptideProphet [85] as implemented in the Trans Proteomic Pipeline v4.3 [86] at an estimated false discovery rate of 1%. Search parameters included a mass tolerance of 200 ppm for the parent ion and 0.2 Dalton (Da) for the fragment ions, allowing up to two missed cleavages, and carbamidomethylation of cysteine residues (+57.02 Da) and dimethylation of lysine Ɛ-amines (+28.03 Da) as fixed modifications. Methionine oxidation (+15.99 Da), dimethylation (+28.03 Da) and thioacylation (+88.00 Da) of peptide N-termini were set as variable.

LC–MS/MS data analysis

Peptides identified by MS/MS represent the prime side MMP cleavage products. The complete cleavage sites were reconstructed using the open web-based program WebPICS [87] (http://clipserve.clip.ubc.ca/pics/). WebPICS generates a non-redundant list of identified cleavage sites in peptides by matching each non-redundant prime-side peptide sequence to the human IPI database (v3.69, 174,784 entries; EMBL-EBI, UK) [88] 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 Arg or Lys in the case of trypsin-generated libraries or the next N-terminal Asp or Glu for GluC-libraries. Non-prime subsite positions with ambiguous information coming from different protein isoforms are omitted, but the prime-side sequence is still analyzed in the pool. Identified cleavage sites were summarized as heat maps (Gnuplot; www.gnuplot.info) showing relative occurrence and fold-change over natural abundance of amino acids and as iceLogos [89], displaying percent difference compared to natural amino acid abundance. Of note, PICS is not designed for substrate discovery; thus provided UniProt identifiers in Supplementary Tables 1 and 2 are for reference only and do not represent substrate candidates. All mass spectrometry raw data have been deposited to ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner

repository [90] with the data set identifier <PXD002265>.

Quenched fluorescence protease activity assay

Stock solutions (1.0 mM) of synthetic quenched fluorescence (QF) peptide substrates (ChinaPeptides Co. Ltd., Shanghai, China) were dissolved in DMSO. Working stocks (100 μM) were prepared in DMSO using the molar extinction coefficient of the conjugated quencher, (2,4)-dinitrophenyl, of 6.985 cm^{-1} mM^{-1} at 400 nm [91]. Fluorogenic protease-activity assays were performed immediately after MMP activation in the presence of protease inhibitor cocktail (HALT™, Life Technologies, in the absence of EDTA) using a multi-wavelength fluorescence scanner (POLARstar OPTIMA, BMG Labtech). Each purified and activated recombinant MMP (1–10 nM) was incubated with 1 μM QF-peptide in 100 μl of 100 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, and 0.05% Brij-35. The excitation and emission wavelengths were set to 320 and 405 nm, respectively. The fluorescence was measured at 45 s intervals for 1 h at 37 °C. All measurements were performed in duplicate. Experiments were repeated at least three times with independent substrate and MMP preparations on different days.

In silico structural analysis of MMPs

Models of the catalytic domains of human MMPs were generated using the following PDB entries and domain boundaries: 2clt (isolated MMP1, F100-R262), 3ayu (MMP2 in complex with beta-amyloid precursor protein-derived inhibitory peptide, Y110-A447), 1c3i (MMP3 complexed with a propeptide-based inhibitor, F100-P265), 1mmq (MMP7 complexed with hydroxamate, carboxylate, and sulfodiimine inhibitors, Y95-K260), 1bzs (isolated MMP8, F99-L263), 4h3x (MMP9 complexed with a hydroxamate-based inhibitor, F107-P445), 1jiz (MMP12 in complex with a hydroxamic acid inhibitor, F100-D264), 2d1n (MMP13 with a hydroxamic acid inhibitor, Y104-P268), and 1bqq (MMP14 complexed with TIMP2, Y112-G285) [75,92–99]. Missing terminal residues of MMP14 were modeled using the SWISS-MODEL server with default settings [100] and MMP12 (PDB ID: 1jiz) as template to obtain a salt-bridge N-terminus to Asp274; active site mutations were reverted within PyMOL (The PyMOL Molecular Graphics System, Version 1.7.1.3, Schrödinger, LLC) [101]. Secondary structures were assigned using KSDSSP [102] and amino acid sequence numberings were corrected within the UCSF Chimera package [103]. PDB files were converted into PQR files containing per-atom charge and radius using the PDB2PQR server [104] and PROPKA [105] to assign protonation states at pH 7.4. Surface charge representations were created within PyMOL [101] using the APBS Plug-in [106,107]. Peptide dockings with the respective substrates were performed using the high resolution Rosetta FlexPepDock server with default settings [72]. All results were evaluated and optimized based on the available complex structures for the prime site and the anti-parallel edge strand for non-prime substrate recognition. Total energies for the obtained complexes were between −206.2 (MMP1 PLG↓L) and −288.0 (MMP13 PAN↓L) Rosetta energy units (REU), with an average of −253.6 REU and a standard deviation of 32.2. The difference in total energies within the top-10 models of each peptide docking scenario was minor, typically between 2.0 and 3.0 REUs, and with 6–8 of the Rosetta top-10 solutions following the same backbone trace and sidechain orientation, of which one representative was chosen. Integrity of the active site was checked manually for all models. The topology representation was created using TopDraw [108] within the CCP4 software suite [109]. All other molecular graphics figures were made using the molecular visualization system PyMOL.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.matbio.2015.09.003.

Conflict of interest

The authors declare to have no competing interests.

Author contributions

U.E., P.F.H., O.S. and C.M.O. designed the research, performed the experiments, and analyzed and interpreted the data. C.L.B., G.S.B., J.H.C., A.D., V.G., R.K., U.a.d.K., T.K., P.F.L., G.M., A.P., D.R., A.E.S., and Y.W. performed the experiments on individual MMPs. U.E. prepared the figures and wrote the manuscript with P.F.H and O.S. C.M.O. conceived and supervised the project, helped with the data interpretation and manuscript editing, and provided grant support. All authors were involved in revising the article critically, and approved the final version of the manuscript for publication. C.M.O. had full access to all data and takes responsibility for data integrity and accuracy of data analysis.

Acknowledgments

This work was supported by (i) a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology (C.M.O.), (ii) project grants from the Canadian Institutes of Health Research (CIHR; MOP-11433, MOP-37937, and MOP-111055), as well as with (iii) infrastructure

grants from the Michael Smith Research Foundation for Health Research (MSFHR) and the Canada Foundations for Innovation (CFI). U.E. and A.D. were supported by post-doctoral fellowships from MSFHR. The German Academic Exchange Service (DAAD) and the MSFHR supported P.F.H.. O.S. and U.a.d.K. were supported by fellowships from the German Research Foundation (DFG). C.L.B. was supported by postdoctoral fellowships of the Swiss National Science Foundation and the Novartis Jubilee Foundation. P.F.L. was supported by a Feodor Lynen Research Fellowship of the Alexander von Humboldt Foundation. A.P. and G.M. were co-funded by the UBC Centre for Blood Research Strategic Training Program in Transfusion Science and a CBR Internal Collaborative Training Award, respectively. J.H.C. and A.E.S. were both supported by Natural Sciences and Engineering Research Council of Canada Graduate Scholarships and CIHR Strategic Training Fellowships, with additional funding from MSFHR. None of the funders had a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors thank all current and former members of the Overall Lab at the University of British Columbia for inspiring discussions, feedback, and support. We thank Wei Chen, Suzanne Perry and Jason Rogalski from the University of British Columbia Proteomics Core Facility and the UBC Centre for Blood Research Mass Spectrometry Suite for excellent LC–MS/MS measurements.

Received 13 August 2015;
Received in revised form 17 September 2015;
Accepted 18 September 2015
Available online 25 September 2015

Keywords:
MMPs;
Specificity profiling;
Proteomics;
Quenched-fluorescent peptides;
PICS;
Peptide docking
‡ contributed equally
# in alphabetical order

References

[1] C. Tallant, A. Marrero, F.X. Gomis-Rüth, Matrix metalloproteinases: fold and function of their catalytic domains, Biochim. Biophys. Acta 1803 (2010) 20-8, http://dx.doi.org/10.1016/j.bbamcr.2009.04.003.

[2] K. Maskos, W. Bode, Structural basis of matrix metalloproteinases and tissue inhibitors of metalloproteinases, Mol. Biotechnol. 25 (2003) 241–266, http://dx.doi.org/10.1385/MB:25:3:241.

[3] G. Murphy, H. Nagase, Progress in matrix metalloproteinase research, Mol. Asp. Med. 29 (2008) 290–308, http://dx.doi.org/10.1016/j.mam.2008.05.002.

[4] H. Nagase, R. Visse, G. Murphy, Structure and function of matrix metalloproteinases and TIMPs, Cardiovasc. Res. 69 (2006) 562–573, http://dx.doi.org/10.1016/j.cardiores.2005.12.002.

[5] H.E. Van Wart, H. Birkedal-Hansen, The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 5578–5582.

[6] G.S. Butler, C.M. Overall, Updated biological roles for matrix metalloproteinases and new "intracellular" substrates revealed by degradomics, Biochemistry (Mosc) 48 (2009) 10830–10845, http://dx.doi.org/10.1021/bi901656f.

[7] R.O. Hynes, A. Naba, Overview of the matrisome—an inventory of extracellular matrix constituents and functions, Cold Spring Harb. Perspect. Biol. 4 (2012) a004903, http://dx.doi.org/10.1101/cshperspect.a004903.

[8] L. Chen, M. Nakai, R.J. Belton, R.A. Nowak, Expression of extracellular matrix metalloproteinase inducer and matrix metalloproteinases during mouse embryonic development, Reprod. Camb. Engl. 133 (2007) 405–414, http://dx.doi.org/10.1530/rep.1.01020.

[9] E.I. Deryugina, J.P. Quigley, Tumor angiogenesis: MMP-mediated induction of intravasation- and metastasis-sustaining neovasculature, Matrix Biol. J. Int. Soc. Matrix Biol. 44–46 (2015) 94–112, http://dx.doi.org/10.1016/j.matbio.2015.04.004.

[10] Y. Itoh, MT1-MMP: a key regulator of cell migration in tissue, IUBMB Life 58 (2006) 589–596, http://dx.doi.org/10.1080/15216540600962818.

[11] A. Aiken, R. Khokha, Unraveling metalloproteinase function in skeletal biology and disease using genetically altered mice, Biochim. Biophys. Acta 1803 (2010) 121–132, http://dx.doi.org/10.1016/j.bbamcr.2009.07.002.

[12] J.D. Mott, Z. Werb, Regulation of matrix biology by matrix metalloproteinases, Curr. Opin. Cell Biol. 16 (2004) 558–564, http://dx.doi.org/10.1016/j.ceb.2004.07.010.

[13] G. Giannelli, J. Falk-Marzillier, O. Schiraldi, W.G. Stetler-Stevenson, V. Quaranta, Induction of cell migration by matrix metalloprotease-2 cleavage of laminin-5, Science 277 (1997) 225–228.

[14] S. Schenk, E. Hintermann, M. Bilban, N. Koshikawa, C. Hojilla, R. Khokha, et al., Binding to EGF receptor of a laminin-5 EGF-like fragment liberated during MMP-dependent mammary gland involution, J. Cell Biol. 161 (2003) 197–209, http://dx.doi.org/10.1083/jcb.200208145.

[15] J. Xu, D. Rodriguez, E. Petitclerc, J.J. Kim, M. Hangai, Y.S. Moon, et al., Proteolytic exposure of a cryptic site within collagen type IV is required for angiogenesis and tumor growth in vivo, J. Cell Biol. 154 (2001) 1069–1079, http://dx.doi.org/10.1083/jcb.200103111.

[16] R. Agnihotri, H.C. Crawford, H. Haro, L.M. Matrisian, M.C. Havrda, L. Liaw, Osteopontin, a novel substrate for matrix metalloproteinase-3 (stromelysin-1) and matrix metalloproteinase-7 (matrilysin), J. Biol. Chem 276 (2001) 28261–28267, http://dx.doi.org/10.1074/jbc.M103608200.

[17] E.H. Sage, M. Reed, S.E. Funk, T. Truong, M. Steadele, P. Puolakkainen, et al., Cleavage of the matricellular protein SPARC by matrix metalloproteinase 3 produces polypeptides that influence angiogenesis, J. Biol. Chem 278

(2003) 37849–37857, http://dx.doi.org/10.1074/jbc.M302946200.
[18] J.C. Monboisse, J.B. Oudart, L. Ramont, S. Brassart-Pasco, F.X. Maquart, Matrikines from basement membrane collagens: a new anti-cancer strategy, Biochim. Biophys. Acta 2014 (1840) 2589–2598, http://dx.doi.org/10.1016/j.bbagen.2013.12.029.
[19] S. Ricard-Blum, R. Salza, Matricryptins and matrikines: biologically active fragments of the extracellular matrix, Exp. Dermatol. 23 (2014) 457–463, http://dx.doi.org/10.1111/exd.12435.
[20] J.M. Wells, A. Gaggar, J.E. Blalock, MMP generated matrikines, Matrix Biol. 44–46 (2015) 122–129, http://dx.doi.org/10.1016/j.matbio.2015.01.016.
[21] G.S. Butler, C.M. Overall, Matrix metalloproteinase processing of signaling molecules to regulate inflammation, Periodontol. 63 (2013) 123–148, http://dx.doi.org/10.1111/prd.12035.
[22] C.J. Morrison, G.S. Butler, D. Rodríguez, C.M. Overall, Matrix metalloproteinase proteomics: substrates, targets, and therapy, Curr. Opin. Cell Biol. 21 (2009) 645–653, http://dx.doi.org/10.1016/j.ceb.2009.06.006.
[23] P. Schlage, U. Auf dem Keller, Proteomic approaches to uncover MMP function, Matrix Biol. J. Int. Soc. Matrix Biol. 44–46 (2015) 232–238, http://dx.doi.org/10.1016/j.matbio.2015.01.003.
[24] N. Fortelny, J.H. Cox, R. Kappelhoff, A.E. Starr, P.F. Lange, P. Pavlidis, et al., Network analyses reveal pervasive functional regulation between proteases in the human protease web, PLoS Biol. 12 (2014), e1001869http://dx.doi.org/10.1371/journal.pbio.1001869.
[25] E. Sturrock, C.P. Sommerhoff, Highlight: the protease web, Biol. Chem. 395 (2014) 1133, http://dx.doi.org/10.1515/hsz-2014-0237.
[26] 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.
[27] C.L. Bellac, A. Dufour, M.J. Krisinger, A. Loonchanta, A.E. Starr, U. Auf dem Keller, et al., Macrophage matrix metalloproteinase-12 dampens inflammation and neutrophil influx in arthritis, Cell Rep. 9 (2014) 618–632, http://dx.doi.org/10.1016/j.celrep.2014.09.006.
[28] H. Haro, H.C. Crawford, B. Fingleton, K. Shinomiya, D.M. Spengler, L.M. Matrisian, Matrix metalloproteinase-7-dependent release of tumor necrosis factor-alpha in a model of herniated disc resorption, J. Clin. Invest. 105 (2000) 143–150, http://dx.doi.org/10.1172/JCI7091.
[29] M. Suzuki, G. Raab, M.A. Moses, C.A. Fernandez, M. Klagsbrun, Matrix metalloproteinase-3 releases active heparin-binding EGF-like growth factor by cleavage at a specific juxtamembrane site, J. Biol. Chem. 272 (1997) 31730–31737.
[30] K. Zhang, G.A. McQuibban, C. Silva, G.S. Butler, J.B. Johnston, J. Holden, et al., HIV-induced metalloproteinase processing of the chemokine stromal cell derived factor-1 causes neurodegeneration, Nat. Neurosci. 6 (2003) 1064–1071, http://dx.doi.org/10.1038/nn1127.
[31] G.A. McQuibban, J.H. Gong, E.M. Tam, C.A. McCulloch, I. Clark-Lewis, C.M. Overall, Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3, Science 289 (2000) 1202–1206.
[32] C.L. Wilson, A.J. Ouellette, D.P. Satchell, T. Ayabe, Y.S. López-Boado, J.L. Stratman, et al., Regulation of intestinal alpha-defensin activation by the metalloproteinase matrilysin in innate host defense, Science 286 (1999) 113–117.
[33] N. Poulalhon, D. Farge, N. Roos, C. Tacheau, C. Neuzillet, L. Michel, et al., Modulation of collagen and MMP-1 gene expression in fibroblasts by the immunosuppressive drug rapamycin. A direct role as an antifibrotic agent? J. Biol. Chem. 281 (2006) 33045–33052, http://dx.doi.org/10.1074/jbc.M606366200.
[34] A. Dufour, C.M. Overall, Missing the target: matrix metalloproteinase antitargets in inflammation and cancer, Trends Pharmacol. Sci. 34 (2013) 233–242, http://dx.doi.org/10.1016/j.tips.2013.02.004.
[35] G. Shay, C.C. Lynch, B. Fingleton, Moving targets: emerging roles for MMPs in cancer progression and metastasis, Matrix Biol. J. Int. Soc. Matrix Biol. 44–46 (2015) 200–206, http://dx.doi.org/10.1016/j.matbio.2015.01.019.
[36] C. Bonnans, J. Chou, Z. Werb, Remodelling the extracellular matrix in development and disease, Nat. Rev. Mol. Cell Biol. 15 (2014) 786–801, http://dx.doi.org/10.1038/nrm3904.
[37] T.R. Cox, J.T. Erler, Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer, Dis. Model. Mech. 4 (2011) 165–178, http://dx.doi.org/10.1242/dmm.004077.
[38] P. Lu, K. Takai, V.M. Weaver, Z. Werb, Extracellular matrix degradation and remodeling in development and disease, Cold Spring Harb. Perspect. Biol. 3 (2011)http://dx.doi.org/10.1101/cshperspect.a005058.
[39] K. Kessenbrock, C.-Y. Wang, Z. Werb, Matrix metalloproteinases in stem cell regulation and cancer, Matrix Biol. J. Int. Soc. Matrix Biol. 44–46 (2015) 184–190, http://dx.doi.org/10.1016/j.matbio.2015.01.022.
[40] Y. Itoh, Membrane-type matrix metalloproteinases: their functions and regulations, Matrix Biol. J. Int. Soc. Matrix Biol. 44–46 (2015) 207–223, http://dx.doi.org/10.1016/j.matbio.2015.03.004.
[41] C.M. Overall, Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites, Mol. Biotechnol. 22 (2002) 51–86, http://dx.doi.org/10.1385/MB:22:1:051.
[42] M.L. Patterson, S.J. Atkinson, V. Knäuper, G. Murphy, Specific collagenolysis by gelatinase A, MMP-2, is determined by the hemopexin domain and not the fibronectin-like domain, FEBS Lett. 503 (2001) 158–162, http://dx.doi.org/10.1016/S0014-5793(01)02723-5.
[43] B. Steffensen, U.M. Wallon, C.M. Overall, Extracellular matrix binding properties of recombinant fibronectin type II-like modules of human 72-kDa gelatinase/type IV collagenase. High affinity binding to native type I collagen but not native type IV collagen, J. Biol. Chem. 270 (1995) 11555–11566.
[44] F.X. Gomis-Rüth, T.O. Botelho, W. Bode, A standard orientation for metallopeptidases, Biochim. Biophys. Acta 2012 (1824) 157–163, http://dx.doi.org/10.1016/j.bbapap.2011.04.014.
[45] J.L. Harris, B.J. Backes, F. Leonetti, S. Mahrus, J.A. Ellman, C.S. Craik, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, Proc. Natl. Acad. Sci. 97 (2000) 7754–7759, http://dx.doi.org/10.1073/pnas.140132697.
[46] P. Kasperkiewicz, A.D. Gajda, M. Drąg, Current and prospective applications of non-proteinogenic amino acids in profiling of proteases substrate specificity, Biol. Chem. 393 (2012) 843–851, http://dx.doi.org/10.1515/hsz-2012-0167.

[47] O. Schilling, C.M. Overall, Proteomic discovery of protease substrates, Curr. Opin. Chem. Biol. 11 (2007) 36–45, http://dx.doi.org/10.1016/j.cbpa.2006.11.037.
[48] 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.
[49] 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.
[50] 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), e105984http://dx.doi.org/10.1371/journal.pone.0105984.
[51] 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 Proteomics 10 (2011), M111.009233http://dx.doi.org/10.1074/mcp.M111.009233.
[52] R. Cruz, P. Huesgen, S.P. Riley, A. Wlodawer, C. Faro, C.M. Overall, et al., RC1339/APRc from *Rickettsia conorii* is a novel aspartic protease with properties of retropepsin-like enzymes, PLoS Pathog. 10 (2014), e1004324http://dx.doi.org/10.1371/journal.ppat.1004324.
[53] 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. Proteome 100 (2014) 102–114, http://dx.doi.org/10.1016/j.jprot.2013.10.004.
[54] 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.
[55] B.E. Turk, L.L. Huang, E.T. Piro, L.C. Cantley, Determination of protease cleavage site motifs using mixture-based oriented peptide libraries, Nat. Biotechnol. 19 (2001) 661–667, http://dx.doi.org/10.1038/90273.
[56] E.I. Chen, W. Li, A. Godzik, E.W. Howard, J.W. Smith, A residue in the S2 subsite controls substrate selectivity of matrix metalloproteinase-2 and matrix metalloproteinase-9, J. Biol. Chem 278 (2003) 17158–17163, http://dx.doi.org/10.1074/jbc.M210324200.
[57] S.J. Kridel, E. Chen, L.P. Kotra, E.W. Howard, S. Mobashery, J.W. Smith, Substrate hydrolysis by matrix metalloproteinase-9, J. Biol. Chem 276 (2001) 20572–20578, http://dx.doi.org/10.1074/jbc.M100900200.
[58] L.A. Reiter, J.P. Rizzi, J. Pandit, M.J. Lasut, S.M. McGahee, V.D. Parikh, et al., Inhibition of MMP-1 and MMP-13 with phosphinic acids that exploit binding in the S2 pocket, Bioorg. Med. Chem. Lett. 9 (1999) 127–132.
[59] S.J. Deng, D.M. Bickett, J.L. Mitchell, M.H. Lambert, R.K. Blackburn, H.L. Carter 3rd, et al., Substrate specificity of human collagenase 3 assessed using a phage-displayed peptide library, J. Biol. Chem. 275 (2000) 31422–31427, http://dx.doi.org/10.1074/jbc.M004538200.
[60] G.M. McGeehan, D.M. Bickett, M. Green, D. Kassel, J.S. Wiseman, J. Berman, Characterization of the peptide substrate specificities of interstitial collagenase and 92-kDa gelatinase. Implications for substrate optimization, J. Biol. Chem. 269 (1994) 32814–32820.
[61] M.M. Smith, L. Shi, M. Navre, Rapid identification of highly active and selective substrates for stromelysin and matrilysin using bacteriophage peptide display libraries, J. Biol. Chem. 270 (1995) 6440–6449.
[62] H. Nagase, C.G. Fields, G.B. Fields, Design and characterization of a fluorogenic substrate selectively hydrolyzed by stromelysin 1 (matrix metalloproteinase-3), J. Biol. Chem. 269 (1994) 20952–20957.
[63] H. Nagase, G.B. Fields, Human matrix metalloproteinase specificity studies using collagen sequence-based synthetic peptides, Biopolymers 40 (1996) 399–416, http://dx.doi.org/10.1002/(SICI)1097-0282(1996)40:4<399::AID-BIP5>3.0.CO;2-R.
[64] S. Netzel-Arnett, Q.X. Sang, W.G. Moore, M. Navre, H. Birkedal-Hansen, H.E. Van Wart, Comparative sequence specificities of human 72- and 92-kDa gelatinases (type IV collagenases) and PUMP (matrilysin), Biochemistry (Mosc) 32 (1993) 6427–6432.
[65] W. Bode, C. Fernandez-Catalan, H. Tschesche, F. Grams, H. Nagase, K. Maskos, Structural properties of matrix metalloproteinases, Cell Mol. Life Sci. 55 (1999) 639–652.
[66] 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. U. S. A. 111 (2014) E4148–E4155, http://dx.doi.org/10.1073/pnas.1406134111.
[67] R.A. Dean, J.H. Cox, C.L. Bellac, A. Doucet, A.E. Starr, C.M. Overall, Macrophage-specific metalloelastase (MMP-12) truncates and inactivates ELR+ CXC chemokines and generates CCL2, -7, -8, and -13 antagonists: potential role of the macrophage in terminating polymorphonuclear leukocyte influx, Blood 112 (2008) 3455–3464, http://dx.doi.org/10.1182/blood-2007-12-129080.
[68] 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.
[69] C.M. Overall, G.A. McQuibban, I. Clark-Lewis, Discovery of chemokine substrates for matrix metalloproteinases by exosite scanning: a new tool for degradomics, Biol. Chem. 383 (2002) 1059–1066, http://dx.doi.org/10.1515/BC.2002.114.
[70] Q. Li, P.W. Park, C.L. Wilson, W.C. Parks, Matrilysin shedding of syndecan-1 regulates chemokine mobilization and transepithelial efflux of neutrophils in acute lung injury, Cell 111 (2002) 635–646.
[71] 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.
[72] N. London, B. Raveh, E. Cohen, G. Fathi, O. Schueler-Furman, Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions, Nucleic Acids Res. 39 (2011) W249–W253, http://dx.doi.org/10.1093/nar/gkr431.
[73] M. Kukreja, S.A. Shiryaev, P. Cieplak, N. Muranaka, D.A. Routenberg, A.V. Chernov, et al., High-throughput multiplexed peptide-centric profiling illustrates both substrate

cleavage redundancy and specificity in the MMP family, Chem. Biol. 22 (2015) 1122–1133, http://dx.doi.org/10.1016/j.chembiol.2015.07.008.
[74] 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.
[75] T. Kohno, H. Hochigai, E. Yamashita, T. Tsukihara, M. Kanaoka, Crystal structures of the catalytic domain of human stromelysin-1 (MMP-3) and collagenase-3 (MMP-13) with a hydroxamic acid inhibitor SM-25453, Biochem. Biophys. Res. Commun. 344 (2006) 315–322, http://dx.doi.org/10.1016/j.bbrc.2006.03.098.
[76] 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.
[77] 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.
[78] G.B. Fields, New strategies for targeting matrix metalloproteinases, Matrix Biol. J. Int. Soc. Matrix Biol. 44–46 (2015) 239–246, http://dx.doi.org/10.1016/j.matbio.2015.01.002.
[79] N. Sela-Passwell, A. Trahtenherts, A. Krüger, I. Sagi, New opportunities in drug design of metalloproteinase inhibitors: combination between structure–function experimental approaches and systems biology, Expert Opin. Drug Discov. 6 (2011) 527–542, http://dx.doi.org/10.1517/17460441.2011.560936.
[80] 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.
[81] 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.
[82] D.V. Rozanov, A.Y. Strongin, Membrane type-1 matrix metalloproteinase functions as a proprotein self-convertase. Expression of the latent zymogen in *Pichia pastoris*, autolytic activation, and the peptide sequence of the cleavage forms, J. Biol. Chem 278 (2003) 8257–8260, http://dx.doi.org/10.1074/jbc.M213246200.
[83] The UniProt Consortium, Activities at the Universal Protein Resource (UniProt), Nucleic Acids Res. 42 (2014) D191–D198, http://dx.doi.org/10.1093/nar/gkt1140.
[84] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinforma Oxf. Engl. 2004;20: 1466–7. doi:http://dx.doi.org/10.1093/bioinformatics/bth092.
[85] 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.
[86] A. Keller, J. Eng, N. Zhang, X. Li, R. Aebersold, A uniform proteomics MS/MS analysis platform utilizing open XML file formats, Mol. Syst. Biol. 1 (2005) 2005.0017, http://dx.doi.org/10.1038/msb4100024.
[87] 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.
[88] P.J. Kersey, J. Duarte, A. Williams, Y. Karavidopoulou, E. Birney, R. Apweiler, The International Protein Index: an integrated database for proteomics experiments, Proteomics 4 (2004) 1985–1988, http://dx.doi.org/10.1002/pmic.200300721.
[89] N. Colaert, K. Helsens, 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/nmeth1109-786.
[90] J.A. Vizcaíno, 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–D1069, http://dx.doi.org/10.1093/nar/gks1262.
[91] 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.
[92] C. Antoni, L. Vera, L. Devel, M.P. Catalani, B. Czarny, E. Cassar-Lajeunesse, et al., Crystallization of bi-functional ligand protein complexes, J. Struct. Biol. 182 (2013) 246–254, http://dx.doi.org/10.1016/j.jsb.2013.03.015.
[93] M.F. Browner, W.W. Smith, A.L. Castelhano, Matrilysin-inhibitor complexes: common themes among metalloproteases, Biochemistry (Mosc) 34 (1995) 6602–6610.
[94] C. Fernandez-Catalan, W. Bode, R. Huber, D. Turk, J.J. Calvete, A. Lichte, et al., Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor, EMBO J. 17 (1998) 5238–5248, http://dx.doi.org/10.1093/emboj/17.17.5238.
[95] H. Hashimoto, T. Takeuchi, K. Komatsu, K. Miyazaki, M. Sato, S. Higashi, Structural basis for matrix metalloproteinase-2 (MMP-2)-selective inhibitory action of β-amyloid precursor protein-derived inhibitor, J. Biol. Chem. 286 (2011) 33236–33243, http://dx.doi.org/10.1074/jbc.M111.264176.
[96] S. Iyer, R. Visse, H. Nagase, K.R. Acharya, Crystal structure of an active form of human MMP-1, J. Mol. Biol. 362 (2006) 78–88, http://dx.doi.org/10.1016/j.jmb.2006.06.079.
[97] H. Matter, W. Schwab, D. Barbier, G. Billen, B. Haase, B. Neises, et al., Quantitative structure–activity relationship of human neutrophil collagenase (MMP-8) inhibitors using comparative molecular field analysis and X-ray structure analysis, J. Med. Chem. 42 (1999) 1908–1920, http://dx.doi.org/10.1021/jm980631s.
[98] H. Nar, K. Werle, M.M. Bauer, H. Dollinger, B. Jung, Crystal structure of human macrophage elastase (MMP-12) in complex with a hydroxamic acid inhibitor, J. Mol.

Biol. 312 (2001) 743–751, http://dx.doi.org/10.1006/jmbi.2001.4953.
[99] D.L. Steele, O. El-Kabbani, P. Dunten, L.J. Windsor, R.U. Kammlott, R.L. Crowther, et al., Expression, characterization and structure determination of an active site mutant (Glu202-Gln) of mini-stromelysin-1, Protein Eng. 13 (2000) 397–405.
[100] L. Bordoli, T. Schwede, Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal, Methods Mol. Biol. 857 (2012) 107–136, http://dx.doi.org/10.1007/978-1-61779-588-6_5.
[101] W.L. DeLano, The case for open-source software in drug discovery, Drug Discov. Today 10 (2005) 213–217, http://dx.doi.org/10.1016/S1359-6446(04)03363-X.
[102] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637, http://dx.doi.org/10.1002/bip.360221211.
[103] Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C.C. Huang, E.F. Pettersen, et al., UCSF chimera, MODELLER, and IMP: an integrated modeling system, J. Struct. Biol. 179 (2012) 269–278, http://dx.doi.org/10.1016/j.jsb.2011.09.006.
[104] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667, http://dx.doi.org/10.1093/nar/gkh381.
[105] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. 7 (2011) 525–537, http://dx.doi.org/10.1021/ct100578z.
[106] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041, http://dx.doi.org/10.1073/pnas.181342398.
[107] M. Lerner, H. Carlson, APBS Plugin for Pymol, Ann Arbor Univ Mich, 2008.
[108] C.S. Bond, TopDraw: a sketchpad for protein structure topology cartoons, Bioinforma Oxf. Engl. 19 (2003) 311–312.
[109] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, et al., Overview of the CCP4 suite and current developments, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 235–242, http://dx.doi.org/10.1107/S0907444910045749.
[110] 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, http://dx.doi.org/10.1016/0003-2697(84)90782-6.
[111] D.J. Marchant, C.L. Bellac, T.J. Moraes, S.J. Wadsworth, A. Dufour, G.S. Butler, et al., A new transcriptional role for matrix metalloproteinase-12 in antiviral immunity, Nat. Med. 20 (2014) 493–502, http://dx.doi.org/10.1038/nm.3508.