Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot
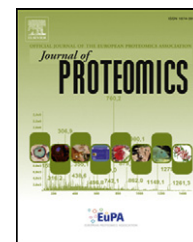
Proteomic protease specificity profiling of clostridial collagenases reveals their intrinsic nature as dedicated degraders of collagen☆☆

CrossMark

Ulrich Eckhard[a,b], Pitter F. Huesgen[a], Hans Brandstetter[b], Christopher M. Overall[a,*]

[a]Centre for Blood Research, Department of Oral Biological and Medical Sciences, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada
[b]Division of Structural Biology, Department of Molecular Biology, University of Salzburg, Billrothstr, 11, 5020 Salzburg, Austria

ARTICLE INFO

ABSTRACT

Clostridial collagenases are among the most efficient degraders of collagen. Most clostridia are saprophytes and secrete proteases to utilize proteins in their environment as carbon sources; during anaerobic infections, collagenases play a crucial role in host colonization. Several medical and biotechnological applications have emerged utilizing their high collagenolytic efficiency. However, the contribution of the functionally most important peptidase domain to substrate specificity remains unresolved. We investigated the active site sequence specificity of the peptidase domains of collagenase G and H from *Clostridium histolyticum* and collagenase T from *Clostridium tetani*. Both prime and non-prime cleavage site specificity were simultaneously profiled using Proteomic Identification of protease Cleavage Sites (PICS), a mass spectrometry-based method utilizing database searchable proteome-derived peptide libraries. For each enzyme we identified >100 unique-cleaved peptides, resulting in robust cleavage logos revealing collagen-like specificity patterns: a strong preference for glycine in P3 and P1′, proline at P2 and P2′, and a slightly looser specificity at P1, which in collagen is typically occupied by hydroxyproline. This specificity for the classic collagen motifs Gly-Pro-X and Gly-X-Hyp represents a remarkable adaptation considering the complex requirements for substrate unfolding and presentation that need to be fulfilled before a single collagen strand becomes accessible for cleavage.

Biological significance
We demonstrate the striking sequence specificity of a family of clostridial collagenases using proteome derived peptide libraries and PICS, Proteomic Identification of protease Cleavage Sites. In combination with the previously published crystal structures of these proteases, our results represent an important piece of the puzzle in understanding the complex mechanism underlying collagen hydrolysis, and pave the way for the rational design of specific test substrates and selective inhibitors.

---

1.   Introduction

Collagens are by far the most abundant proteins in mammals that constitute up to 90% of the extracellular matrix of a tissue, thus imparting its structural integrity. Collagen is an astonishing molecule of remarkable primary, secondary, tertiary and quaternary structure, fascinating scientists for decades [1–4]. Interstitial collagen is built up by three polypeptide strands with high proline and hydroxyproline (Hyp) content (combined occurrence >20%) and a glycine at every third position. Pro-Hyp-Gly thus forms the most common triplet (10.5%). Each strand possesses the conformation of a left-handed polyproline II helix, twisted together into a rigid, rod shaped, right-handed super helix of approximately 300 nm in length and 1.5 nm in diameter, burying the peptide bonds within the interior of the helix. The three alpha-chains are held together by numerous inter-chain hydrogen bonds, and highly ordered hydration networks surround the triple helices. Collagen molecules spontaneously self-assemble into larger quaternary assemblies of great tensile strength, first as fibrils and then as bundles of fibrils to fibers up to 2 μm in length and 300 nm in diameter. Additional intermolecular interactions increase collagen thermal stability and afford its great tensile strength and long half-life of 15 years for skin collagen and of over 100 years for cartilage collagen [5–9].

Equally fascinating as collagen is the repertoire of proteases, often termed collagenases, capable of degrading native triple helical collagen. Human proteases such as the zinc-dependent matrix metalloproteinases (MMP) 1, 8, 13 and 14, and the cysteine protease cathepsin K, play crucial roles in the dynamic remodeling of connective tissue under physiological and/or pathological conditions [10,11]. Notably, all enzymes show distinct preferences towards the different fibrillar collagen substrates and cleave only at well-defined sites; *e.g.* MMP1, MMP8, MMP13 and MMP14 all share the unique ability to cut the native super-helix into 3/4 and 1/4 length fragments at a single peptide bond (775–776) between a glycine residue at P1 and a leucine (on type I collagen α2) or isoleucine (on type I collagen α1) at P1′. However, their distinct preferences for different types of interstitial collagens indicate different functional roles in the human body [12]. Notably, the hemopexin carboxyl-domain exosites bind collagen and are essential for cleavage [13–16]. Exosites represent specialized substrate binding sites located on domains outside the active site cleft, providing substrate interactions not influenced by the primary specificity pockets, thus refining substrate affinity and proteinase specificity [11]. For cathepsin K, five distinct cleavage sites have been identified within type I collagen, and one in type II collagen. In all cases, subsequent collagenolysis is performed by gelatinases (MMP2, MMP9) and other synergistic proteinases only after the initial dismantling of the collagen triple helix that exposes more generally susceptible proteolytic sites upon denaturation [10,17,18]. However, MMP2 also has a weak native collagenase activity imparted by its three fibronectin type II repeats [15] which function as exosites to localize the enzyme on the substrate.

Bacterial collagenases are secreted by saprophytic clostridia to utilize collagen as a carbon source. Pathogenic strains such as *Clostridium histolyticum* and *Clostridium tetani* use these enzymes to facilitate host invasion, colonization and toxin diffusion during anaerobic infections [19,20]. In contrast to the tightly regulated natural turnover of collagen in the host organism, clostridial collagenases circumvent all constraints for collagen degradation and are capable of efficiently digesting most, if not all, types of collagen [21–23]. Their enzymatic activity is independent of the quaternary collagen assembly, which is especially remarkable given collagen's triple helical structure that renders it resistant to most other proteases [24–26]. Clostridial collagenases not only initiate hydrolysis by multiple cleavages within the triple helical region, but also completely degrade the obtained fragments into a mixture of small oligopeptides [22,24,27]. This ability is often referred to as broad substrate specificity of clostridial collagenases [21–23], although no natural substrate other than collagen has been reported to our knowledge. These properties place clostridial collagenases among the most efficient enzymes degrading all types of collagen, which is also exploited by a diverse spectrum of medical and biotechnological applications. For instance, collagenases from *C. histolyticum* are approved therapeutic agents for breaking down the tough collagen cords in Morbus Dupuytren [28,29] and are widely used in enzymatic wound debridement [30,31] and tissue dissociation experiments [32,33].

Clostridial collagenases are large multi-modular zinc-metalloproteinase of approximately 115 kDa, consisting of four to six domains [19,26,34–36] (Fig. 1). As members of the gluzincin superfamily of metalloproteinases they share the common HExxH zinc binding motif within the peptidase domain of the N-terminal collagenase unit, complemented by an additional zinc-coordinating glutamate 28 to 30 amino acid downstream. Two to four accessory domains of approximately 10 kDa each form a C-terminal collagen recruitment unit of variable composition (Fig. 1), providing important exosites for native collagenolysis, responsible for collagen binding [37] and swelling [38] (Fig. 1). Although most clostridial strains possess only one collagenase, *C. histolyticum* encodes for two with complementary characteristics: collagenase G (ColG) exhibits high collagenolytic and low peptidolytic activity, whereas its homologue collagenase H (ColH) shows low collagenolytic and high peptidolytic activity [39].

Recently, we established a flexible expression and purification platform for clostridial collagenases [35], undertook a 'one substrate, one inhibitor'-based biochemical characterization of the collagenase units of ColG, ColH and the *C. tetani* collagenase (ColT) [40], and reported the crystal structures of the ColG collagenase unit [36], the ColG Polycystic Kidney Disease (PKD)-like domains [37], and of the ColH and ColT peptidase domains [41]. The ColG crystal structure showed a saddle-shaped architecture of the N-terminal collagenase unit (Tyr119-Gly790) capable of degrading native triple-helical collagen even in the absence of its accessory collagen recruitment domain [36]. The N-terminal activator domain (Tyr119-Asp388) was required for
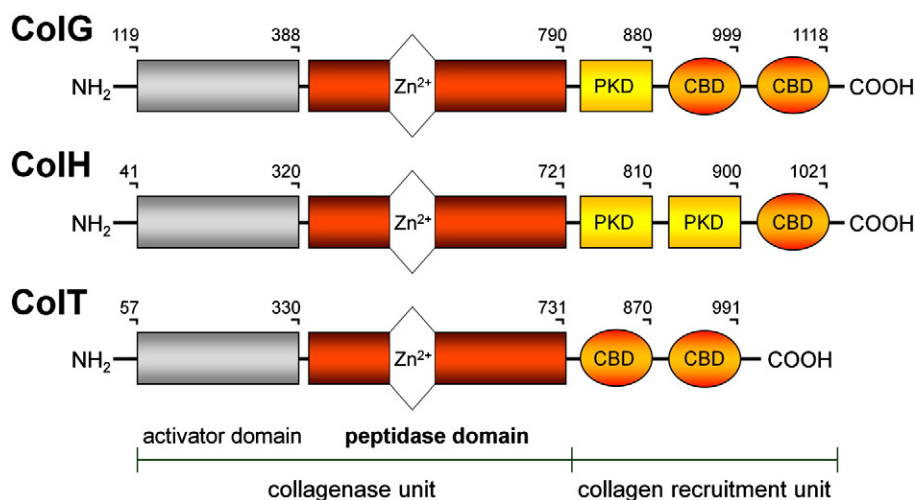
Fig. 1 – Schematic representation of the domain organization of *Clostridium histolyticum* collagenases G (ColG) and H (ColH), and collagenase T from *C. tetani* (ColT). The location of the catalytic zinc ion is indicated within the peptidase domain. PKD, polycystic kidney disease-like domain; CBD, collagen binding domain.

degrading native collagen, but dispensable for the peptidolytic activity. The crystal structures of the peptidase domains of ColH and ColT revealed a calcium binding site in close proximity to the active site, an aspartate switch for zinc stabilization and a conformational selectivity filter [41]. The calcium close to the active site explained for the first time its importance for structural integrity and enzymatic activity and the aspartate switch stabilized the active site zinc analogous to the tyrosine switch in active astacin [42]. The selectivity filter regulated the substrate access to the active site by steric shielding, explaining their opposing peptidolytic and collagenolytic efficiency [41].

The dynamic interactions between proteases and substrates are also reflected in the cleavage site specificity, which relies largely on the recognition of the amino acid sequences flanking the scissile peptide bond by the protease active site. In 1967, Schechter and Berger introduced the standard nomenclature for describing interactions between a peptide substrate and the active site of a protease [43]. Residues N-terminal to the scissile peptide bond are designated as the non-prime (P) side, and residues C-terminal as prime (P′) side, both interacting with complementary substrate recognition sites in the protease active site cleft, termed S and S′, respectively. Protease specificity can be profiled using a variety of approaches [44]. Uniquely among these, Proteomic Identification of protease Cleavage Sites (PICS) determines both the full prime and non-prime cleavage site preferences in a single experiment and in a high throughput manner [45]. PICS uses proteome-derived peptide libraries, generated by digestion of a suitable proteome with a highly specific protease *e.g.* trypsin and GluC. For cleavage specificity analysis, primary amines in the substrate library are blocked prior to incubation with the test protease of interest. The blocking step is critical as it allows for selective biotin labeling of the newly generated neo-N-termini. This enables selective streptavidin capture and mass spectrometry-based identification of the prime side cleavage products. Due to the non-random sequence of the proteome-derived test substrates, the complete cleavage site can be reconstructed by bioinformatics of the non-prime side

sequences from proteome databases. Alignment of dozens to hundreds of cleaved peptides then provides a detailed profile of the test protease cleavage site specificity.

Here, we present the cloning, expression and purification of the peptidase domain of ColG, and the first detailed active site specificity profiles of the peptidase domains of three different pathogenic clostridial collagenases, *C. histolyticum* ColG, ColH and *C. tetani* ColT. Our results showed a striking match to the cleaved sequence in collagen, demonstrating that the exquisite specificity of clostridial collagenases for collagen arises primarily from the active site, and does not depend on exosite contributions from the ancillary domains. This sheds additional light on the complex mechanism of hydrolyzing large quaternary protein molecules such as collagen and provides new opportunities for the design of sensitive assays and specific peptidic inhibitors.

2.   Material and methods

2.1.   Materials

All restriction enzymes were purchased from Fermentas (St. Leon-Rot, Germany); primers were obtained from Sigma-Aldrich (Vienna, Austria), and sequence analysis was performed at Eurofins MWG Operon (Ebersberg, Germany). All reagents were of the highest standard available from AppliChem (Darmstadt, Germany) or Sigma-Aldrich (St. Louis, MO, USA), unless stated otherwise.

2.2.   Cloning

Based on the crystal structure of the collagenase unit of collagenase G from *C. histolyticum* (pdb entries 2y3u) [36], we designed a peptidase domain construct comprising the amino acid sequence Asp398 to Gly790 (numbering according to Swiss-Prot entry Q9X721). After PCR amplification from the full length construct [35], the peptidase domain was cloned

into a modified pET-15b vector, encoding for an N-terminal His$_6$-tag followed by a TEV-protease recognition site (ENLYFQxGGT), and the peptidase starting sequence. Primers used were: ColG-fwd: 5′-ACGTggtacc GATCACGATAAGTTC TTAGA-3′ and ColG-rev: 5′-ACGTggatccTTACCCATTATCTG TTAAAACCC-3′; restriction sites encoding for *Kpn I* and *Bam HI* are represented in lowercase letters. PCR products were purified with MinElute PCR Purification Kit (Qiagen, Germany) and digested with the appropriate restriction enzymes, ligated overnight under standard conditions, and transformed into XL2-Blue (Stratagene, Germany) cells *via* electroporation. Nucleotide sequence was confirmed by DNA sequencing prior to protein expression (Eurofins MWG Operon, Ebersberg, Germany). Cloning of the peptidase domain constructs of collagenase H from *C. histolyticum* (Leu331-Gly721; Swiss-Prot entry Q46085) and collagenase T from *C. tetani* (D340-K731; numbering according to Swiss-Prot entry Q899Y1) has been described previously [41].

2.3. Expression and purification

After induction with 1 mM IPTG at OD$_{600}$ 0.8, the peptidase domain construct of ColG was expressed in soluble form in *Escherichia coli* BL21(DE3) cells for 4 h at 37 °C in 2 L baffled flasks, containing 500 mL LB-medium with 100 μg/mL ampicillin. Cells were collected by centrifugation (4000 *g*, 10 min, 4 °C) and resuspended in 1/100 culture volume of Buffer A (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0). Cells were lysed by sonication (4 cycles of 45 s, at 45 W power, and with 0.5 s intervals), and clarified by two cycles of centrifugation (15,000 *g*, 10 min, 4 °C). To reduce viscosity, we enzymatically removed DNA from the sample with 1.0 μg/mL DNase I and 5.0 mM MgCl$_2$ and incubation for 15 min on ice. All subsequent purification steps were carried out at 4 °C. The three-step purification scheme included immobilized metal affinity chromatography *via* pre-equilibrated Ni–NTA resin (Qiagen), His tag removal by in-house prepared TEV protease in a molar ratio of 1:100 overnight, re-chromatography, and size exclusion chromatography Superdex 75 10/300 GL, Amersham Bioscience for re-buffering (10 mM Tris, 10 mM NaCl, pH 7.5) and as final polishing step, as described elsewhere in great detail for other clostridial collagenase constructs [35,36,40]. Protein samples were concentrated before and after size exclusion chromatography to a concentration of approximately 10 mg/mL using washed and pre-equilibrated 10 kDa molecular-weight cutoff Centricons (Amicon-Ultra, Millipore). Protein concentration was assayed by UV$_{280}$, using the molar extinction coefficient of 77,700 M^{-1} cm^{-1}, calculated by the ExPASy ProtParam tool [46]. After purification, samples were snap-frozen in liquid nitrogen and stored as 25 μl aliquots at −80 °C until use. Expression and purification of the peptidase domains of ColH and ColT has been described previously [41].

2.4. Enzymatic assay

Enzymatic assays with N-[3-(2-Furylacryloyl)]-ʟ-leucyl-glycyl-ʟ-prolyl-ʟ-alanine (FALGPA) as substrate were performed as described by van Wart and Steinbrink [47] and detailed in the manufacture's protocol (AppliChem, Germany). A 1.0 mM stock solution was prepared in reaction buffer, and the concentration verified in solution *via* UV absorbance at 305 nm (ε_{305} = 24.70 mM^{-1} cm^{-1}). All measurements were carried out at 25 °C, and the decrease in absorbance upon substrate cleavage was monitored at 345 nm, as described in detail [40].

2.5. PICS peptide library preparation

Human whole proteome-derived peptide libraries were essentially prepared as described in detail previously [48]. In brief, cell pellets collected from HEK293 cell cultures were lysed in 20 mM Hepes, pH 7.5, supplemented with 0.1% sodium dodecyl sulfate (SDS) and protease inhibitors (1× Roche complete protease inhibitor cocktail, supplemented with 1 mM PMSF and 10 mM EDTA), 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 reduced with DTT (5 mM, 1 h, 37 °C). Free sulfhydryl groups were alkylated with iodoacetamide (20 mM, 3 h, 20 °C, in the dark). Labeling was stopped by adding further DTT (5 mM, 15 min, 20 °C) and the reaction was cleaned-up by chloroform/methanol precipitation [21]. The protein pellet was air-dried, resuspended in 50 μl of 0.1 M NaOH and diluted with ddH$_2$O to an estimated protein concentration of 2 mg/mL. Samples were neutralized with 100 mM Hepes, 5 mM CaCl$_2$, pH 7.2, and digested overnight at 37 °C with trypsin (TPCK-treated trypsin, Sigma) at a protease-to-protein weight ratio of 1:100. After inactivation of trypsin with 1 mM PMSF (30 min, 20 °C), undigested protein aggregates were removed by centrifugation (20,000 *g*, 10 min, 4 °C). Reduction and alkylation steps were repeated (5 mM DTT, 1 h, 37 °C; 40 mM iodoacetamide, 1.5 h, 20 °C, in the dark) to ensure complete protection of free sulfhydryl groups. Primary amines were protected by reductive dimethylation with formaldehyde (30 mM) and sodium cyanoborohydride (15 mM) for 16 h overnight at 20 °C. To ensure completeness of amine labeling, another 15 mM formaldehyde and 15 mM sodium cyanoborohydride were added and reaction was incubated for additional 2 h at 37 °C. The resulting peptides were desalted by size exclusion chromatography (Sephadex G-10 columns, 10 mM potassium phosphate buffer, pH 2.7, 10% methanol) and, after vacuum evaporation of methanol, further purified by reversed-phase chromatography (RESOURCE RPC column, GE Healthcare) using an FPLC system (Äkta Explorer, GE Healthcare). Eluates were concentrated to a minute amount of solvent by vacuum concentration, and resuspended in water to an approximate concentration of 2.0 mg/mL. Peptide concentration was estimated using the bicinchoninic assay (BCA, Pierce). Purified tryptic human whole proteome peptide libraries were stored in aliquots of 200 μg at −80 °C until use.

2.6. PICS cleavage site specificity assay

Cleavage assays were performed by incubating 250 μg of human whole-proteome peptide library with active recombinant clostridial collagenase at a protease to peptide library ratio of 1:50 (w/w) in 50 mM Hepes, 150 mM NaCl, 5 mM CaCl$_2$ at pH 7.4, incubated for 16 h overnight, and stopped by heat inactivation at 70 °C for 30 min. Prime-side cleavage products were isolated by positive enrichment essentially as described [48]. In brief, cleaved peptides presenting neo-amino termini generated by

collagenase activity were biotinylated by incubation with 0.5 mM sulfosuccinimidyl 2-(biotinamido)-ethyl-1,3-dithiopropionate (amine-reactive biotin with a cleavable disulfide linker; Pierce) for 2 h at ambient temperature. Biotinylated prime-side cleavage products were separated 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 with 50 mM Hepes, pH 7.2, biotinylated peptides were eluted with 20 mM DTT (2 h, 20 °C), desalted using reversed-phase solid phase extraction (Sep-Pak, Waters) with binding and washing in 0.1% (vol/vol) formic acid and elution in 80% (vol/vol) acetonitrile. The eluates were evaporated to near dryness in a SpeedVac concentrator, brought to 15 μl with 0.1% formic acid, and stored at −20 °C until LC–MS/MS analysis.

2.7. Nanoflow HPLC–MS/MS analysis and spectrum to sequence assignment

LC–MS/MS analysis was performed on a nano-LC system (Thermo) coupled to a linear ion trap-orbitrap (LTQ-Orbitrap XL, Thermo Finnigan, San Jose, CA) using a nanospray ionization source, a 2-cm-long, 100-μm-inner diameter fused silica trap column, a 20 cm long, 50-μm-inner diameter fused silica fritted analytical column, and a 20-μm-inner diameter fused silica gold coated spray tip (6-μm-diameter opening, pulled on a P-2000 laser puller from Sutter Instruments, coated on Leica EM SCD005 Super Cool Sputtering Device). The trap column was packed with 5 μm-diameter Aqua C-18 beads (Phenomenex), while the analytical column was packed with 3 μm-diameter Reprosil-Pur C-18-AQ beads (Dr. Maisch, Ammerbuch, Germany). Buffer A consisted of 0.5% acetic acid, and buffer B of 0.5% acetic acid and 80% acetonitrile. Gradients were run from 0% B to 15% B over 15 min, then from 15% B to 40% B in the next 65 min, then increased to 100% B over 10 min period, and held at 100% B for 30 min. Sample pick up was at 10 μL/min, sample loading volume was 20 μL (for 5 μL samples), pre-column equilibration volume 20 μL and analytical column equilibration volume 4 μL. The LTQ-Orbitrap was set to acquire a full-range scan at 60,000 resolution from 350 to 1800 m/z in the Orbitrap, and to simultaneously fragment the top five peptide ions in each cycle in the LTQ (minimum intensity 200 counts). Parent ions were then excluded from MS/MS for the next 180 s. Single charged ions were excluded since in ESI mode peptides usually carry multiple charges. The Orbitrap was continuously recalibrated using the lock-mass function [49]. Error of mass measurement was usually within 5 ppm and was not allowed to exceed 10 ppm (see Supplementary Fig. 1).

2.8. Data analysis

Acquired Orbitrap data were centroided and converted to the mzXML and mgf format using ms-convert [50]. Peptides were identified from the human UniProt database (version 2012–11) using the search engines X! TANDEM [51] and Mascot [52]. Peptides were analyzed with PeptideProphet [53] and iProphet [54], both implemented within the Trans Proteomic Pipeline v4.6 [55]. Peptides were analyzed at a false discovery rate of 1% for each peptide by using a decoy search strategy. Search parameters included a mass tolerance of 10 ppm for parental

ions and 0.5 Da for fragment ions, allowing up to three missed cleavages, and for the following peptide modifications: Carbamidomethylation of cysteine residues (+57.02 Da), dimethylation of lysine amines (+28.03 Da), variable methionine oxidation (+15.99 Da), and thioacylation (+88.00 Da) and dimethylation (+28.03 Da) of peptide N-termini. N-terminally thioacylated peptides identified by MS/MS represent prime-side cleavage products of the proteases of interest. The complete cleavage sites were reconstructed by bioinformatic determination of the non-prime-side sequences from the databases using WebPICS [56], available at http://clipserve.clip.ubc.ca/pics/. WebPICS generates a non-redundant list of identified prime side peptide sequences, matches these to the human UniProt protein database, and extracts the non-prime cleavage side sequence to the next cleavage site of the enzyme used for library generation, i.e. to the next N-terminal Arg or Lys for a trypsin-generated library. Subsite positions with ambiguous information coming from different protein isoforms are omitted and replaced by X, the one-letter notation for an unknown amino acid, for further analysis. Reconstructed cleavage sites were aligned and summarized as iceLogos [57].

2.9. In-silico structural analysis

Based on the crystal structures of collagenase G (pdb entries 4are [41] and 2y6i [36]) and human matrix metalloproteinase 1 (hMMP1; pdb entry 1hfc [58]) we performed peptide docking with the respective ideal substrates using the Rosetta FlexPepDock web server [59,60], allowing full flexibility to the peptide and side-chain flexibility to the protease. The results were optimized based on the available complex structures for the prime site, and the anti-parallel edge strand for nonprime substrate recognition. Molecular graphic figures were made using the molecular visualization system PyMOL (The PyMOL Molecular Graphics System, Version 1.5.03, Schrödinger, LLC) [61].

3. Results

3.1. Expression and purification of the peptidase domains of collagenase G, H and T

We cloned and purified the peptidase domain of ColG, spanning amino acids Asp398-Gly790 (Fig. 1A), based on the published crystal structures of the collagenase unit of ColG (pdb entries 2y3u, 2y50, 2y6i, and 4are) and the peptidase domains of ColH (pdb entries 4ar1, 4arf) and ColT (pdb entries 4ar8, 4ar9). Expression typically yielded 25 mg of highly pure and monodisperse protein after a 3-step purification scheme (from 1 L of cell culture). The purified recombinant protein migrated on a denaturing, non-reducing SDS-PAGE gel with an apparent molecular mass of 40 kDa, and was found to be at least 95% homogeneous (Fig. 2A). Dynamic light scattering analysis proved the protein to be highly monodisperse and suitable for downstream applications such as protease specificity profiling or crystallization (Supplementary Fig. 2). Although the obtained crystals were of insufficient quality for structure determination, they demonstrate the high quality and purity of the protein preparation. A 3D model based on the structure of the ColG
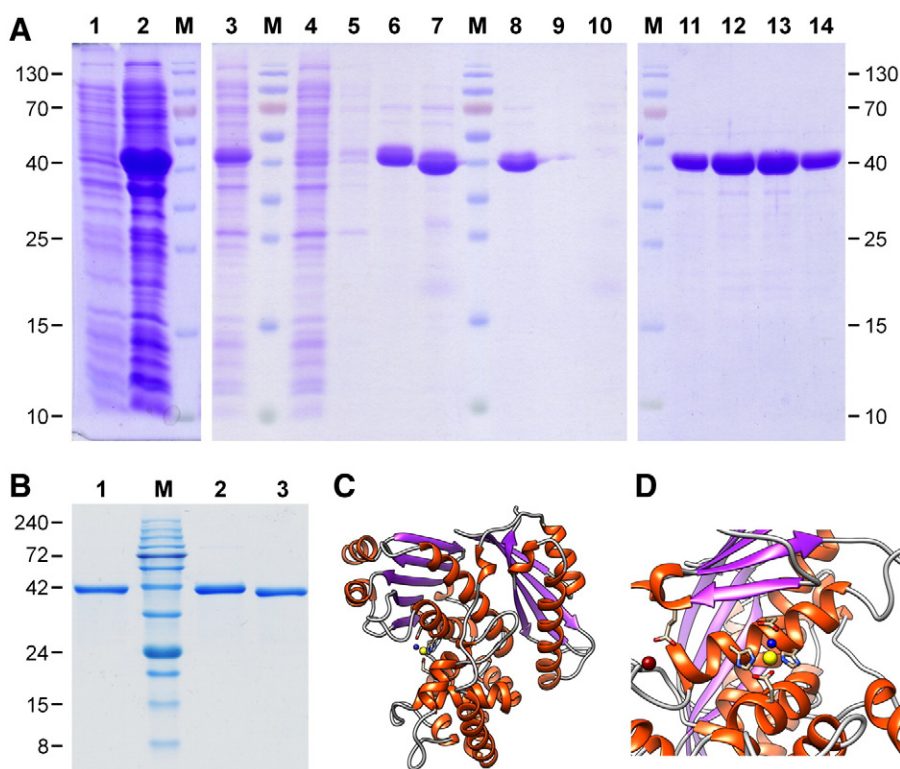
Fig. 2 – (A) SDS-PAGE monitoring of the expression and purification of the peptidase domain of ColG. M: PageRuler Prestained Protein Ladder (Fermentas); lane 1: BL21 DE3 cells before induction with 1 mM IPTG; lane 2: clarified lysate; lane 3: sample before immobilized metal ion affinity chromatography (IMAC) purification; lane 4: IMAC flow-through; lane 5: IMAC wash; lane 6: IMAC elution; lane 7: sample after His-tag removal; lane 8: flow-through of IMAC re-chromatography; lane 9: wash of re-chromatography; lane 10: elution of re-chromatography; lanes 11 to 14: sample after size exclusion chromatography. (B) SDS-PAGE analysis of the peptidase domains of ColG (lane 1), ColH (lane 2), and ColT (lane 3) used for protease specificity profiling. M: BLUeye Prestained Protein Ladder (GeneDireX). (C) Ribbon representation of the ColG peptidase domain based on the crystal structure of its collagenase unit and the crystal structure of the peptidase domain of ColT. (D) View towards the active side. Zinc binding ligands are shown in dark red, zinc and calcium are shown as yellow and red spheres, respectively.

collagenase unit (pdb entry 4are) and the ColT peptidase domain (pdb entry 4ar9) is shown in Figs. 2C and 2D to illustrate the overall shape and active site geometry of the target protein. Enzymatic activity was verified with the standard substrate for clostridial collagenases, N-[3-(2-Furyl)acryloyl]-Leu-Gly-Pro-Ala (FALGPA), and was in the same range as for the collagenase module of ColG (data not shown). The protein batches of the peptidase domains of ColH and ColT were of similar or even superior quality, and were previously used to determine their crystal structures [41].

3.2. Specificity profiling by Proteomic Identification of protease Cleavage Sites (PICS)

In a first step, we generated a PICS peptide library by digesting human cell lysates. This peptide library contained oligopeptides representing the biological amino acid diversity, with the constraint of Lys and Cys being modified by dimethylation and carbamidomethylation during library preparation, respectively. Purified recombinant peptidase domains of ColG, ColH and ColT were incubated with aliquots of the tryptic peptide library, prime-side cleavage products enriched by biotin-mediated affinity purification and identified by high-resolution mass

spectrometry. Following spectrum-to sequence assignment at a false discovery rate (FDR) of 1% and bioinformatic inference of the non-prime side sequences from the proteome database (Supplemental Tables S1, S2, and S3; Supplementary Fig. 1) we obtained 108, 153, and 183 unique cleaved P5 to P5′ sequences for ColG, ColH and ColT, respectively. For each protease, the cleaved peptide sequences were aligned at the cleavage site, occurrences of each amino acid at each position counted and summarized as heat map (Fig. 3, left panels). Normalization to the natural amino acid abundance in human protein database revealed the subsite preferences (Fig. 3, middle panels). The over- and under-representation of each amino acid was further analyzed by iceLogos [57], which additionally tested the significance of the observed changes from natural amino acid occurrence while accounting for sample size (Fig. 3, right panels). It should be noted that, due to the substrate specificity of trypsin, the employed library lacked peptides with internal Arg and Lys. However, since clostridial collagenases were previously shown not to prefer Arg/Lys in subsites S1′ and S2′ [62], we anticipate that our results accurately represent the overall sequence specificity. The number of interacting sites depends on the unique structure of the active site, but usually extends up to four or five amino acids in each direction.

All three analyzed clostridial collagenases showed comparable active site specificities, perfectly matching the repetitive amino acid sequence of collagen thus highlighting their collagenolytic nature (Fig. 3). Distinct preferences at positions P3 to P4′ were revealed, with common strong preferences for Gly in P3 and P1′ and for Pro in P2 and P2′, reminiscent of the typical collagen motif Gly-Pro-X. In addition, Pro was more than 4-fold enriched in P1, and at least 2.4-fold in P3′, indicating the capability to also accommodate the second most prominent collagen sequence pattern, Gly-X-Hyp. In line with these observations, P1′ was occupied by Gly in 57%, 68%, and 57% of all sequences cleaved by ColG, ColH, and Colt, respectively, with Ala accounting for additional 26%, 19%, and 26% at this position. In P2′, Pro was by far the most dominant amino acid, with 6 to 11 fold enrichment over its natural abundance. Besides Pro, large hydrophobic residues such as Phe, Trp, and Met, or large slightly polar ones such as Tyr and Gln, were moderately preferred in P1. In contrast, small polar residues like Ser or Thr as well as negatively charged residues were significantly underrepresented relative to their natural abundance at nearly every position throughout the specificity profiles.

Next, we investigated cooperative effects between Pro and the occurrence of other amino acids in subsites close to the cleavage sites. When only substrate peptides with Pro in P1 were considered, 60% of those identified for ColH and over 86% of the ColG and ColT cleavage sites showed Gly in P3, and more than 80% showed either Gly or Ala in P1′. In the case of ColH, 90% of the cleavage sites had Gly in P1′, whereas ColT allowed His/Met/Ser (19%) and Ala (27%) although Gly was the most favored (54%). The same picture held true when we considered sites with Pro in P2; 89%, 97%, and 83% of the identified substrate peptides for ColG, ColH, and ColT had Gly or Ala in P1′, and 92%, 81%, and 88% in P3 for ColH, ColG and ColT, respectively. Also 94%, 94%, 86% of the respective ColH, ColG and ColT identified cleavage sites with Pro in P2′ showed Gly or Ala in P1′. Remarkably, a reduced occurrence of Pro in P1 was observed for ColG and ColH cleavage sites where Ala was in P1′, while Pro increased in P2. In ColH this effect was most pronounced: The Pro occurrence dropped to 3% in P1, but increased to over 70% in P2. The opposite observation was made in peptides with Gly in P1′; Pro was more strongly enriched in P1 in all three cases. All the three obtained specificity profiles show the perfect adaption of clostridial proteases to collagen as their substrate. Stunningly, this collagen-like active site specificity pattern was possible even though none of the identified peptides originated from collagen itself, as it was not present in the proteome used for peptide library preparation.

3.3. Structural *in-silico* active site comparison of clostridial collagenases and their human counterparts, the matrix metalloproteinases

Up to now, 23 members of the human matrix metalloproteinase (MMP) family, sharing significant sequence homology, have been identified [63,64]. They are traditionally subdivided into: (i) collagenases (MMP1, 8, 13, and 18), (ii) gelatinases (MMP2 and 9), (iii) stromelysins (MMP3, 10, and 11), matrilysins (MMP7, and 26), and (v) membrane-type MMPs (MT-MMPs; MMP14, 15, 16, 17, 24, and 25), even though some of the MMPs (*e.g.* MMP12) do not fit into any of these functional groups [64,65]. We chose MMP1, also known as interstitial or fibroblast collagenase, for our structural comparison, as it was considered as the prototype for human collagenases [66–68].

First, we superimposed the active sites of collagenase G and MMP1, to provide an impression of their similarities and differences (Fig. 4A and B). Both zinc metalloproteinases share the zincin motif (HExxH) with the five-stranded β-sheet of the thermolysin-like peptidase (TLP) fold and present the canonical non-prime site substrate-recognition motif, the edge strand. Whereas clostridial collagenases belong to the gluzincin clan employing a glutamate as the third proteinaceous zinc ligand, MMPs are metzincins with a histidine as the third zinc ligand, and a conserved methionine containing 1,4-β-turn (Met-turn) forming the hydrophobic basement to the catalytic zinc [69,70]. In clostridial collagenases the structurally equivalent position is taken by a conserved alanine positioned on the so called glutamate helix [41,69]. Another noteworthy difference position of the third zinc ligand: In MMPs the third zinc ligand is located just 6 amino acids after the central HExxH motif, whereas clostridial collagenases contain a stretch of up to 30 amino acids long that not only shapes the active site cleft, but also builds the recently discovered calcium binding site crucial for enzymatic activity [41].

In a second step we performed peptide docking experiments using hexapeptide sequences. For collagenase G we used the sequence Gly-Pro-Pro-Gly(S1′)-Pro-Ala (Fig. 4C), as identified by our PICS analysis, which also resembles the prime site of the classical substrate for clostridial collagenase, FALGPA. For MMP1 we used the sequence Pro-Leu-Gly-Leu(S1′)-Ala-Gly (Fig. 4D), mimicking the general MMP substrate (QF-24; [71]). Most striking is their difference in the substrate recognition site S1′. MMPs contain a well-defined hydrophobic pocket within the lower catalytic subdomain, typically accommodating leucine or isoleucine at P1′ (Fig. 4F), resembling a critical determinant for substrate binding [72]. In contrast, clostridial collagenases possess a double Gly motif preceding the edge strand of the upper subdomain, forming a secondary oxyanion


Fig. 3 – PICS specificity profiles for collagenase G, H, and T. Alignments of 108 cleavage sites for the peptidase domain of ColG (A), 153 cleavage sites for ColH (B), and 183 cleavage sites for ColT are summarized as heat maps showing their relative occurrence (left panels) and the fold-change over the natural abundance of amino acids (middle panels), and as iceLogos [57] showing percent difference compared to natural amino acid abundance (right panels). Heat maps were created using Gnuplot (www.gnuplot.info). Single letter code for amino acid residues plotted is on the y axes. P and P′ subsite positions are shown on the x axes. In iceLogos, significantly over-represented amino acids are shown above the axis, under-represented residues below the axis, and amino acids that have not been identified are depicted in pink. For easier comparison, the same scales have been applied to all three proteases.

pocket by both backbone amides [36]. This helps to fix the carbonyl oxygen of a preferably small P1′ residue (Fig. 4E), such as glycine or alanine, to the top of the active site.

Notably, small amino acids are typically found at position P1 in MMP substrates [72]. In consequence, the subsites S2′ and S3′ in clostridial collagenases are horizontally mirrored
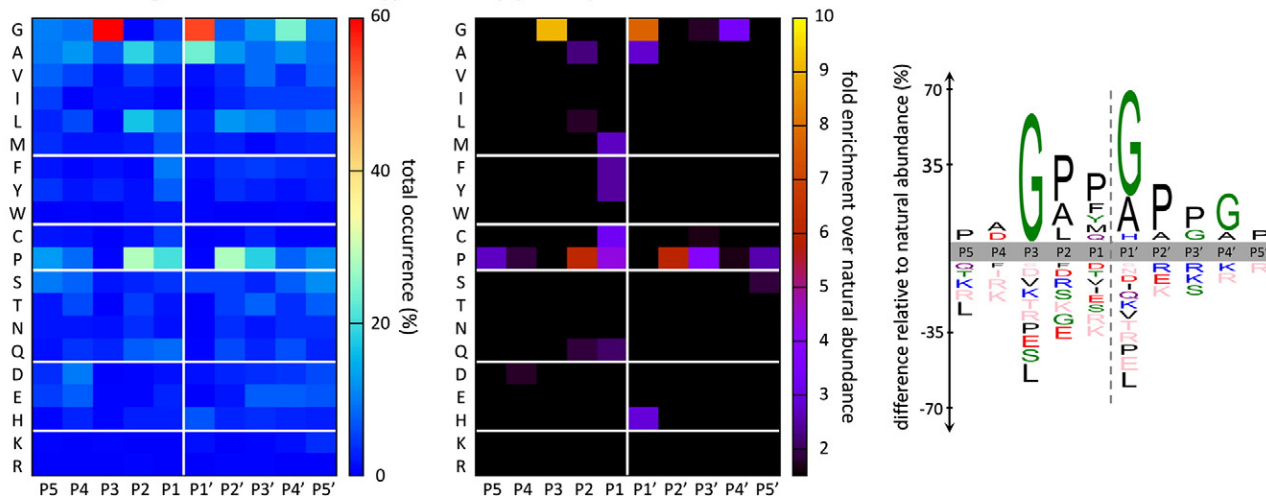

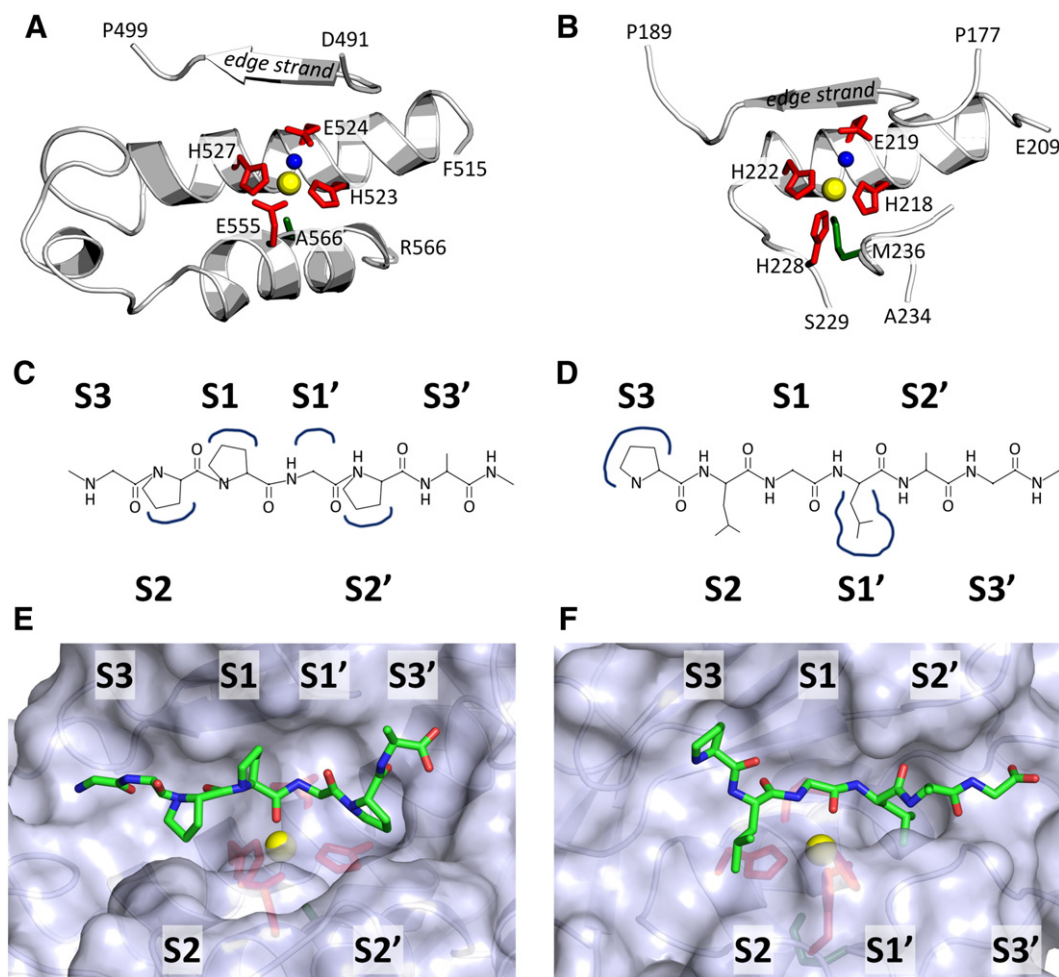**A ColG cleavage sites in a human tryptic library (n=108)**
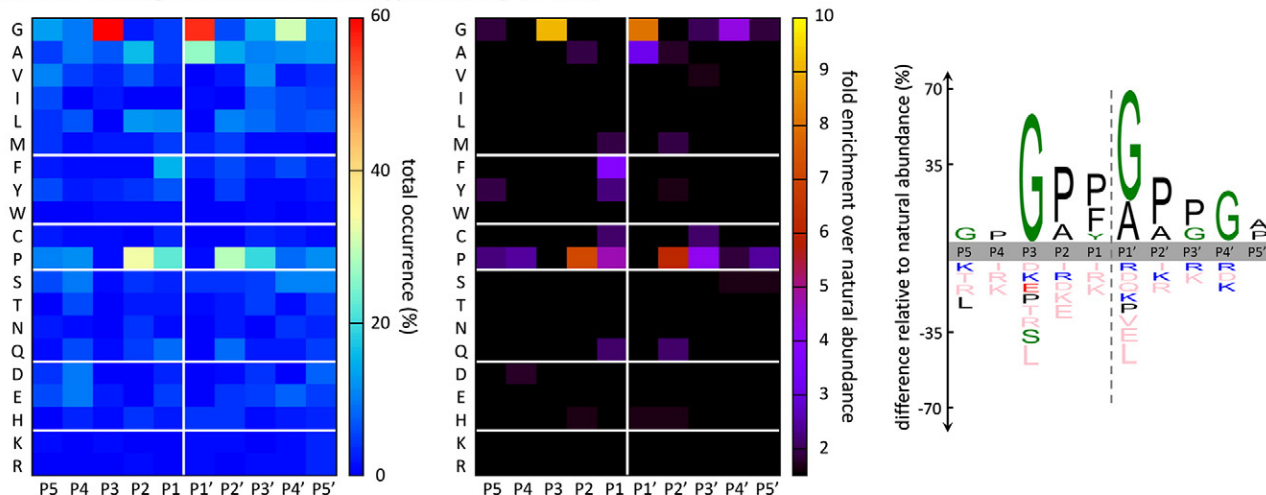
B ColH cleavage sites in a human tryptic library (n=153)
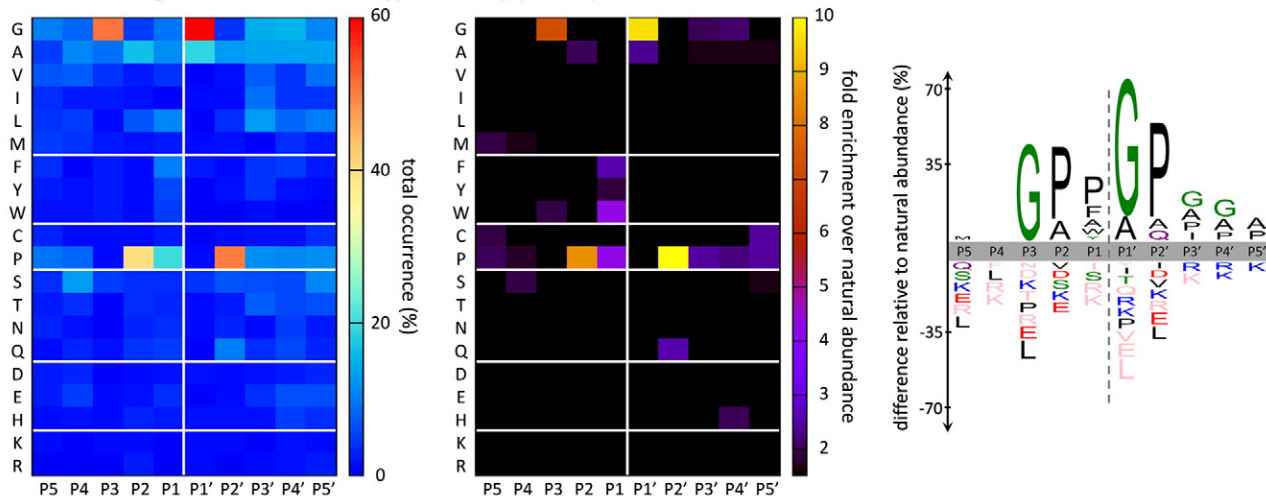
C ColT cleavage sites in a human tryptic library (n=183)

Fig. 4 – *In-silico* comparison of collagenase G (left) and human MMP1 (right). (A, B) Ribbon representation of the unliganded active site clefts. Key active site elements, such as the HExxH motif presenting central helix, edge strand, and hydrophobic basement are depicted. The active site zinc, tetrahedrally coordinated by three proteinaceous residues (ball and stick representation) and a water molecule (blue sphere) is shown in yellow. (C, D) Schematic representation of the hexapeptides used for active site docking. Left: Gly-Pro-Pro-Gly(P1′)-Pro-Ala, used for MMP1; right: Pro-Leu-Gly-Leu(P1′)-Ala-Gly, used for collagenase G. The corresponding substrate binding pockets are indicated. (E, F) Surface representation of the P3 to P3′ liganded active site clefts. Prominent hydrophobic bulges and the S1′ pocket of MMP1 shape the active site clefts.

compared to MMPs; subsite S2′ is provided by the lower catalytic subdomain, and S3′ by the upper. In addition, in the surface representation of our peptide docking result for collagenase G, two prominent bulges built up by a big aliphatic residue of the edge strand (Leu495) and an aromatic residue of the wall forming motif (Phe515), respectively, represent crucial determinants for subsites S1 and S2′, typically occupied by proline.

4. Discussion

In this study we demonstrate that the collagen specificity of clostridial collagenases is imprinted within the active site to an unusually high degree by matching active site subsites precisely with the commonly cleaved collagen Gly-Pro-X triplet motif. The unique and exquisite selectivity for these triplets likely explains the high kcat/kM of clostridial

collagenases for native collagen that augments collagen exosite binding by the collagen recruitment domain or the activator domain. Thus, the cleavage specificity profiles for the peptidase domains of the three different clostridial collagenases, namely of collagenases G and H from *C. histolyticum*, and ColT from *C. tetani*, showed a clear preferences for Pro in P2 and P2′, and Gly dominating in P3 and P1′.

Recent research has revealed the complex mechanism of clostridial collagenolysis on a molecular level and highlighted the different levels of action involved [7,27,36–38,41,73–78]: In a first step, C-terminally located collagen binding domains (CBDs) of clostridial collagenases recognize the triple-helical conformation of collagen, thus allowing binding and localization. In the second step, collagen fibers are swollen and disintegrated by cooperative action of up to two PKDs together with the CBDs, thereby exposing microfibrils and collagen monomers. Microfibrils up to 40 Å in diameter are clamped by the collagenase unit, mechanically expelling and unwinding

the triple-helical collagen molecules by a so-called "chew and digest" mechanism. One strand after the other is then aligned along the edge strand in the active site cleft of the peptidase domain and cleaved at the scissile bond, between hydroxyproline/proline (P1) and glycine (P1′). Subsequently these are further degraded to oligopeptides due to the additional tripeptidyl-carboxypeptidase activity of clostridial collagenases. In contrast, mammalian collagenolytic MMPs cleave collagen only at a single site between glycine (P1) and leucine/isoleucine (P1′), that is located in a temporarily unfolded, hypersensitive region of reduced hydroxyproline content, and otherwise buried in the heart of the collagen super helix [12,79,80]. Notably, unlike clostridial collagenases targeting the bond lying on the amino side of the glycine, MMPs target the bond nearest to the center of the triple helix which is more difficult to access and hydrolyze. This is reflected by particularly slow cleavage kinetics, such as kcat/km of just 66 $\mu M^{-1} h^{-1}$ for type I collagen and human MMP1 [81,82]. The question of why nature selected a more difficult site to cleave when a structurally more accessible bond is nearby cannot be empirically determined. We speculate that the importance of maintaining structural tissue integrity by preventing unwarranted extracellular cleavage of collagen outweighs the need for a high rate of cleavage. Extracellular cleavage of proteins is inherently under less cellular control than for intracellular or phagocytosed substrates. Thus, a slower rate of cleavage imparts a greater degree of control in maintaining fibril strength, despite active remodeling.

In our present study, we focused on the substrate specificity of the peptidase domains of the clostridial collagenases G, H, and T, as the activator domain and the collagen recruitment domains were previously shown to be dispensable for peptidolytic activity [36,40]. Further, exclusion of the collagen binding and activator domains served to minimize unexpected and undesirable exosite effects, such as broadening substrate specificity by providing additional contact areas. Strikingly, our analysis revealed eye-catching collagen-like specificity patterns harbored within the peptidase domain, with Gly dominating P3 and P1′, and Pro dominating P2 and P2′. A slightly looser specificity was observed at positions P1, typically occupied by hydroxyproline in collagen, with enrichment for large bulky amino acids such as phenylalanine, tyrosine and tryptophan next to the dominating proline. Bearing in mind the complexity of the prerequisites needing to be fulfilled before a single collagen strand becomes accessible for cleavage [13,36,80], this high sequence specificity for the typical collagen motifs Gly-Pro-X and Gly-X-Hyp is striking and reveals for the first time that the high selectivity for native collagen is specifically harbored within the peptidase domain of clostridial collagenases. This adaption to substrates utilizing side by side proline rigidity and glycine flexibility can be explained by the active site structure. The edge strand guided the non-prime substrate recognition sites S3 to S1 to form a wide cleft, allowing for residues such as Pro and Hyp that are not capable of adopting an ideal extended beta-strand conformation due to the cyclic nature of the side chains and thereby caused reduced backbone flexibility [36,83,84]. A prominent bulge in S1 induces a kink in the substrate upon binding, which can be perfectly adapted by proline and hydroxyproline even though their steric limitations. This rigidity, which also restricts the backbone

conformation of neighboring residues [85], is relaxed by the flexibility of Gly in P1′, which itself is fixed by the secondary S1′ oxyanion pocket, leading to substrate binding in a highly distorted conformation around the scissile bond needed for efficient hydrolysis [36,41]. Thus we conclude, that the combination of the secondary oxyanion pocket in S1′ and the hydrophobic bulge in S1 are in combination the primary active site determinants of substrate recognition, imprinting the tight specificity for collagen-like peptides in clostridial collagenases.

The differences between clostridial collagenases and mammalian MMP mechanisms of action are striking, but reflect the different *in vivo* roles of these proteases. Whereas clostridial collagenases are designed specifically to cleave and degrade collagens, MMPs regulate a variety of processes by precise proteolytic processing of many target proteins and so regulate cell migration, proliferation, adhesion and signaling, but not so much through matrix degradation rather through modulation of extracellular signaling networks [86]. Deregulation of these processes can result in pathological conditions *e.g.* cancer or fibrotic diseases such as liver and lung fibrosis, or Morbus Dupuytren [87–89]. Noteworthy, for the latter disease, clostridial collagenase injections were recently approved as therapeutics to break down the tough fibrous cords [29,90]. Contrary to the tightly regulated endogenous host processes, clostridia secrete collagenases to rapidly degrade collagen in a fast, highly processive and unrestricted manner, by cleaving collagens at multiple sites. This allows clostridia not only to utilize collagen as additional carbon resource but also to assists host invasion and colonization during anaerobic infection by tissue degradation.

Our study supports other recent reports highlighting the suitability of PICS for assaying protease specificity [56,91,92]. By appropriately choosing proteome sources and the digestion enzyme during the peptide library generation step, specificity of all protease classes can be determined in an unbiased and cost efficient manner. Despite the unique and exquisite cleavage site specificity of clostridial collagenases which restricts the number of peptides that can be cleaved and identified, PICS identified more than 100 unique cleavage sites for each of the peptidase domains, thus yielding robust sequence specificity logos.

Taken together, our active site specificity profiling of clostridial collagenases revealed their substrate specificity to be not only encoded by the collagen recruitment domains but also within the active site of the peptidase domains, indicating the molecular adaptation of these proteases to their peculiar substrate on several levels. Furthermore, our results point towards clostridial collagenases as dedicated extracellular matrix degraders, whether it is for decomposition of biomaterials or as virulence factors in pathogenesis, and towards a co-evolution of clostridial collagenases with their substrate, collagens.

5.   Conclusion

We used PICS and tandem mass spectrometry to define the prime and non-prime cleavage site specificities of the clostridial collagenases ColG and ColH from *C. histolyticum*, and ColT from *C. tetani*. Our results highlight the dedication of the peptidase domains to collagen-related peptides as substrates and represent an important piece in the complex puzzle of collagenolysis.

Together with previous and ongoing structural biology efforts, our findings pave the way for the rational development of better peptidic substrates and inhibitors, and guide the engineering of collagenase variants specific to the needs of the respective biotechnological and biomedical application.

Author contributions

Competing interests

The authors declare to have no competing interests.

Funding

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.10.004.

R E F E R E N C E S

[1] Traub W, Yonath A, Segal DM. On the molecular structure of collagen. Nature 1969;221:914–7.

[2] Hulmes DJ, Miller A. Quasi-hexagonal molecular packing in collagen fibrils. Nature 1979;282:878–80.

[3] Hulmes DJ. The collagen superfamily—diverse structures and assemblies. Essays Biochem 1992;27:49–67.

[4] Perumal S, Antipova O, Orgel JPRO. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. Proc Natl Acad Sci U S A 2008;105:2824–9.

[5] Shoulders MD, Raines RT. Collagen structure and stability. Annu Rev Biochem 2009;78:929–58.

[6] Ramshaw JA, Shah NK, Brodsky B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. J Struct Biol 1998;122:86–91.

[7] Orgel JPRO, Irving TC, Miller A, Wess TJ. Microfibrillar structure of type I collagen in situ. Proc Natl Acad Sci U S A 2006;103:9001–5.

[8] Wagenaar-Miller RA, Engelholm LH, Gavard J, Yamada SS, Gutkind JS, Behrendt N, et al. Complementary roles of intracellular and pericellular collagen degradation pathways in vivo. Mol Cell Biol 2007;27:6309–22.

[9] Verzijl N, DeGroot J, Thorpe SR, Bank RA, Shaw JN, Lyons TJ, et al. Effect of collagen turnover on the accumulation of advanced glycation end products. J Biol Chem 2000;275:39027–31.

[10] Fields GB. Interstitial collagen catabolism. J Biol Chem 2013;288:8785–93.

[11] Overall CM. Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. Mol Biotechnol 2002;22:51–86.

[12] Balbín M, Fueyo A, Knäuper V, Pendás AM, López JM, Jiménez MG, et al. Collagenase 2 (MMP-8) expression in murine tissue-remodeling processes. Analysis of its potential role in postpartum involution of the uterus. J Biol Chem 1998;273:23959–68.

[13] Manka SW, Carafoli F, Visse R, Bihan D, Raynal N, Farndale RW, et al. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. Proc Natl Acad Sci U S A 2012;109:12461–6.

[14] Nagase H, Fushimi K. Elucidating the function of non catalytic domains of collagenases and aggrecanases. Connect Tissue Res 2008;49:169–74.

[15] Tam EM, Moore TR, Butler GS, Overall CM. Characterization of the distinct collagen binding, helicase and cleavage mechanisms of matrix metalloproteinase 2 and 14 (gelatinase A and MT1-MMP): the differential roles of the MMP hemopexin c domains and the MMP-2 fibronectin type II modules in collagen triple helicase activities. J Biol Chem 2004;279:43336–44.

[16] Pelman GR, Morrison CJ, Overall CM. Pivotal molecular determinants of peptidic and collagen triple helicase activities reside in the S3' subsite of matrix metalloproteinase 8 (MMP-8): the role of hydrogen bonding potential of ASN188 and TYR189 and the connecting cis bond. J Biol Chem 2005;280:2370–7.

[17] Garnero P, Borel O, Byrjalsen I, Ferreras M, Drake FH, McQueney MS, et al. The collagenolytic activity of cathepsin K is unique among mammalian proteinases. J Biol Chem 1998;273:32347–52.

[18] Kafienah W, Brömme D, Buttle DJ, Croucher LJ, Hollander AP. Human cathepsin K cleaves native type I and II collagens at the N-terminal end of the triple helix. Biochem J 1998;331(Pt 3):727–32.

[19] Hatheway CL. Toxigenic clostridia. Clin Microbiol Rev 1990;3:66–98.

[20] Wells CL, Wilkins TD. Clostridia: sporeforming anaerobic bacilli. In: Baron S, editor. Med MicrobiolGalveston (TX): University of Texas Medical Branch at Galveston; 1996.

[21] Wessel D, Flügge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 1984;138:141–3.

[22] Bond MD, Van Wart HE. Purification and separation of individual collagenases of *Clostridium histolyticum* using red dye ligand chromatography. Biochemistry (Mosc) 1984;23:3077–85.

[23] Toyoshima T, Matsushita O, Minami J, Nishi N, Okabe A, Itano T. Collagen-binding domain of a *Clostridium histolyticum* collagenase exhibits a broad substrate spectrum both in vitro and in vivo. Connect Tissue Res 2001;42:281–90.

[24] French MF, Bhown A, Van Wart HE. Identification of *Clostridium histolyticum* collagenase hyperreactive sites in type I, II, and III collagens: lack of correlation with local triple helical stability. J Protein Chem 1992;11:83–97.

[25] Mookhtiar KA, Van Wart HE. Clostridium histolyticum collagenases: a new look at some old enzymes. Matrix Suppl 1992;1:116–26.

[26] Brüggemann H, Gottschalk G. Insights in metabolism and toxin production from the complete genome sequence of *Clostridium tetani*. Anaerobe 2004;10:53–68.

[27] Mookhtiar KA, Steinbrink DR, Van Wart HE. Mode of hydrolysis of collagen-like peptides by class I and class II *Clostridium histolyticum* collagenases: evidence for both endopeptidase and tripeptidylcarboxypeptidase activities. Biochemistry (Mosc) 1985;24:6527–33.

[28] Kaplan FTD. Collagenase clostridium histolyticum injection for the treatment of Dupuytren's contracture. Drugs Today (Barc) 2011;47:653–67.

[29] Desai SS, Hentz VR. The treatment of Dupuytren disease. J Hand Surg 2011;36:936–42.

[30] Shi L, Carson D. Collagenase Santyl ointment: a selective agent for wound debridement. J Wound Ostomy Continence Nurs 2009;36:S12–6.

[31] Ramundo J, Gray M. Collagenase for enzymatic debridement: a systematic review. J Wound Ostomy Continence Nurs 2009;36:S4–S11.

[32] Brandhorst H, Brandhorst D, Hesse F, Ambrosius D, Brendel M, Kawakami Y, et al. Successful human islet isolation utilizing recombinant collagenase. Diabetes 2003;52:1143–6.

[33] Balamurugan AN, Breite AG, Anazawa T, Loganathan G, Wilhelm JJ, Papas KK, et al. Successful human islet isolation and transplantation indicating the importance of class 1 collagenase and collagen degradation activity assay. Transplantation 2010;89:954–61.

[34] Matsushita O, Okabe A. Clostridial hydrolytic enzymes degrading extracellular components. Toxicon 2001;39:1769–80.

[35] Ducka P, Eckhard U, Schönauer E, Kofler S, Gottschalk G, Brandstetter H, et al. A universal strategy for high-yield production of soluble and functional clostridial collagenases in *E. coli*. Appl Microbiol Biotechnol 2009;83:1055–65.

[36] Eckhard U, Schönauer E, Nüss D, Brandstetter H. Structure of collagenase G reveals a chew-and-digest mechanism of bacterial collagenolysis. Nat Struct Mol Biol 2011;18:1109–14.

[37] Eckhard U, Brandstetter H. Polycystic kidney disease-like domains of clostridial collagenases and their role in collagen recruitment. Biol Chem 2011;392:1039–45.

[38] Wang YK, Zhao GY, Li Y, Chen XL, Xie BB, Su HN, et al. Mechanistic insight into the function of the C-terminal PKD domain of the collagenolytic serine protease deseasin MCP-01 from deep sea Pseudoalteromonas sp. SM9913: binding of the PKD domain to collagen results in collagen swelling but does not unwind the collagen triple helix. J Biol Chem 2010;285:14285–91.

[39] Bond MD, Van Wart HE. Characterization of the individual collagenases from *Clostridium histolyticum*. Biochemistry (Mosc) 1984;23:3085–91.

[40] Eckhard U, Schönauer E, Ducka P, Briza P, Nüss D, Brandstetter H. Biochemical characterization of the catalytic domains of three different clostridial collagenases. Biol Chem 2009;390:11–8.

[41] Eckhard U, Schönauer E, Brandstetter H. Structural basis for activity regulation and substrate preference of clostridial collagenases G, H, and T. J Biol Chem 2013;288:20184–94.

[42] Gomis-Rüth FX, Stöcker W, Huber R, Zwilling R, Bode W. Refined 1.8 A X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. J Mol Biol 1993;229:945–68.

[43] Schechter I, Berger A. On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 1967;27:157–62.

[44] Auf dem Keller U, Schilling O. Proteomic techniques and activity-based probes for the system-wide study of proteolysis. Biochimie 2010;92:1705–14.

[45] Schilling O, Overall CM. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 2008;26:685–94.

[46] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999;112:531–52.

[47] Van Wart HE, Steinbrink DR. A continuous spectrophotometric assay for *Clostridium histolyticum* collagenase. Anal Biochem 1981;113:356–65.

[48] Schilling O, Huesgen PF, Barré O, Auf dem Keller U, Overall CM. Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. Nat Protoc 2011;6:111–20.

[49] Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 2005;4:2010–21.

[50] Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 2008;24:2534–6.

[51] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 2004;20:1466–7.

[52] Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999;20:3551–67.

[53] Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002;74:5383–92.

[54] Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics 2011;10 [M111.007690].

[55] Keller A, Eng J, Zhang N, Li X, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol 2005;1 [2005.0017].

[56] Schilling O, Auf dem Keller U, Overall CM. Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites. Biol Chem 2011;392:1031–7.

[57] Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. Nat Methods 2009;6:786–7.

[58] Spurlino JC, Smallwood AM, Carlton DD, Banks TM, Vavra KJ, Johnson JS, et al. 1.56 A structure of mature truncated human fibroblast collagenase. Proteins 1994;19:98–109.

[59] London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O. Rosetta FlexPepDock web server—high resolution modeling of peptide-protein interactions. Nucleic Acids Res 2011;39: W249–53.

[60] Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins 2010;78:2029–40.

[61] DeLano WL. The case for open-source software in drug discovery. Drug Discov Today 2005;10:213–7.

[62] Hu Y, Webb E, Singh J, Morgan BA, Gainor JA, Gordon TD, et al. Rapid determination of substrate specificity of *Clostridium histolyticum* beta-collagenase using an immobilized peptide library. J Biol Chem 2002;277:8366–71.

[63] Dufour A, Overall CM. Missing the target: matrix metalloproteinase antitargets in inflammation and cancer. Trends Pharmacol Sci 2013;34:233–42.

[64] Scozzafava A, Ilies MA, Manole G, Supuran CT. Protease inhibitors. Part 12. Synthesis of potent matrix metalloproteinase and bacterial collagenase inhibitors incorporating sulfonylated N-4-nitrobenzyl-beta-alanine hydroxamate moieties. Eur J Pharm Sci 2000;11:69–79.

[65] Nagase H, Visse R, Murphy G. Structure and function of matrix metalloproteinases and TIMPs. Cardiovasc Res 2006;69:562–73.

[66] Bauer EA, Eisen AZ, Jeffrey JJ. Immunologic relationship of a purified human skin collagenase to other human and animal collagenases. Biochim Biophys Acta 1970;206:152–60.

[67] Goldberg GI, Wilhelm SM, Kronberger A, Bauer EA, Grant GA, Eisen AZ. Human fibroblast collagenase. Complete primary structure and homology to an oncogene transformation-induced rat protein. J Biol Chem 1986;261:6600–5.

[68] Pardo A, Selman M. MMP-1: the elder of the family. Int J Biochem Cell Biol 2005;37:283–8.

[69] Gomis-Rüth FX, Botelho TO, Bode W. A standard orientation for metallopeptidases. Biochim Biophys Acta 1824;2012:157–63.

[70] Hooper NM. Families of zinc metalloproteases. FEBS Lett 1994;354:1–6.

[71] Knight CG, Willenbrock F, Murphy G. A novel coumarin-labelled peptide for sensitive continuous assays of the matrix metalloproteinases. FEBS Lett 1992;296:263–6.

[72] Maskos K. Crystal structures of MMPs in complex with physiological and pharmacological inhibitors. Biochimie 2005;87:249–63.

[73] Ohbayashi N, Matsumoto T, Shima H, Goto M, Watanabe K, Yamano A, et al. Solution structure of clostridial collagenase H and its calcium-dependent global conformation change. Biophys J 2013;104:1538–45.

[74] Ohbayashi N, Yamagata N, Goto M, Watanabe K, Yamagata Y, Murayama K. Enhancement of the structural stability of full-length clostridial collagenase by calcium ions. Appl Environ Microbiol 2012;78:5839–44.

[75] Bauer R, Wilson JJ, Philominathan STL, Davis D, Matsushita O, Sakon J. Structural comparison of ColH and ColG collagen-binding domains from *Clostridium histolyticum*. J Bacteriol 2013;195:318–27.

[76] Philominathan STL, Koide T, Matsushita O, Sakon J. Bacterial collagen-binding domain targets undertwisted regions of collagen. Protein Sci 2012;21:1554–65.

[77] Wilson JJ, Matsushita O, Okabe A, Sakon J. A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation. EMBO J 2003;22:1743–52.

[78] Eckhard U, Nüss D, Ducka P, Schönauer E, Brandstetter H. Crystallization and preliminary X-ray characterization of the catalytic domain of collagenase G from *Clostridium histolyticum*. Acta Crystallogr Sect F Struct Biol Cryst Commun 2008;64:419–21.

[79] Lu KG, Stultz CM. Insight into the degradation of type-I collagen fibrils by MMP-8. J Mol Biol 2013;425:1815–25.

[80] Chung L, Dinakarpandian D, Yoshida N, Lauer-Fields JL, Fields GB, Visse R, et al. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. EMBO J 2004;23:3020–30.

[81] Welgus HG, Jeffrey JJ, Eisen AZ. The collagen substrate specificity of human skin fibroblast collagenase. J Biol Chem 1981;256:9511–5.

[82] Hasty KA, Jeffrey JJ, Hibbs MS, Welgus HG. The collagen substrate specificity of human neutrophil collagenase. J Biol Chem 1987;262:10048–52.

[83] Ramachandran GN, Lakshminarayanan AV, Balasubramanian R, Tegoni G. Studies on the conformation of amino acids. XII. Energy calculations on prolyl residue. Biochim Biophys Acta 1970;221:165–81.

[84] MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. J Mol Biol 1991;218:397–412.

[85] Schimmel PR, Flory PJ. Conformational energies and configurational statistics of copolypeptides containing L-proline. J Mol Biol 1968;34:105–20.

[86] Butler GS, Overall CM. Updated biological roles for matrix metalloproteinases and new "intracellular" substrates revealed by degradomics. Biochemistry (Mosc) 2009;48:10830–45.

[87] Cox TR, Erler JT. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Dis Model Mech 2011;4:165–78.

[88] Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol 2012;196:395–406.

[89] Shi F, Harman J, Fujiwara K, Sottile J. Collagen I matrix turnover is regulated by fibronectin polymerization. Am J Physiol Cell Physiol 2010;298:C1265–75.

[90] Gilpin D, Coleman S, Hall S, Houston A, Karrasch J, Jones N. Injectable collagenase *Clostridium histolyticum*: a new nonsurgical treatment for Dupuytren's disease. J Hand Surg 2010;35:2027–2038.e1.

[91] Becker-Pauly C, Barré O, Schilling O, Auf dem Keller U, Ohler A, Broder C, et al. Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates. Mol Cell Proteomics 2011;10 [M111.009233].

[92] Biniossek ML, Nägler DK, Becker-Pauly C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. J Proteome Res 2011;10:5363–73.