Developmental and Comparative Immunology 129 (2022) 104337

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/devcompimm

Chicken cathepsin G-like - A highly specific serine protease with a peculiar tryptase specificity expressed by chicken thrombocytes

Zhirong Fu [a], Srinivas Akula [a], Anna-Karin Olsson [b], Lars Hellman [a,*]

[a] Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden
[b] Department of Medical Biochemistry and Microbiology, BMC, Box 589, SE-751 23, Uppsala, Sweden

ARTICLE INFO

Keywords:
Chicken
Birds
Serine protease
Cleavage specificity
Tryptase
Thrombocyte
Evolution

ABSTRACT

Serine proteases are major granule constituents of cells from several mammalian hematopoietic cell lineages. Despite the relatively extensive knowledge about these mammalian proteases, very little is known about their bird, reptile and amphibian homologs. In order to close this gap in our understanding of the evolution of these proteases, we have characterized the extended cleavage specificity and hematopoietic expression pattern of the chicken serine protease cathepsin G-like. This protease, which clusters in a separate subfamily of serine proteases among the vertebrate hematopoietic serine proteases, has been characterized using substrate phage display and further validated by using a panel of recombinant substrates. A preference for a lysine in the P1 position of a substrate, arginines in positions P2 and P3, and the aromatic amino acid tryptophane in the P4 position was observed. Based on the sequence alignment we could identify a consensus sequence for this protease as being PGGWRRK$^{\downarrow}$ALSV. Mass spectrometry analysis of a peptide with the consensus sequence obtained by phage display showed that cleavage of this peptide occurred after the conserved Lys (K) residue. A screening of potential *in vivo* substrates based on the derived P5–P3' consensus sequence resulted in a relatively limited number of potential substrates, due to the high selectivity of this enzyme. The most interesting of these were PDGF-A, coagulation factor V and low-density lipoprotein receptor like-8. Immunohistochemical analysis of chicken white blood cells with antisera produced against chicken cathepsin G-like and chicken egg lysozyme, as a reference protein known to be expressed by hematopoietic cells, showed presence of chicken cathepsin G-like almost exclusively in thrombocytes whereas lysozyme was found at very high amounts in heterophils, and lower amounts in monocytes and thrombocytes.

1. Introduction

Serine proteases are important granule components of a number of mammalian hematopoietic cells. Human and murine mast cells express and store large amounts of both chymotryptic and tryptic serine proteases named chymases and tryptases (Hellman and Thorpe, 2014; Akula et al., 2015; Caughey, 2011; Galli and Tsai, 2010; Pejler et al., 2010). Mammalian neutrophils have primarily serine proteases with elastase and chymase specificities but also a tryptic enzyme whereas the T and NK cells express tryptases, asp-ases, chymases and metases (Hellman and Thorpe, 2014; Akula et al., 2015). The picture in mammals is becoming relatively detailed concerning the presence, amounts and specificities of these mammalian proteases. However, in non-mammalian vertebrates the information is still very fragmentary. We have recently analysed the specificity of two distantly related hematopoietic serine proteases expressed by NK-cell like cells from the Channel catfish (*Ictalurus punctatus*), but except for these very little is known about non-mammalian serine proteases expressed by hematopoietic cells (Thorpe et al., 2016). In order to close this gap in our understanding of the appearance and diversification of these enzymes, we have in the current study characterized the extended cleavage specificity and hematopoietic expression pattern of a serine protease belonging to a small subfamily of serine proteases encoded from a locus that so far only have been found in the genomes of amphibians, reptiles and birds (Fig. 1) (Akula et al., 2015). Members of this locus form a separate branch in the phylogenetic tree of hematopoietic serine proteases (Figs. 1 and 2) (Akula et al., 2015). No information concerning primary or extended specificity of any of these proteases have so far been available and nothing is known about in which cell or cell type they are expressed. We therefore decided to start the analysis of this new

* Corresponding author. Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE 751 24, Uppsala, Sweden.
E-mail address: Lars.hellman@icm.uu.se (L. Hellman).

https://doi.org/10.1016/j.dci.2021.104337
Received 22 November 2021; Received in revised form 13 December 2021; Accepted 13 December 2021
Available online 15 December 2021

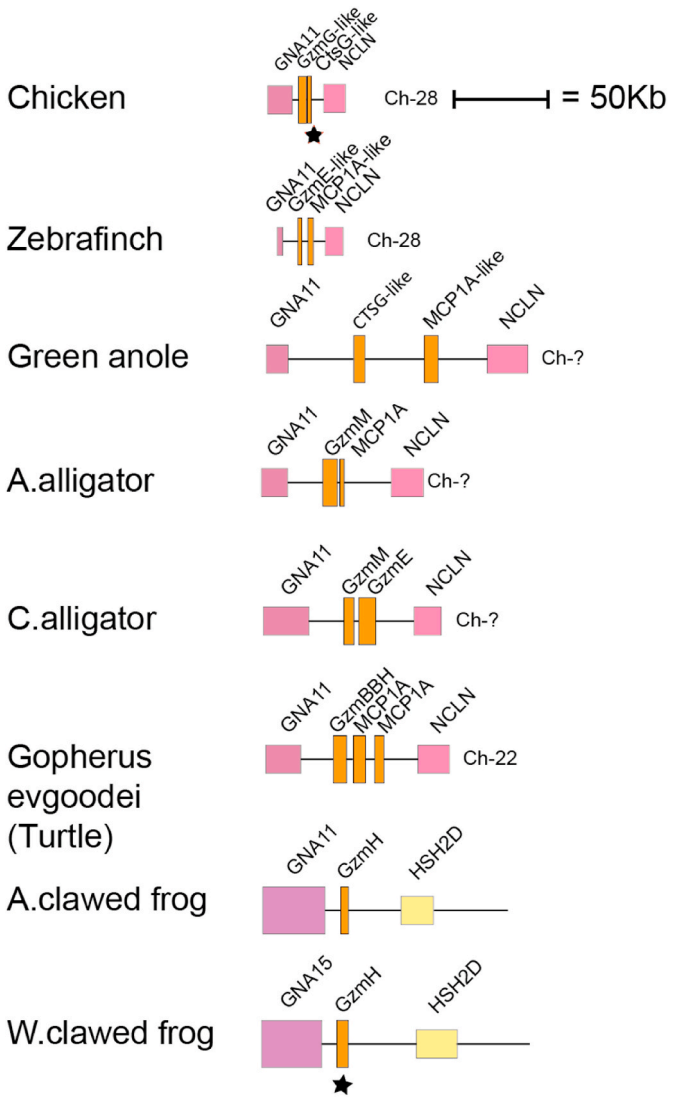

Fig. 1. The Amphibian, reptile and bird locus encoding chicken cathepsin G-like.
An in-scale figure of a locus present in birds, reptiles and amphibians but that has been lost in mammals. The protease genes within this locus are shown in orange. The genes for which we have produced recombinant proteins are marked by black stars. The protease of interest in this study, chicken cathepsin G-like, is one of the two proteases marked by a black star.

subfamily by the characterization of the chicken cathepsin G-like. In spite of its name, this protease is not more related to cathepsin G in mammals than the other mammalian hematopoietic serine proteases (Fig. 2). The name is likely derived from the initial alignments made by the sequencing consortium. This protease is a classical trypsin/chymotrypsin-related serine protease with the characteristic catalytic trial His-Asp-Ser (Schechter and Berger, 1967). Information concerning primary specificity can often be obtained by looking at amino acid residues that form the active site (S1) pocket of a serine protease of this family. By comparison with the corresponding region in bovine chymotrypsin, focusing on the residues 189, 216 and 226 chymotrypsinogen numbering, one can obtain clues to its primary specificity. These three residues have been found to be positioned in the bottom and the sides of the pocket thereby determining the type of amino acid that is favoured as primary target for cleavage by an enzyme (Schechter and Berger, 1967). However, for more distantly related members of this subfamily of proteases the analogy with the well characterized mammalian representatives becomes very weak and does not result in any information of value. This situation was seen for related fish proteases (Thorpe et al., 2016). This indicates that the structure is so different that the position of these residues not fully match the position they have in the mammalian enzyme. During the analysis of two catfish proteases, we observed this situation very clearly. The triplets of these two catfish enzymes are very similar but the specificity is completely different. One of them is a highly specific met-ase whereas the other is a highly specific tryptase, with preference for several consecutive basic amino acids, three or four in a row ((Thorpe et al., 2016) and manuscript in preparation). We have previously made several attempts to analyse several amphibian, reptile and bird proteases without success. The information concerning residues 189, 216 and 226 of enzymes encoded from this amphibian, reptile and bird locus did thereby not seem trustworthy, even if the triplets of all members of this locus presented in Fig. 1 display the DGG triplet indicating a tryptase specificity, and a detailed analysis had therefore to be performed by other methods. We here present the analysis of an enzyme named chicken cathepsin G-like by phage display and a new type of recombinant substrates. The result shows that this enzyme is a highly specific tryptase with preference for two arginines and a lysine residue preceded by the aromatic amino acid tryptophane (WRRK). The high selectivity of this enzyme made it possible to screen the chicken proteome for potential targets. The most interesting potential targets were PDGF-A, coagulation factor V and low-density lipoprotein receptor like-8. We also produced an antiserum in rats against this protease and chicken egg lysozyme as reference protein. Using these antisera in an immunohistochemical analysis of chicken white blood cells we could detect expression of chicken cathepsin G-like exclusively in chicken thrombocytes. In contrast chicken lysozyme was expressed and stored in high amounts in granules of chicken heterophils but also found in lower amounts in monocytes and thrombocytes. The results from the analysis of cathepsin G-like were unexpected as none of the mammalian hematopoietic serine proteases is found in platelets or megacaryocytes, the mammalian corresponding cells. The role of chicken cathepsin G-like in chicken thrombocyte biology is therefore still largely unresolved and needs in depth analysis and screening for biological relevant *in vivo* substrates.

2. Materials and methods

2.1. Construction of expression vector and production of recombinant protein for chicken cathepsin G-like

The chicken cathepsin G-like sequence was extracted from the NCBI database (Accession number XP_423728) and an expression construct was designed and ordered from GenScript (Piscataway, NJ, USA). The synthesized open reading frame for the active chicken cathepsin G-like starting with the sequence encoding the IIGGHE of the active protease was cloned into the EcoRI and XhoI sites of the pCEP-Pu2 vector, used for expression in mammalian cells (Vernersson et al., 2002). The enzyme was produced as an inactive recombinant protein, with an *N*-terminal His_6-tag followed by an EK site. The HEK 293 cells were grown to 70% confluency in 25 cm^2 tissue culture flasks (BD VWR) with Dulbecco's Modified Eagles Medium (DMEM) (GlutaMAX, Invitrogen, Carlsbad, CA, USA) supplemented with 5% fetal bovine serum (FBS) and 50 μg/ml gentamicin. Following transfection using 25 μg of chicken cathepsin G-like in pCEP-Pu2 with lipofectamine (Invitrogen, Carlsbad, CA, USA), the cells were allowed to rest for a few days to express the selection marker, until puromycin was added (0.5 μg/ml) to select for cells which had taken up the DNA. Heparin (5 μg/ml) was added to the culture medium to enhance yield of the secreted protein. Cells were then expanded and the conditioned media was collected for purification of the enzyme.

To purify the recombinant enzyme, 750 ml conditioned media was filtered to remove cell debris and other small insoluble particles (Munktell 00H 150 mm, Falun, Sweden) and 500 μl nickel nitrilotriacetic acid (Ni-NTA) beads was added. Following mixing by rotation for

45 min at 4 °C, the Ni-NTA beads were collected by centrifugation and transferred to a column containing a glass filter (Sartorius, Goettingen, Germany). After washing with PBS containing 0.05% Tween, 10 mM imidazole and 1 M NaCl, the recombinant protein was eluted in PBS Tween (0.05%) containing 100 mM imidazole. The volume of the first fraction was 200 µl and for the subsequent fractions we used 400 µl. Individual fractions were run on SDS-PAGE gel. Bovine serum albumin (BSA) was used as standard to estimate the concentration of the protein in each fraction. The fractions with high mount of protein was pooled and the concentration was estimated by separation on SDS-PAGE gels using BSA as standard, where-after the protein was kept at 4 °C or frozen at −80 °C to be used in subsequent experiments.

2.2. Activation of recombinant chicken cathepsin G-like

The concentration of the recombinant chicken cathepsin G-like was determined by SDS-PAGE as described above and the level of EK (Roche, Mannheim, Germany) was adjusted for proper activation of the enzyme. In the experiment shown in Fig. 3, 100 µl of the eluted recombinant enzyme was digested with 3 µl EK (Roche, Mannheim, Germany) for 3 h at 37 °C. The activated enzyme was stored at 4 °C until use.

2.3. Substrate phage display

A T7 phage library containing 5×10^7 variants, each displaying a unique nine amino acid sequence was used to determine the extended cleavage specificity of the chicken cathepsin G-like enzyme. The nine amino acid-region has been inserted into the *C*-terminal of the capsid 10 protein, followed by a His_6-tag. Approximately 10^9 plaque forming units (pfu) were bound to 125 µl Ni-NTA agarose beads via their His_6-tags for 1 h at 4 °C with gentle rotation. Unbound phages were removed by washing ten times with PBS containing 0.05% Tween 20 and 1 M NaCl, followed by two washes with PBS. The beads were re-suspended in 375 µl PBS and approximately 250 ng of recombinant chicken cathepsin G-like was added. This reaction was incubated for 2 h at 37 °C with gentle rotation, allowing cleavage of the susceptible phages and their subsequent detachment from the Ni-NTA beads. The supernatant containing released phages was then recovered after centrifugation. Thirty µl was used in a plaque assay to determine the number of released phages. Briefly, ten-fold serial dilutions were made, mixed with (*E.coli*) BLT5615 (for propagation and visualization of plaques on a bacterial lawn) and plated on LA-Amp (50 µg/ml) plates, incubated for 2.5 h at 37 °C and then counted. The remaining supernatant was added to a 10 ml culture of *E. coli* BLT5615 (OD_{600} 0.5) and the culture was incubated under

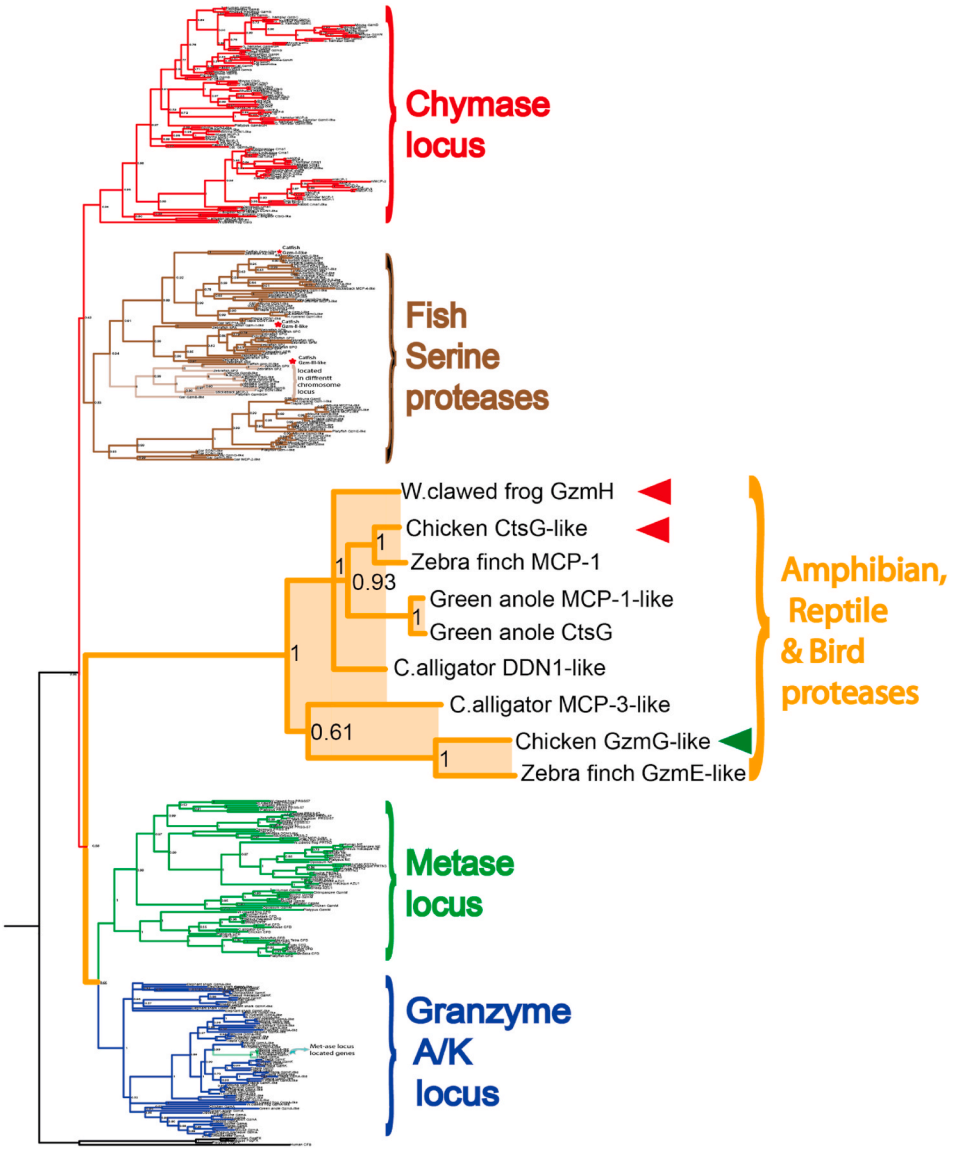

Fig. 2. Phylogenetic analysis of a panel of vertebrate hematopoietic serine proteases. A panel of hematopoietic serine proteases originating from fish, amphibian, reptiles, birds and mammals were aligned using the MrBase alignment program. The proteases originating from different chromosomal loci appear in separate branches on the tree. The different branches represent proteases from the mast cell chymase locus, the fish locus, the met-ase locus, the T cell tryptase locus also named the granzyme A/K locus, and the relatively newly identified locus encoding related sequences to the other hematopoietic serine proteases, a locus only identified in amphibians, reptiles and birds but not in mammals (Akula et al., 2015). It is in this locus, which is enlarged in the figure, we find the protease of this study, chicken cathepsin G-like and also corresponding proteases in other birds, in reptiles as represented by Green anole, and Gopherus turtle, and in amphibians as represented by two frog species, American and Western clawed frog. The two proteases within this locus that we have produced recombinant proteins of are marked with red arrows, the chicken cathepsin G-like and the Western clawed frog GzmH. The chicken cathepsin G-like is the protease in focus of this study. The Western clawed frog enzyme we have not yet succeeded in determining its specificity. A green star marks another Chicken enzyme encoded from this locus that is located very near cathepsin G-like in the chicken locus, but which has not yet been analysed for its cleavage specificity (Fig. 1).

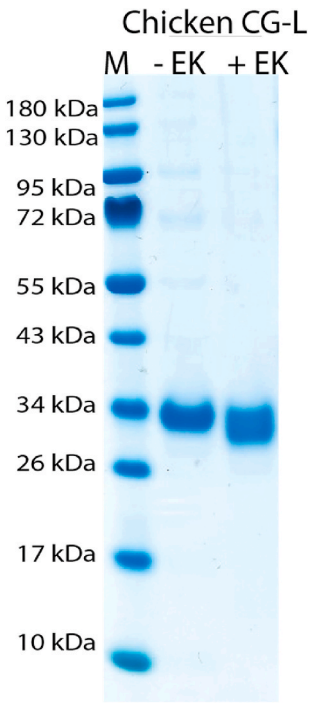

Fig. 3. Recombinant chicken cathepsin G-like. The enzyme was produced as an inactive protein (left lane) in HEK293-EBNA cells with a *N*-terminal His_6-tag and an enterokinase (EK) site, facilitating purification and activation, respectively. Addition of EK results in cleavage and the removal of the His_6 -tag and the EK site, generating an active enzyme and a subsequent reduction in size (right lane). Ten ul of the active and ten of the inactive enzyme was run on a 4–12% pre-cast SDS-PAGE gel (Invitrogen, Carlsbad, USA) and stained with Coomassie brilliant blue. As can be seen from the figure, approximately 60% of the protein has been activated by EK cleavage in this experiment.

vigorous shaking for 75 min at 37 °C for phage expansion. From this culture 1.5 ml was removed and centrifuged to remove bacterial debris. 800 μl of this supernatant was then transferred to a microcentrifuge tube containing 100 μl PBS and 100 μl 5M NaCl. This solution was placed at 4 °C overnight and bound to 125 μl Ni-NTA beads the following morning (i.e., start of the next selection cycle). The complete process was repeated 6 times, constituting in total 7 rounds of selection. Individual plaques were isolated from the plates of the final selection round as plugs using a Pasteur pipette and transferred into 100 μl phage buffer before vortexing for 30 min and stored at 4 °C. The random nine amino acid regions contained in these phages were amplified by PCR (T7-Select primers, Novagen, USA) and sequenced at GATC Biotech Sequencing centre (Cologne, Germany). The resulting sequences were translated using CLC viewer and aligned using Adobe Illustrator. A parallel control reaction without enzyme (only PBS) was also run under the same conditions and plaque numbers were compared to the enzyme sample.

2.4. Peptide cleavage analysis

A consensus sequence derived from the phage display data was produced as a synthetic peptide with the sequence Pro-Gly-Gly-Trp-Arg-Arg-Lys-Ala-Leu-Ser-Val, estimated >95% pure by HPLC (GeneCust, Dudelange, Luxembourg). The peptide was dissolved in distilled water to 5 mg/ml and 7.5 μg was used in a cleavage reaction. The cleavage reaction consisted of approximately 100 ng of active chicken cathepsin G-like in a 10 μl volume, in PBS buffer. The reaction was analysed using mass spectrometry after a 37 °C incubation taken at 0 min and 60 min, with subsequent enzyme inactivation by heating for 15 min at 65 °C. A control run without enzyme was also used.

2.5. Phage display sequence verification using a two-thioredoxin approach

In order to verify the phage display data, a recombinant 2xTrx system developed in our laboratory was used. Here, the random nine amino acid region from the phage display was introduced between two adjacent *E. coli* thioredoxin coding regions. Originally, a pET21 vector containing the coding region for a single trx protein was modified to contain a second trx with BamHI and SalI sites in the intervening region. Here the random region was synthesized as oligonucleotides (Sigma, St Louis, USA), which were ligated and inserted between a BamHI and a SalI restriction site. This resulted in a vector containing a first trx followed by the random cleavable region then a second trx with His_6-tag (facilitating purification).

This construct was expressed in *E.coli* Rosetta gami (Novagen). Ten ml of an overnight culture was added to 90 ml LB + Amp (50 μg/ml) and 500 μl 20% glucose. After approximately 1 h (reaching OD_{600} 0.5), 100 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and the culture placed on a shaker (moderate shaking) at 37 °C for 3 h. The culture was pelleted by centrifugation at 10,000 rpm for 3 min and the supernatant was discarded. The pellet was washed in 10 ml PBS containing 0.05% Tween, centrifuged and pelleted again followed by resuspension in 1/100th starting volume (i.e., 1 ml) of PBS. To obtain the intracellularly expressed protein, the resuspended pellet was sonicated for 6 × 30 s on ice. The supernatant was transferred to a new microcentrifuge tube after centrifugation at 10,000 rpm for 10 min at 4 °C. To purify the recombinant protein 125 μl Ni-NTA beads were added and incubated for 45 min at 4 °C with gentle agitation. The solution with Ni-NTA beads was transferred to a 2 ml column (Terumo, Leuven, Belgium) containing a glass filter (Sartorius, Goettingen, Germany) and subsequently washed with 3 × 2 ml and 2 × 1 ml PBS containing 0.05% Tween and 10 mM imidazole. To elute the protein from the column we used PBS containing 0.05% Tween and 100 mM imidazole. Fractions were collected after passing through the column. The first fraction volume was half the Ni-NTA bead volume (75 μl) and further fractions eluted with a full bead volume (150 μl). Individual fractions were run on SDS-PAGE gel and their concentrations were estimated based on the staining intensity of a BSA standard. The fractions containing high amounts of protein were pooled and the final protein concentration was determined by the use of a Bradford assay (Bio-rad, CA, USA) and then kept at 4 °C for direct use or frozen in −80 °C for long term storage.

For cleavage analysis, 2 μl of the EK activated recombinant chicken cathepsin G-like was added to 20 μg of the pooled 2xTrx protein in 50 μl total volume and aliquots of 5 μg removed after 0, 15, 45 and 150 min after enzyme addition. The reactions were run at room temperature and the samples taken at different time points were analysed on SDS-PAGE gel under denaturing conditions using pre-cast 4–12% Bis-Tris gels (Invitrogen, Carlsbad, CA, USA) and 1x MES buffer (Invitrogen, Carlsbad, CA, USA). Gels were stained with colloidal Coomassie blue solution to visualize the protein bands (Neuhoff et al., 1988).

2.6. Screening for potential in vivo substrates

The derived consensus sequence (Pro-Gly-Gly-Trp-Arg-Arg-Lys-Ala-Leu-Ser-Val) and variants of this sequence was used in a screening for potential *in vivo* substrates using a standard protein Basic Local Alignment Search Tool (BLASTp).

2.7. Immunohistochemical analysis of the expression pattern of chicken cathepsin G-like and hen egg lysozyme

To determine the site of expression for chicken cathepsin G-like an antiserum against this protease was produced by injection of approximately 50 μg of this enzyme emulsified with Freunds complete adjuvant subcutaneously in rats. Animal work was approved by the local animal ethics committee (C11/14). Two booster doses with the same amount of

protein were given two weeks apart in Freunds incomplete adjuvant. After 6 weeks of immunization 5–10 ml of blood was collected from the rats during anaesthesia and the rats were subsequently euthanized. The blood was allowed to coagulate over night at +4 °C. The tubes were centrifuged and the serum was transferred to new tubes, aliquoted and stored at −80 °C until use. In parallel with the generation of the anti-serum against chicken cathepsin G-like we also made an anti-serum against hen egg lysozyme as a reference for another protein expressed by hematopoietic cells, one rat per protein.

The majority of blood cells are the erythrocytes. To be able to look at expression of chicken cathepsin G-like and lysozyme in white blood cells we therefore needed to remove the majority of the red blood cells. This was done by magnetic cell sorting using an anti-chicken CD45 antibody and MACS beads (Miltenyi Biotech, Bergisch Gladbach, Germany). Fresh blood in Alsevers solution (50:50) was delivered to us from Håtunalab (Uppsala, Sweden). One ml of the blood cells in Alsevers was centrifuged at 1500 rpm for 5 min and the supernatant was removed. The pellet was then resuspended in 400ul of staining buffer (PBS pH 7.2 and 1%BSA) containing 4ug/ml of a mouse IgM anti-chicken CD45-APC antibody (MA5-28677, Thermofisher, Waltham MA, USA) and left at RT for 20 min in dark. The tube was then centrifuged at 1500 rpm and the supernatant removed. Six ml washing buffer, was then added, the same as staining buffer, mixed followed by centrifugation at 1500 rpm for 5 min. The pellet was resuspended in 480 μl staining buffer. 120 μl of anti-mouse IgM beads was then added (130-047-302, Miltenyi Biotech, Bergisch Gladbach, Germany), and incubated for 20 min at RT in dark. The cells were pelleted at 1500 rpm for 5 min and then resuspend in 5 ml staining buffer. The cell suspension was then passed over a LS column, that had been pre-equilibrated by passing 3 ml staining buffer over the column, followed by a washing step with 3 times 3 ml of staining buffer. The column was the removed from the magnet and the cells were flushed out of the column by pushing 5 ml staining buffer through the column. The cells were the concentrated by a centrifugation step, 300g for 10 min at +4 °C and the pellet was then dissolved in 600 μl of staining buffer. Ten μl of the cells were analysed with trypan blue for viability.

Approximately 100 μl of the cell suspension from the MACS purification was spun onto a single glass slides with a cytospin centrifuge, 500 rpm at RT for 5 min. After cytospin the slides were allowed to dry for a few minutes. The cells were then fixated directly by submerging the slides in 4% PFA (paraformaldehyde) in PBS pH7.4 for 10 min at RT. The slides were then washed 2 times with cold PBS and stored at +4 °C in humid chamber in a fridge until use.

Before histochemical staining the cells were permeabilized by submerging the slides in a PBS pH 7.4 solution containing 0.25% Triton X-100 for 10 min at RT. The slides were then washed 3 times on a shaker table for 5 min with PBST (PBS + 0.05% Tween 20). To block endogenous peroxidase activity the slides were incubated with a 3% H_2O_2 solution in distilled water for 20 min at RT followed by three washes in PBST for 5 min each on a shaker. Then the slides were incubated with blocking buffer for 1 h at RT in PBST containing 1% BSA and 22.52 mg glycine/ml followed by three washes in PBST for 5 min each at a shaker table. The primary antibody was added, initially at a 1/50 dilution and later also tested at a higher concentration by using a 1/10 dilution in PBS+1%BSA, and incubated at RT for 1 h, followed by three washes in PBST for 5 min each on a shaker. The secondary antibody, anti-rat IgG (H + L) (Vector laboratories, Burlington, CA, USA), diluted 1/500 in PBS+1%BSA was added and incubated at RT for 45 min, followed by two washes in PBST for 5 min each on a shaker. Streptavidin horse radish peroxidase (SA-HRP) (Vector laboratories, Burlington, CA, USA) was then added at a dilution of 1/500 in PBS and incubated for 30 min at RT, followed by two washes in PBST for 5 min each on a shaker. AEC (3-amino-9-ethylcarbazole) staining was then performed for 15 min at RT with Vector laboratories AEC kit (SK-4200), followed by a 5 min wash in water. The cells were then counter stained for 20 s with Mayer's Hematoxylin (Histolab. Västra Frölunda, Sweden) followed by rinsing under running water for 5–10 min. The glasses were then mounted with Flouromount-G (Southern Biotech, Birmingham, AL, USA).

3. Results

3.1. Production, purification and activation of chicken cathepsin G-like

The coding region for the active chicken cathepsin G-like was designed and ordered from Genscript and cloned into the mammalian expression vector pCEP-Pu2 for expression in the human embryonic kidney cell line, HEK-293 EBNA. The coding region for the chicken cathepsin G-like was in this clone preceded by an *N*-terminal His_6-tag for easy purification, followed by an enterokinase (EK) site to enable removal of the His_{6x}-tag after purification. By this *N*-terminal addition an inactive enzyme could be purified from the cell supernatant, and the purified protease could be activated by EK cleavage. Here, the His_6-tag facilitates purification on Ni^{2+} chelating immobilized metal ion affinity chromatography columns and cleavage with EK results in the activation of the enzyme by removing both the His_6-tag and the EK site (Fig. 3). Based on the result in Fig. 3 we estimate that approximately 60% of the enzyme of the original preparation was successfully cleaved, resulting is sufficient amount of active enzyme for the analysis of its cleavage specificity (Fig. 3).

3.2. Substrate phage display

To determine the extended cleavage specificity of chicken cathepsin G-like a phage T7-based system was used where individual peptide sequences are displayed on the surface of the phage. This system enables the characterization of a region covering 4–5 amino acids both upstream and downstream of the cleavage site. The library used has a complexity of approximately 50 million different peptide sequences. After seven rounds of selection, the phages selected by chicken cathepsin G-like showed a 3352-fold increase compared to the PBS control. One hundred and twenty phage plaques from the last selection round were picked and the region encoding the peptide sequence was amplified using PCR. Ninety-six of the most clean PCR fragments were sequenced. After decoding the random linker region of the sequenced clones, the amino acid sequences were aligned. Each row in Fig. 4 represents an individually sequenced random region. The alignment showed an apparent preference for several positively charged amino acids in and around the cleavage site, indicating tryptase activity (Fig. 4). Tryptophane was also found to be strongly favoured in what later was shown to be in the P4 position. We also observed a strong preference for Arg in the P3, Arg, Ala or Ser in the P2 position, and a preference for Lys over Arg in the P1 position. There was also a relatively strong preference for Leu in what later was found to be the P2′position (Fig. 4). A consensus sequence could be extracted from the phage display result being PGGWRRKALSV or PGGWRAKVLSV.

3.3. Peptide cleavage analysis

To identify the exact cleavage site, a peptide was synthesized with one of the consensus regions determined from phage display (PGGWRRKALSV). This peptide was cleaved for various durations and analysed using mass spectrometry (Fig. 5). After 1 h, the full-length peptide was cleaved into two smaller products with respective sizes of 388.46 and 855.99. The larger peptide migrated as a comparatively smaller size fragment due to its high charge +3, +4, which complicates the analysis slightly as the sizes not directly match but need to be charge compensated (Fig. 5). However, the size of the smaller fragments gives strong additional support for the cleavage after the Lys residue resulting in two fragments PGGWRRK and ALSV (Fig. 5).

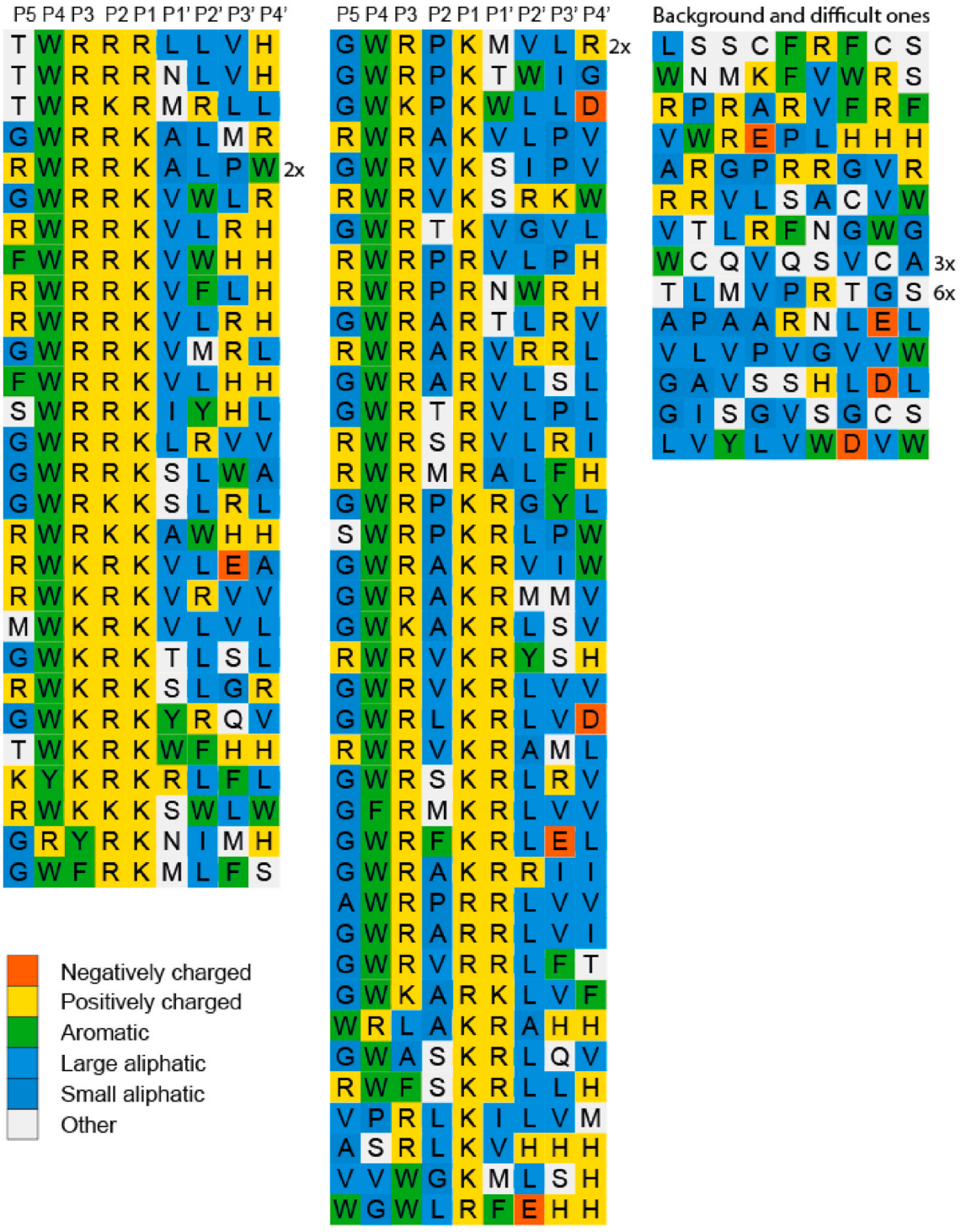

Fig. 4. Phage display analysis of the cleavage specificity of chicken cathepsin G-like. The sequences originating from the nine amino acid random region of the individual phages after seven cycles of pannings with chicken cathepsin G-like and the A T7 phage library were aligned to obtain the extended cleavage specificity of this enzyme. The sequences show a strict preference for Lys or Arg with a strong preference for Lys in the P1 position, indicating tryptase activity. Multiples of identical sequences are indicated to the right as 'x' number of times. A consensus (from P5–P3′) of Pro-Gly-Gly-Trp-Arg-Arg-Lys-Ala-Leu-Ser-Val was identified.

3.4. Verification of the sequence pattern obtained from the phage display by using recombinant substrates

To validate the phage display sequence data and to address variations of amino acids in the aligned phages we used a new system based on recombinant protein substrates that has been developed in our lab. In this system the consensus sequence obtained from the phage display analysis is introduced in the linker region between two *Escherichia coli* thioredoxin molecules by ligation of double stranded oligonucleotides in a vector containing the coding regions for these two thioredoxin molecules, sitting in tandem (Fig. 6A). This protein and a number of mutants originating from this consensus sequence were analysed by the cleavage of recombinant substrates in this two-thioredoxin (2xTrx) system (Figs. 6 and 7). This system has been used successfully to study a number of enzymes, resulting in that we have a library of substrates for many different types of enzymes with different primary specificities including

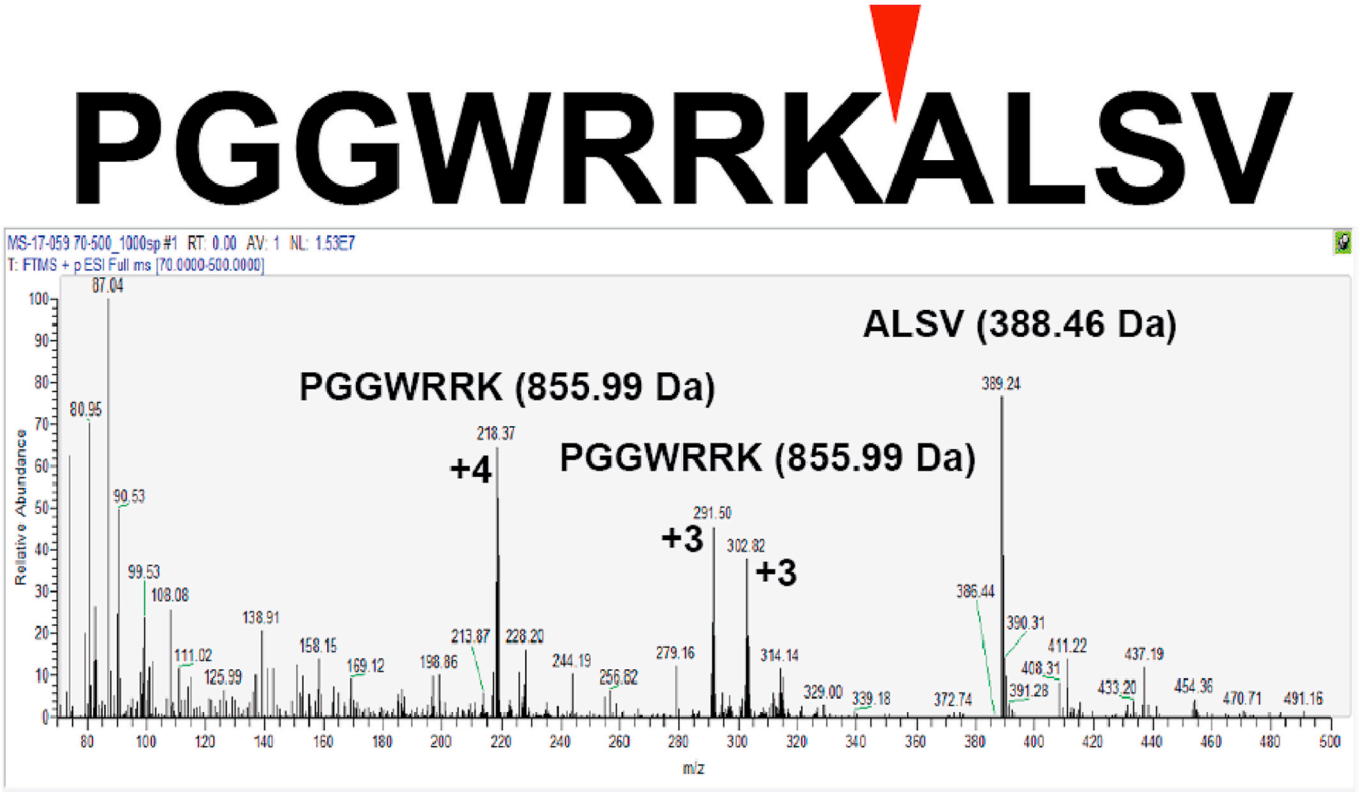

Fig. 5. Mass spectrometry peptide and potential *in vivo* substrate analyses. A) Mass spectrometry of a consensus-derived peptide. A synthetic peptide with the consensus-derived sequence, PGGWRRKALSV, was cleaved for 60 min at 37 °C and the resulting mixture was run using mass spectrometry. After 60 min, several peaks at MW 855.99 and 388.46 were identified as PGGWRRK and ALSV, respectively, indicating tryptase specificity with preference for Lys (lower panel). No cleavage in the control without enzyme was seen (not shown).

chymases, tryptases, elastases and asp-ases (Thorpe et al., 2012, 2016; Gallwitz et al., 2010, 2012; Andersson et al., 2010; Chahal et al., 2015; Fu et al., 2015). However, due to the high specificity of the chicken Cathepsin G-like enzyme, most of the 2xTrx substrates had to be newly designed and produced. Exchanging Ala for Arg in the P2 position and Lys with Arg in the P1 position resulted in a reduced cleavage rate by 2–3 fold (Fig. 6C). A similar reduction was seen when exchanging P3 Arg and P2 Ala with Lys (Fig. 6C). Moving the P4 Trp into position P3 resulted in a complete block in cleavage (Fig. 6C). As also seen from the phage display results, exchanging the Lys in the P1 position with an Arg resulted in a drop in cleavage by 4–5 times, showing a relatively strong preference for Lys in the P1 position (Figs. 4–6D). Introducing an Arg in the P1′position for the Val in the consensus sequence did not affect the cleavage rate (Fig. 6E). The same situation was observed when changing the Ala in the P2 position for a Ser (Fig. 6E). In contrast, when changing the Ala in the P2 position with the negatively charged glutamic acid (Glu) the cleavage dropped by more than 20 times (Fig. 6E). No difference was seen with Arg or Lys in the P3 position indicating that both amino acids are equally preferred (Fig. 6F). However, changing the Arg in the P3 position to an Ala resulted in a drop in cleavage by 4–5 fold, indicating that a positively charged residue in this position is highly preferred (Fig. 6F). We could also see that Trp was preferred over the other two aromatic amino acids, Phe and Tyr, by a factor of 6–7 and that Leu was even less favoured in this position (Fig. 7A). Exchanging this Trp for a Ser resulted in an even stronger reduction in cleavage rate (Fig. 7B). Interestingly also was the strong effect of changing the Leu in the P2′position for a Ser or a Glu, which almost completely blocked cleavage (Fig. 7B). Changing the position of the Trp or only having two Arg residues in the substrate was found to result in no cleavage, showing a high specificity of this enzyme (Fig. 7C).

3.5. *Screening for potential in vivo substrates*

The chicken proteome was screened with the consensus sequence PGGWRRKALSV or PGGWRAKVLSV and variants of these two sequences. This screening resulted in the identification of a list of potential targets. However, most of them were intracellular proteins and therefore less likely to be the targets of this presumably secreted protease, as it contains a signal sequence. The most interesting candidates among the extracellular proteins were chicken PDGF-A with the sequence KKWKRKKL, the low-density lipoprotein receptor related protein 8 with the sequence RNWKRKNT and two isoforms of coagulation factor V with the sequences NAWRAKSN and NAWRAKAQ.

3.6. *Histochemical analysis of the cell type specificity of chicken cathepsin G-like and of chicken lysozyme*

Due to almost complete absence of information concerning the tissue specificity of the different members of the amphibian, reptile and bird locus where chicken cathepsin G-like is located we decided to produce an antiserum against chicken cathepsin G for a histochemical analysis of its pattern of expression. As a reference protein we also produced an antiserum against chicken egg lysozyme. The recombinant chicken cathepsin G-like and chicken egg lysozyme were therefore injected into rats. After two booster injections with two-week intervals blood and subsequently serum was collected from the rats. The sera against chicken cathepsin G-like and chicken egg lysozyme were used to study the expression of these two proteins in chicken white blood cells. Chicken blood was obtained in Alsevers solution from commercial lab outside Uppsala. Red blood cells constitute the absolute majority of cells why the majority of them need to be removed for a better analysis of the white blood cells. Broiler chicken blood at week 7 has been estimated to

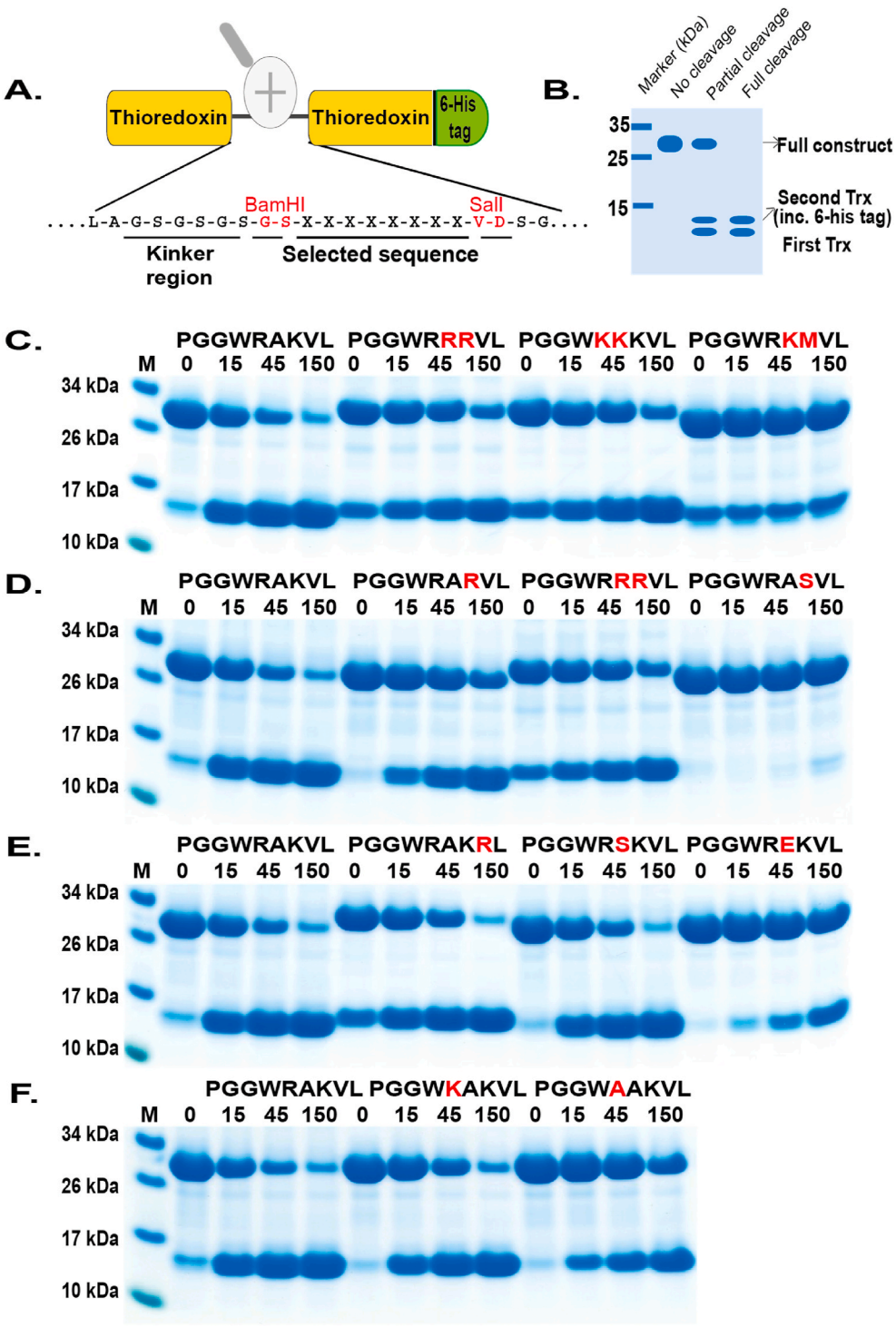

Fig. 6. Verification of phage display sequences using a two-trx system. The phage display-derived consensus sequence and a number of variants of this sequence were added in between two adjacent trx proteins (panel **A**), expressed in *E.coli* and subjected to chicken cathepsin G-like (panels ***C*–F**). The protein samples following cleavage at 0, 15, 45 and 150 min were run on pre-cast 4–12% SDS-PAGE gels (Invitrogen, Carlsbad, USA). Hypothetical cleavage is shown (panel **B**) to highlight possible cleavage patterns. The individual lanes represent various time points after addition of the enzyme, in minutes. The residues that have been changed in the various substrates from the consensus sequence PGGWRRKALSV are marked in red for more easy recognition.

contain 2.2×10^9 red blood cells and 8.9×10^6 white blood cells per ml (Onyishi et al., 2017). The white blood cells were positively selected for by MACs separation using a mouse anti-chicken CD45 antibody, which has been found to bind the majority of all chicken white blood cells, including the thrombocytes, but not to erythrocytes (De Boever et al., 2010; Seliger et al., 2012; Hofmann and Schmucker, 2021). This step removed the absolute majority of the red blood cells. The white blood cells were then spun onto glass slides by the use of a cytospin centrifuge. Following fixation by 4% paraformaldehyde (PFA) the samples were tested for different concentration of the two antisera to obtain the best concentration for reliable staining. Using a 1/50 times dilution of the anti-chicken cathepsin G-like serum and a 30 min incubation time at RT, the chicken cathepsin G-like serum almost exclusively stained chicken thrombocytes, whereas the lysozyme serum showed very strong staining for lysozyme in chicken heterophils and weaker staining in both monocytes and thrombocytes (Fig. 8D and F). Interestingly, when reducing the dilution of the serum to 1/10 we could see staining for chicken cathepsin G-like also in heterophils, indicating cross reactivity to another protease with similar amino acid sequence, most likely the granzyme G-like (Figs. 2 and 8E, green arrow head). The expression in the thrombocytes of both cathepsin G-like and lysozyme was confined to a ring around the nucleus indication ER staining and not granule storage

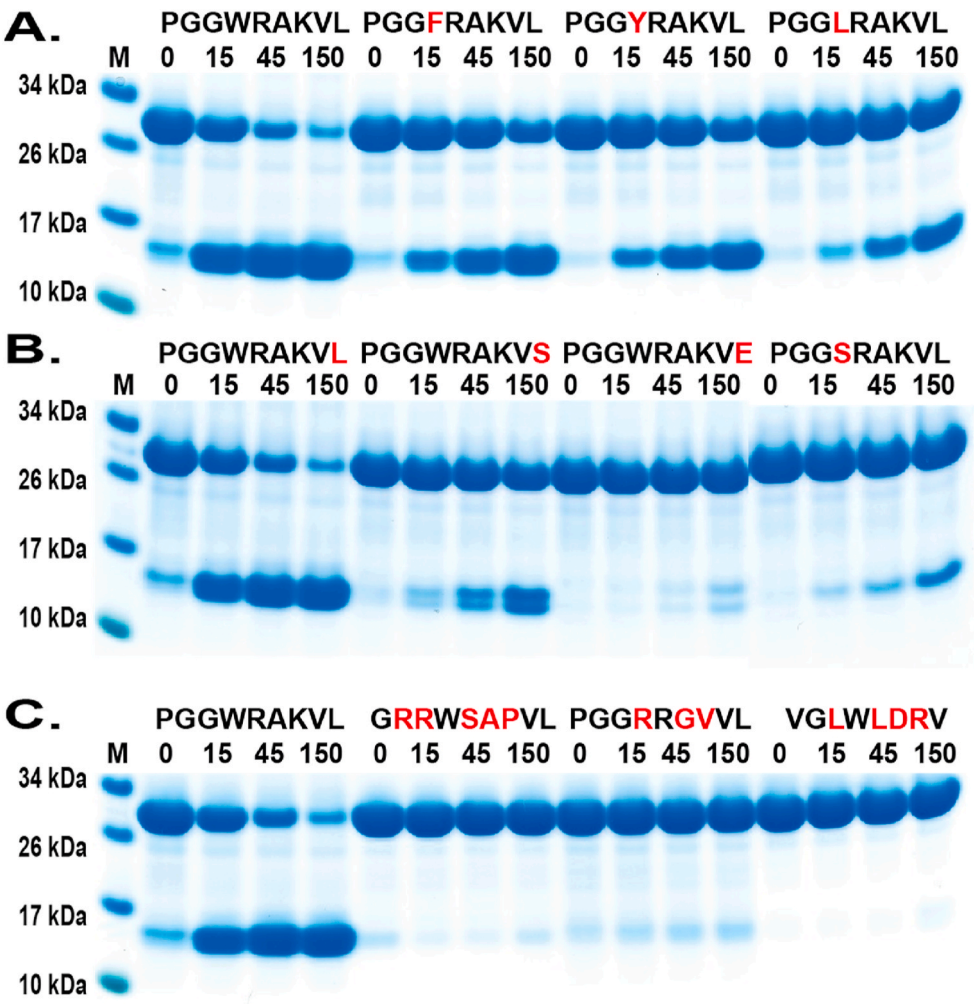

Fig. 7. Verification of phage display sequences using a two-trx system. The phage display-derived consensus sequence and a number of variants of this sequence were added in between two adjacent trx proteins expressed in *E.coli* and subjected to chicken cathepsin G (panels **A-C**). The protein samples following cleavage at 0, 15, 45 and 150 min were run on pre-cast 4–12% SDS-PAGE gels (Invitrogen, Carlsbad, USA). The individual lanes represent various time points after addition of the enzyme, in minutes. The residues that have been changed in the various substrates from the consensus sequence PGGWRAKVL are marked in red for easier recognition.

as seen for lysozyme in the heterophils (Fig. 8D and F). In parallel with the histochemical analysis, white blood cells were also analysed by Giemsa staining of non-fixated cells, clearly showing the prominent granule staining of the heterophils, but absence of cytoplasmic granules in both thrombocytes and monocytes (Fig. 8A and B).

4. Discussion

As a first step in the analysis of the presence, expression patterns and *in vivo* functions of the hematopoietic serine proteases in birds we have characterized one of the proteases expressed in this amphibian, reptile and bird specific locus (Akula et al., 2015). No information on any of the members of this locus has been presented previously, most likely because this locus has been lost in mammals and therefore not being part of any analysis of mammalian immunity. Surprisingly, this protease was found to be almost exclusively expressed by chicken thrombocytes. When reducing the dilution of the anti-chicken cathepsin G-like antiserum from 1/50 to 1/10 in the histochemical analysis we could observe staining also in heterophils and a faint staining of monocytes (Fig. 8E). This indicates cross-reactivity to another protein. The most likely candidate is here the granzyme G-like protease, that is marked with a green arrow in Fig. 2. The gene for this protease is located in the same locus as the cathepsin G-like gene, actually very close to each other (Fig. 1). The chicken granzyme G-like seems in contrast to chicken cathepsin G-like to primarily be expressed in chicken heterophiles and at lower levels in monocytes. This pattern of expression is similar to the mammalian situation where neutrophils, the counterpart to heterophiles in humans, express and store massive amounts of serine proteases, primarily cathepsin G, *N*-elastase and proteinase 3. Based on the histochemical analysis we can conclude that one of these two proteases, the cathepsin G-like, is expressed almost exclusively in thrombocytes and the other, the granzyme G-like, most likely primarily in heterophiles. We have also produced recombinant protein for a protease located in the same locus in Western clawed frog, the granzyme H gene (Figs. 1 and 2). We have tried several times with phage display to determine the cleavage specificity of this frog enzyme, however without success. We have also used the 2xTrx substrates shown in Figs. 6 and 7 that work well for the analysis of the chicken cathepsin G-like to test the frog enzyme, however also without success. This indicates that the frog enzyme has a different cleavage specificity compared to the chicken cathepsin G-like, despite their high sequence similarity (Fig. 2). We have experienced a similar situation when analysing a new subfamily of serine proteases, the duodenases, located in the chymase locus, another locus for hematopoietic serine proteases found in all tetrapods except birds (Akula et al., 2015, 2021). These genes have diversified extensively in cleavage specificity despite their very similar primary sequence (Fu et al., 2021). Among the duodenases we find two highly specific tryptases, one highly specific asp-ases, one pure chymase, and two dual enzymes with both tryptase and chymase specificity (Fu et al., 2021). A similar situation may exist in this amphibian, reptile and bird locus. Coming analyses of the cleavage specificity of additional members of this locus will hopefully shed light on this issue.

Chicken thrombocytes do, in contrast to their mammalian counterparts, have a nucleus. This is also a major difference between bird and

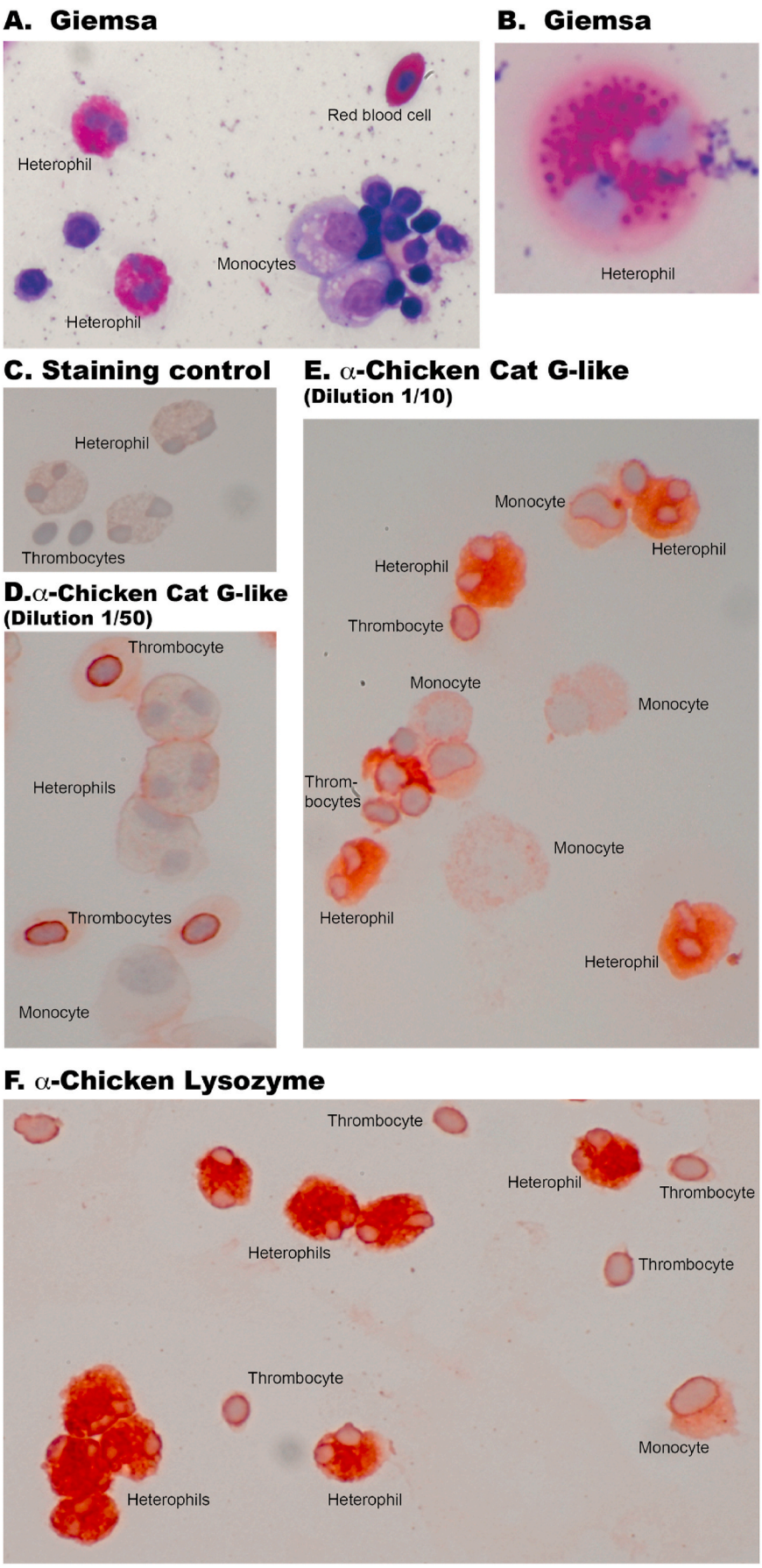

Fig. 8. Histochemical analysis of chicken white blood cells. Chicken white blood cells were purified by magnetic cell sorting using the MACS system and an antibody against chicken CD45. The cells were spun onto glass slides using a cytospin centrifuge. The cells were the fixated in a 4% PFA solution and stained with two different antisera one against chicken cathepsin G-like and one against chicken egg lysozyme. Cells from the MACs enrichment was also analysed by Giemsa staining of non-fixated cells as shown in panels **A** and **B**. Panel **C** shows the staining control with pre-immune rat serum. Panel **D** shows staining using the antiserum against chicken cathepsin G-like using a 1/50 dilution of the primary antibody, panel **E** shows staining using the antiserum against chicken cathepsin G-like using a 1/10 dilution of the primary antibody, and panel **F** shows the staining against chicken lysozyme (dilution 1/50).

mammalian red blood cells. Mammalian red blood cells lack a nucleus that is being expelled during the maturation of red blood cells at the erythroblast stage before becoming a reticulocyte (Moras et al., 2017; Menon and Ghaffari, 2021). Another difference is now the expression of a hematopoietic serine protease in chicken thrombocytes, something which seems to be lacking in mammalian platelets. The chicken cathepsin G-like was found to be a highly specific tryptase. Screening for potential *in vivo* substrates using the consensus target sequence, and variants of this sequence, resulted in only very few potential extracellular targets. One of them was coagulation factor V, another the low-density lipoprotein receptor related protein 8 and a third was PDGF-A. None of them can easily be seen as obvious *in vivo* targets. Further analysis will therefore be needed to identify biologically most relevant targets for this highly specific serine protease expressed almost exclusively as it appears by chicken thrombocytes. Does this protease have a role in enhancing coagulation, stabilizing a clot or another role in sealing wounds and blood leakage as this is one of the major functions of platelets in mammals? Or does it have a role in regulating angiogenesis, another role of mammalian platelets, or possibly a completely new function specific for thrombocytes in non-mammalian tetrapods? The future will hopefully shed light in this and numerous other questions concerning the role of the different hematopoietic serine proteases in chicken, reptile and amphibian immunity.

Acknowledgments

This study was supported by grants from the Knut and Alice Wallenberg Foundation KAW 2017.0022.

References

Akula, S., Thorpe, M., Boinapally, V., Hellman, L., 2015. Granule associated serine proteases of hematopoietic cells - an analysis of their appearance and diversification during vertebrate evolution. PLoS One 10 (11), e0143091.

Akula, S., Fu, Z., Wernersson, S., Hellman, L., 2021. The evolutionary history of the chymase locus -a locus encoding several of the major hematopoietic serine proteases. Int. J. Mol. Sci. 22 (20).

Andersson, M.K., Thorpe, M., Hellman, L., 2010. Arg143 and Lys192 of the human mast cell chymase mediate the preference for acidic amino acids in position P2' of substrates. FEBS J. 277 (10), 2255–2267.

Caughey, G.H., 2011. Mast cell proteases as protective and inflammatory mediators. Adv. Exp. Med. Biol. 716, 212–234.

Chahal, G., Thorpe, M., Hellman, L., 2015. The importance of exosite interactions for substrate cleavage by human thrombin. PLoS One 10 (6), e0129511.

De Boever, S., Croubels, S., Demeyere, K., Lambrecht, B., De Backer, P., Meyer, E., 2010. Flow cytometric differentiation of avian leukocytes and analysis of their intracellular cytokine expression. Avian Pathol. 39 (1), 41–46.

Fu, Z., Thorpe, M., Hellman, L., 2015. rMCP-2, the major rat mucosal mast cell protease, an analysis of its extended cleavage specificity and its potential role in regulating intestinal permeability by the cleavage of cell adhesion and junction proteins. PLoS One 10 (6), e0131720.

Fu, Z., Akula, S., Qiao, C., Ryu, J., Chahal, G., de Garavilla, L., et al., 2021. Duodenases are a small subfamily of ruminant intestinal serine proteases that have undergone a remarkable diversification in cleavage specificity. PLoS One 16 (5), e0252624.

Galli, S.J., Tsai, M., 2010. Mast cells in allergy and infection: versatile effector and regulatory cells in innate and adaptive immunity. Eur. J. Immunol. 40 (7), 1843–1851.

Gallwitz, M., Enoksson, M., Thorpe, M., Ge, X., Hellman, L., 2010. The extended substrate recognition profile of the dog mast cell chymase reveals similarities and differences to the human chymase. Int. Immunol. 22 (6), 421–431.

Gallwitz, M., Enoksson, M., Thorpe, M., Hellman, L., 2012. The extended cleavage specificity of human thrombin. PLoS One 7 (2), e31756.

Hellman, L., Thorpe, M., 2014. Granule proteases of hematopoietic cells, a family of versatile inflammatory mediators - an update on their cleavage specificity, in vivo substrates, and evolution. Biol. Chem. 395 (1), 15–49.

Hofmann, T., Schmucker, S., 2021. Characterization of chicken leukocyte subsets from lymphatic tissue by flow cytometry. Cytometry 99 (3), 289–300.

Menon, V., Ghaffari, S., 2021. Erythroid enucleation: a gateway into a "bloody" world. Exp. Hematol. 95, 13–22.

Moras, M., Lefevre, S.D., Ostuni, M.A., 2017. From erythroblasts to mature red blood cells: organelle clearance in mammals. Front. Physiol. 8, 1076.

Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., 1988. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis 9 (6), 255–262.

Onyishi, G.C., Oguine, C.C., Nwani, S.I., Aguzie, I.O., Cd, N., 2017. Haematological parameters dynamics of developing Gallus Gallus domesticus. Animal Research International 14 (2), 2769–2776.

Pejler, G., Ronnberg, E., Waern, I., Wernersson, S., 2010. Mast cell proteases: multifaceted regulators of inflammatory disease. Blood 115 (24), 4981–4990.

Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27 (2), 157–162.

Seliger, C., Schaerer, B., Kohn, M., Pendl, H., Weigend, S., Kaspers, B., et al., 2012. A rapid high-precision flow cytometry based technique for total white blood cell counting in chickens. Vet. Immunol. Immunopathol. 145 (1–2), 86–99.

Thorpe, M., Yu, J., Boinapally, V., Ahooghalandari, P., Kervinen, J., Garavilla, L.D., et al., 2012. Extended cleavage specificity of the mast cell chymase from the crab-eating macaque (Macaca fascicularis): an interesting animal model for the analysis of the function of the human mast cell chymase. Int. Immunol. 12, 771–782.

Thorpe, M., Akula, S., Hellman, L., 2016. Channel catfish granzyme-like I is a highly specific serine protease with metase activity that is expressed by fish NK-like cells. Dev. Comp. Immunol. 63, 84–95.

Vernersson, M., Ledin, A., Johansson, J., Hellman, L., 2002. Generation of therapeutic antibody responses against IgE through vaccination. Faseb. J. 16 (8), 875–877.