



UPPSALA
UNIVERSITET

Evolution and Cleavage Specificity of Hematopoietic Serine Proteases

Vamsi Boinapally

Degree project in biology, Master of science (2 years), 2012

Examensarbete i biologi 45 hp till masterexamen, 2012

Institutionen för biologisk grundutbildning och Department of Cell and Molecular Biology, Uppsala universitet

Handledare: Lars Hellman

Abbreviations

Amp	Ampicillin
Gzm	Granzyme
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LA	Luria Agar
Mc	Mast cell
MC	Macaque Chymase
Mcpt	Mast cell protease
MC_{TC}	Mast cell Tryptase, Chymase
MC_T	Mast cell Tryptase
mMCP	Mouse mast cell protease
Ni-NTA	Nick-nitriotriacetic acid
NK cell	Nature killer cell
PBS	Phosphate buffer saline
T cell	T lymphocytes

Abstract

Several of the major cell types of the immune system store large amounts of serine proteases in their cytoplasmic granules. Many different functions have been described for these hematopoietic serine proteases, like induction of apoptosis in cells infected with intracellular parasites, remodeling of extracellular matrix and triggering of inflammation by recruiting inflammatory cells. These serine proteases are encoded from four different loci, the mast cell chymase locus, the mast cell tryptase locus, the methase locus and T cell tryptase locus. Detailed analyses of the chymase locus and the T cell tryptase locus have been the focus of this thesis as well as the analysis of the extended cleavage specificity of the macaque chymase. The macaque may serve as an excellent animal model to study the *in vivo* function of various inhibitors against the human mast cell chymase. Such a model system is urgently needed to take these inhibitors into the clinic for the treatment of excessive inflammation and high blood pressure.

My results on the evolution of the chymase and granzyme A-K loci have shown that the chymase locus has expanded quite dramatically during the past 150 million years of mammalian evolution, particularly in rodents. In contrast the granzyme A-K locus has suffered relatively few changes during 450 million years of vertebrate evolution.

My analysis of the extended cleavage specificity of the macaque chymase has resulted in the identification of a new model to study *in vivo* effects of chymase inhibitors aimed for human clinical use.

The work on granzyme A and K has also passed an important first level, the production of pure and enzymatically active protease to be used in phage display analysis of their extended cleavage specificities, which in the future will serve as tools to identify novel *in vivo* targets for these important immune proteins.

1.0 Introduction:

1.1 Proteases.

Proteins are made up of a chain of amino acids. The chain of amino acids is joined together by peptide bonds forming a polypeptide. “Peptide bond formed when the carboxyl group of one amino acid condenses with the amino group of the next to eliminate water” [1]. The peptide bond can be broken (hydrolyzed) upon encounter with the protease, which contains a catalytic amino acid in their active site responsible for this hydrolysis. According to the chemical group responsible for catalysis, there are several types of proteases. One such type of proteases are the serine proteases [2].

1.1.1 Serine proteases.

Almost one third of all proteases are serine proteases. They are found in a wide variety of different species and they have got their name from the serine residue (S) that is responsible for catalytic mechanism [3]. Serine is along with histidine (H) and aspartic acid (D) residues, essential for the catalysis mechanism. Serine proteases are involved in a large number of different biological functions, such as regulators of precursor protein activation [4], processing of inactive trypsinogen to trypsin by enterokinase [5], activation of blood coagulating, regulator of zymogen precursor activation [6] and in immune responses by ingesting and killing of bacteria [7].

1.2 Granule associated serine proteases.

Several of the major cell types of our immune system store massive amounts of serine proteases in their cytoplasmic granules, i.e. mast cells, neutrophils, natural killer (NK) cells and T lymphocytes. Most of these proteases are phylogenetically closely related. For example the mast cell chymases, neutrophil cathepsin G, and the T and NK cell Granzyme B, H (B, H and N in mouse) cluster together in one locus, the chymase locus, on chromosome 14 in human and mouse [8, 9]. The genes for another set of hematopoietic serine proteases the NK and T cell expressed granzymes A and K are encoded from another locus, the Gzm A-K locus on chromosome 5 in humans and chromosome 13 in mouse. All the proteases encoded from the chymase locus have lost one of the cysteine bridges found in the majority of other serine proteases, the Cys₁₉₁-Cys₂₂₀ [9]. This thesis focuses on the evolution of the chymase and the Gzm A-K loci in different vertebrates and also on the determination of the extended cleavage specificity of one of these proteases the macaque mast cell chymase. The macaque chymase was found to be an excellent model to test protease inhibitors *in vivo* for human use.

1.2.1 Mast cell serine proteases.

Mast cells are well known for their role in allergic reactions. However, they also have a prominent role in host defense against parasite infections and in recruiting other immune cell to the area of infection by inducing local inflammation. Based on tissue location and protease content in their granules two subsets of mast cells have been described in humans. One type is primarily located in skin, in connective tissue and the submucosa of the esophagus. This subtype contains neutral proteases of both the chymase and tryptase type, hence have got the name MC_{TC}. Another type of mast cells is distributed primarily in the mucosa of the intestine and the lung and contains only tryptases and has therefore got the name MC_T [10, 11]. These proteases are involved in airway remodeling in lungs and in remodeling of the skin connective tissue during allergic reactions [12, 13].

1.2.2 Neutrophil serine proteases.

Neutrophils are known to be first line of host defense against infections. They are well known for their phagocytic activity, they engulf and kill invading microorganisms by highly reactive oxygen species and other non-oxidative strategies [14, 15]. Neutrophils store massive amounts of serine protease in their granules. Cathepsin G, N-elastase and proteinase 3 are stored in their primary granules (Azurophil granules) [14, 16]. Cathepsin G has been shown to be of importance for the resistance against *Staphylococcus aureus* and certain fungal infections [15, 17]. Cathepsin G is also involved in intracellular killing of phagocytosed bacteria in combination with other microbicidal molecules [18]. In addition, cathepsin G have the capacity to degrade various extracellular matrix components [16].

1.2.3 NK and T cell serine proteases.

Cytotoxic lymphocytes (NK cells and cytotoxic T cells) are well known for their role in inducing apoptosis in tumor or virally infected cells. Their cytoplasmic granules contain several serine protease called granzymes. Five different granzymes have been identified in human NK and cytotoxic T cells, the granzymes A, B, H, K and M. Gzm A is known to be involved in induction of cell death by a caspase-independent pathway. Gzm B and H act through a caspase-dependent mechanism. However, very little is known about Gzm K [19].

1.2.4 Nomenclature of Enzyme – substrate complex.

Schechter and Burger have defined the enzyme - substrate complex. In an enzyme the amino acids of the active site responsible for the interaction with the substrate is composed of 7 subsites. The subsites are shaped as a cleft. The subsites located on N-terminal side of the catalytic site are designated S1, S2, S3 and S4 and on the C- terminal side of the catalytic site they are designated are S1', S2' and S3'. Catalysis occur between S1 and S1' in the cleft accommodating the polypeptide substrates. The region of the substrate that fit into that cleft is designated as P1 and P1' positions of the substrate. The N-terminal side of the cleaved peptide bond of the substrate is designated as P1, P2, P3 and P4 and C- terminal side of the cleaved peptide bond as P1', P2', P3'[20].

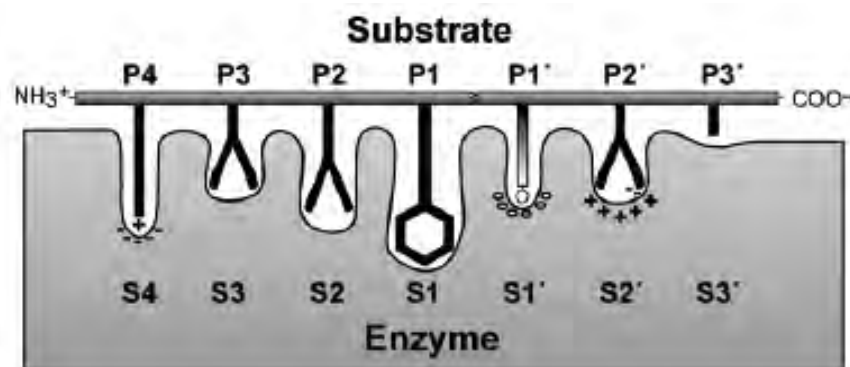


Figure 1. Schematic representation of the enzyme-substrate complex. Letters with numbers (P4 to P3') indicates amino acid residues of the substrate. Letters (S4 to S3') with cleft indicates the subsites of the enzyme. (Drawing by Mattias K. Andersson)

1.3 Cleavage specificity:

Serine proteases got their name by one of the three amino acids that form the catalytic triad, responsible for peptide bond hydrolysis, His57, Asp102 and Ser195, numbering according to chymotrypsin [21]. However, the subfamilies are determined by the amino acid side chain accommodated in S1 pocket of the enzyme (Fig. 1). Peptide bond hydrolysis occurs after the P1 amino acid of the substrate (Fig. 3). According to the cleavage specificity the serine proteases have got their name as chymotrypsin-like, trypsin-like and elastase-like. The amino acids that form the S1 pocket are 189-192, 214-216 and 224-228 [22]. Among them amino acids 189, 216 and 226 are of special importance, where amino acid 189 is located at the base of the S1 cleft, and 216 and 226 are located on the wall. Chymotrypsin-like enzymes typically hold Gly in positions 216, 226 and Ser in 189. These or other small amino acids create a preference for large hydrophobic amino acids, like aromatic side chains of Phe, Tyr and Trp of P1 substrate residue [23]. In trypsin an acidic amino acid Asp189 allow basic or positively charged amino acid side chains like Lys and Arg to enter this pocket. Elastases have Ser in position 189 and larger or often non-polar amino acids in positions 216 and 226 that create a shallow pocket to allow small aliphatic amino acids of P1 substrate residue, like Ala and Val. However, some structural elements outside of the cleavage site are also important as cleavage site determinants (S1pocket residues). These outlying interactions are important for stabilizing the substrate with the enzyme [24].

1.4 Evolution of the chymase locus and the cleavage specificity of enzymes encoded from this locus.

A relatively detailed characterization of the chymase and the tryptase locus has previously been performed. However, during the time for the analysis of the chymase locus several of the genomes included in the study were very incomplete. A major revision is therefore needed to get a more correct picture of the number of duplications involved in the expansion of this locus during mammalian evolution. The cleavage specificity for many of the proteases encoded from the chymase locus have also been analysed for their extended cleavage specificity [9, 25, 26] However, none of the previously analysed chymases have proven to be suitable as model for in vivo function of the human chymase and as model to study the effect of chymase inhibitors aimed for human use. We have therefore concentrated our effort on performing a more detailed analysis of the chymase locus and to study the extended cleavage specificity of the macaque chymase, a potential such model.

1.5 Evolution of granzyme A-K locus serine proteases and their cleavage specificity.

As mentioned above, a relatively detailed analysis have been performed of both the mast cell chymase and tryptase loci, however, no such analysis has yet been performed on the methase (granzyme M locus) and the T cell tryptase locus. This latter locus is also named the granzyme A-K locus. In order to close parts of this gap in our analysis of the evolution of the four loci encoding hematopoietic serine proteases I have performed a detailed analysis of the granzyme A-K locus. Although no detailed evolutionary analysis of this locus has been performed quite a lot of information on these enzymes have accumulated during the past 20 years. For example, granzyme A and K are situated on human chromosome 5 [27, 28] and chromosome 13 in mouse [29, 30]. The chromosomal organization of the locus and initial phylogenetic studies has suggested that Gzm K evolved from early gene duplication of Gzm A [31]. Humans and mouse have a similar orientation and tryptase activity of Gzm A and K, which means that they prefer to cleave after Arg and Lys. This former finding indicates that a gene duplication has occurred before the separation of primates and rodents [30, 32]. In birds like zebra finch and chicken only a Gzm A gene has been observed. No Gzm K related gene

has been identified [33]. In fishes like catfish and Atlantic cod Gzm like sequences have been identified, which share 50% of the similarity with each Gzm A and K [34, 35]. This similarity indicates that Gzm A and K genes had evolved before the separation of tetrapods from bony fish, which did occur approximately 420 million years ago [35], a conclusion that we now question. No detailed analysis of the cleavage specificity of granzymes A or K has been performed. We here present the first steps in such an analysis.

2. 0 Materials and Methods

2.1 Mapping of the chymase locus and the granzyme A-K locus in vertebrates.

Genes for the chymase locus and the granzyme A/K were analyzed for a broad range of vertebrates. The genes were obtained from <http://www.ncbi.nlm.nih.gov/> and <http://www.ensembl.org/index.html> by using tblastn with known human genes as bait. The genes were aligned with Clustal X (ver.2)[36], followed by phylogenetic analysis with 1000 independent boot straps with the Neighbour-Joining method. The bootstrap tree was generated with NJ plot [37].

2.2 Determination of extended cleavage specificity of the macaque chymase by the method of substrate phage display.

2.2.1 Source of Macaque Chymase.

In my project I have worked on the macaque mast cell chymase. The active protease was provided by a research group at Johnson & Johnson in the USA. This protease was produced in baculovirus-infected cells (High Five™) (Invitrogen, Carlsbad, CA). The accession number for the macaque chymase is P56435 (macaque). This was obtained from SwissProt/TrEMBL.

2.2.2 Determination of cleavage specificity by substrate phage display method.

A phage library with approximately fifty million different phage variants was used to determine the cleavage specificity. In this library, the T7 phage capsid protein 10 was modified to generate nine random amino acid peptide ends followed by a six histidine residue tag that made it possible to attach the phages to a matrix (Ni IMAC-Sepharose). The sequence of the random peptide region was PGG (X₉)HHHHHH, where (X₉) is the random amino acids sequence, followed by six histidine residues as tag at the C terminal end. This His₆ tag has high affinity towards Ni-NTA beads, which makes it possible to immobilize the phages to the Ni-NTA beads. His₆-free and low affinity phages were eliminated by several washing steps. Immobilized phages with cleavage specific peptides were released into supernatant upon treating with protease. These specific phages were amplified by infecting a bacterial culture for the next selection round. By repeating this selection round five to six times only phages that carry a sequence highly susceptible to cleavage by the protease will remain in the sublibrary. Sequencing of the random region of individual phages will then make it possible to determine the cleavage specificity of the protease.

2.2.3 The phage display method - experimental procedure.

In the first selection round 300 µl (about 10⁹ plaque forming units) of the phage library was added to 100 µl of Ni-NTA beads along with 100 µl 5 M NaCl and 600 µl PBS in a 1.5 ml eppendorf tube. The mixture of phages and Ni-NTA beads was then incubated at 4°C by rotating for 1 hour. During this time phages with His-tag were allowed to bind to the beads.

Unbound or low affinity phages were removed by ten washing steps using 1ml of washing buffer per cycle (1 M NaCl, 0.1% Tween-20 in PBS, pH 7.2) followed by washing twice with 1 ml PBS. The beads were then resuspended in 500 μ l PBS. Seven μ l of protease was added and the sample and the mixture was incubated at 37°C for 2 hours by gently rotating. Protease free PBS was used as control. During incubation phages with nonameric peptides, specific to protease, were cleaved and thereby released into the supernatant. After 2 hrs of incubation the beads were pelleted by centrifugation. Thirty μ l of the supernatant was taken to perform a limiting dilution to determine the number of detached phages in each selection round. To ensure that phages with His-tag still remaining in the supernatant were removed by 15 μ l fresh Ni-NTA beads and 100 μ l PBS was added to the supernatant. To calculate the total number of immobilized phages on Ni-NTA beads a control elution was carried out by adding 100 μ l mM imidazole to the beads. Diluted phages were plated on LA amp plates by mixing 100 μ l phage dilution with 100 μ l 100mM isopropyl β -D-1-thiogalactopyranoside (IPTG), 300 μ l of early cultured *Escherichia coli* BLT5615 (concentration up to 0.5 at optical density OD₆₀₀). Different plates were used for different dilutions of each protease treated and PBS control. Two other plates were used for each reaction for the eluted beads. Ten ml of *E. coli* BLT5615 was cultured at 37°C for 30 min followed by the addition of 100 μ l 100 mM IPTG. IPTG induce the production of T7 Phage capsid protein. Followed by adding the left over supernatant to the culture and subsequent incubation at 37°C for 75 min result in bacterial lysis by phage amplification. One and a half ml of the lysate was centrifuged at 10,000-x g for 3 min to remove debris. Eight hundred μ l of this supernatant was taken to a separate tube and 100 μ l of 5 M NaCl and 100 μ l of PBS was added. This solution was then used for the next round of selection. In the next rounds, selection was carried out by adding 100 μ l fresh Ni-NTA beads to the phage sub-library followed by 15 washing steps.

2.2.4 Picking of individual phage plaques for PCR amplification and sequence analysis

After five rounds of selection the released phages were plated onto LA AMP plates and 100 individual plaques were picked with Pasteur pipette and transferred individually into 100 μ l phage extraction buffer (100 mM NaCl and 6 mM MgSO₄ in 20 mM Tris-HCl) in eppendorf tubes. Phage extraction was carried out at 65°C for 30 min by vigorous shaking. After extraction the phage lysates were centrifuged at 14000 rpm in an eppendorf and the supernatant was then kept at 4°C until further analysis.

2.2.5 PCR amplification of the region of the phage genome that encodes the nine amino acid random peptides

Using the extracted phage DNA as template PCR reaction was performed to amplify the region of the phage that encodes the random peptide region. Master mixture prepared with following components to perform the PCR reaction. Two μ l phage lysate, 5 μ l 10x *Taq* buffer E with MgCl₂, 1 μ l T7 Up primer (5'- GTTAAGCTGCGTGA CTTGGCT-3'), 1 μ l T7 Down primer (5'-TTGATACCGGAGGTTACCGA-3'), 1 μ l 10mM deoxyribonucleotide triphosphate (dNTP) mix, 0.5 μ l *Taq* DNA polymerase and 39.5 μ l deionized water.

The PCR reaction consisted of 3 steps. Step 1 was at 94° for 5 minutes. The second step was repeated 35 times. This step consisted of denaturation at 94° for 50 seconds, annealing at 50°C for 1 minute and elongation at 72°C for 1 minute. Following the 35 cycles of amplification a third step was included by incubating the sample at 72°C for 6 minutes to complete the extension. After amplification the PCR samples were loaded onto 2% agarose gel with 0.1 μ l/ml ethidium bromide. Samples with nicely visible PCR fragments were sent for sequencing at GATC Biotech in Germany in 96 well plates.

2.2.6 Alignment of the sequences of the random nonamers to identify a consensus cleavage site

The nucleotide sequences obtained from GATC Biotech were translated into amino acid sequences. The sequences were then aligned starting with sequences that had only one preferred P1 amino acid. Sequences with multiple cleavable sites were then aligned using the parameters obtained from sequences with a single potential cleavage site. Background phages were excluded from the alignment. According to the Schechter and Berger cleavage happens between the P1-P1' sites of the substrate [20], and the entire region surrounding the cleavage site is presented as follows, P5 P4 P3 P2 P1 P1' P2' P3' P4'. In the alignment the amino acids are color coded according to their properties.

2.3 Expression of granzyme A, K in a mammalian expression system

2.3.1 Cloning:

Human granzyme A, K and Codfish A/K coding regions were synthesized by GenScript Company and sent to us as fragments inserted into the pUC57 vector. These protease-encoding genes were transferred into the bacterial expression vector pET21a-d (+)r (Novagen). Following initial test of production in a bacterial host (*E. coli* Rosetta gami) we found that the bacteria produced high amounts of the proteases. However, after purification from inclusion bodies and refolding the enzymes were not enzymatically active and could therefore not be used for further experiments. We therefore decided to reclone the inserts into a mammalian expression vector pCEP-Pu2. The granzyme A, K, and A/K inserts were ligated into pCEP-Pu2 and transformed into *E. coli* competent cells. Ligation reaction was carried out for the three inserts separately by mixing 3 µl extracted pCEP-Pu2 vector with 10 µl of each extracted insert, 2 µl 10x ligase buffer, 1 µl T4 DNA ligase, 4 µl distilled water (dH₂O) followed by incubation at room temperature for 2 hrs. These ligated samples were then transformed into DK1 competent cells and plated on three different LA+Amp plates. Four colonies were cultured from each plate and plasmid DNA was extracted, according to the protocol of (E.Z.N.A plasmid miniprep kit (provided by Omega Bio-Tek). The plasmid samples were cleaved using 5 µl of vector DNA, 2 µl 10 x enzyme buffer, 1 µl EcoRI (Conc.10U/ µl), 1 µl XhoI, 11 µl dH₂O and incubated at 37°C for 1 to 2hrs. After incubation 5 µl of sample buffer (5x Ficol buffer) was mixed with digested samples and run on a 1.2% agarose (SeaKem® GTG® Agarose) gel electrophoresis (0.5X TBE-Tris, Boric acid, EDTA) at 100V.

Before transfection the recombinant plasmid miniprep were sterilized by ethanol precipitation of the plasmid DNA by addition of one tenth of the volume of 3M Sodium acetate, followed by addition of 2 volumes of 99% Ethanol. The mixture was kept at -20°C temperature for 30 min. Then the mixtures were centrifuged at maximum speed and the supernatant was discarded. The samples were then incubated in a Speedvac to completely remove remaining traces of ethanol. The plasmid DNA was then resuspended in one third the original volume of TE (10 mM Tris, pH 7.5, 1 mM EDTA in dH₂O).

2.3.2 Transfection to Human embryonic kidney cell line 293 (HEK293)

Before transfection into the HEK293 EBNA cell line, the cells were grown in 25 cm² cell culture flask by using Dulbecco's Modified Eagles Medium (DMEM). Once the cell confluency reached 70%, the cell media was removed and the cells were washed with neutral PBS, which was followed by addition of fresh DMEM media. Then the cell line was transfected with a solution contain 25 µl sterile plasmid DNA, 25 µl sterile TE, 40 µl

lipofectamine, 710 μ l serum free DMEM mixed with 50 μ g/ml gentamicin. After 24 hrs the cells were expanded to 75 cm² culture flasks in DMEM containing 5% FBS, and 50 μ g/ml gentamicin. Two days later selection media was added to the cells, DMEM with 5% FBS, 50 μ g/ml gentamicin, 0.5 μ g/ml puromycin and 5 μ g/ml heparin. After 1-2 weeks the cells reached confluency in the small flask used for transfection and the cells could be expanded for larger scale production. Once the colour of the media changed from red to light yellow the media was collected and stored in 4^oC until purification.

2.3.3 Purification of recombinant proteases.

The stored medium was filtered through glass filter into 1 liter bottles. After filtration the media was transferred into sterile 50 ml falcon tubes. Forty μ l of NI-NTA agarose beads were added to 100 ml of filtered media followed by incubation at 4^oC for 45 min under rotation. After incubation, tubes were centrifuged (3,000 rpm for 3min at 4^oC) to pellet the beads in the cone shaped bottom. Beads were collected by using Pasteur pipettes and transferred into 1.5 ml eppendorf tubes. The beads were pelleted by centrifugation and the supernatant was discarded. Beads were then transferred into 2 ml columns (a syringe, Terumo Europe N.V., 3001 Leuven, Belgium) with glass filters in the bottom (Whatman GFC, glass filter, England) by resuspending the pellet in protein wash buffer (1 M NaCl, 10 mM imidazole in PBS Tween 0.05%). The beads were then washed with protein wash buffer 4 times, first with 1ml, then with 2 ml washes. After washing the protein was eluted in 5 fractions starting with 125 μ l, followed by 150 μ l by adding protein elution buffer (100 mM imidazole in PBS Tween 0.05%). All the fractions were then analysed on SDS-PAGE gels (NuPAGE® Novex 4-12% Bis-Tris Gel 1.0 mm, 12 well). Five μ l of the fractions were mixing with 2 μ l LDS- sample buffer and 0.5 μ l β -mercaptoethanol and then heated for 4min at 80^oC. The gels were stained with colloidal Coomassie brilliant blue (1.7 ml of 85% Phosphoric acid, 100 g ammonium sulfate, 29 ml of 5% Serva blue G in distilled water).

2.3.4 Activation of proteases

Purified proteases were activated by adding 1 μ l enterokinase (Roche) to 20 μ l of purified protein fractions and incubated at 37^oC for 5 hrs. To confirm that the protein had been cleaved a samples of activated and non-activated samples were run on SDS-PAGE gels, according to the procedure described above (2.3.3).

3. 0 Results

3.1 Analysis of the chymase locus genes

A detailed analysis of the chromosomal arrangement of chymase locus genes in a number of different mammalian species was performed (Fig. 2). The human locus is one of the least complex of all the ones studied and contains only four active genes, the α chymase (*cmal*), cathepsin G (*ctsg*) granzyme H (*GzmH*) and granzyme B (*GzmB*). These four genes are located together in the chymase locus on chromosome 14. A very similar gene arrangement can be observed in several primates, including chimpanzee, gorilla and macaque. Phylogenetic analysis showed that all α -*cmal* genes cluster together in a *cmal* group. Granzyme B and H also cluster in one branch in the phylogenetic tree, suggesting that these granzymes have evolved by duplication from a close common ancestor. In contrast to primates dog has two chymase genes *cmal* and *cmal2*. However, a frameshift mutation in the coding region of *cmal2* has transformed this gene into a pseudogene ([25]). Cattle have also two chymase genes (*cmal1a*, *cmal1b*) and two cathepsin G (*ctsg1*, *ctsg2*) genes where the two

cathepsin G genes are located in between granzyme B and H. In addition cattle has four members of a new sub class of genes within this locus that has changed tissue specificity, the duodenases (*BDMD1*, *BDMD2*, *BDMD3*, *BDMD4*). The organization of the cattle chymase locus indicates that a duplication has occurred, involving the chymase and cathepsin G, and that several duplications have generated the four duodenases. The organization of the cattle chymase locus also indicates that inversions have been involved in shaping this locus in cattle. The two *cmal* genes are very similar in sequence indicate that they have appeared by a relatively recent duplication. Interestingly the duodenases cluster together with cathepsin G. This result indicates that the *BDMD* genes originated from *ctsg* and that relatively recent duplication has lead to the four duodenases in cow. Mouse has a total of 15 functional genes within the chymase locus. It has six chymase genes, *cmal*, *cma2*, *Mcpt1*, *Mcpt9*, *Mcpt2*, and *Mcpt4* [38], one gene for a new basophile specific serine protease *Mcpt8* [39], one gene for neutrophil cathepsin G and seven granzyme genes (*e*, *d*, *g*, *n*, *f*, *c* and *b*)[31]. One of the mouse chymase genes *cmal* (*Mcpt5*) clusters with the α -chymases and the five other chymases *cma 2* (*Mcpt10*), *Mcpt1*, 2, 4 and 9 form a separate chymase group that only seems to exist in rodents. This subfamily of chymases has been named β -chymases. The α - and β -chymases cluster together in a phylogenetic tree which indicates that α -*cmal* and β -*cma* genes are a result of several gene duplications, and that several successive gene duplications has resulted in the different β -chymases *Mcpt1*, 2, 4, and 9. *Mcpt-8* forms a branch together with *ctsg*, *BDMD*, and *GzmB* and *H* [40]. In the rat a total of 30 serine protease genes are found within the chymase locus. Analysis of an amino acid alignment, numbered after chymotrypsin [21] shows that 28 of the 30 genes in rat have a functional catalytic triad. Two of them, *mcpt8 like3* (XP_573781.1) and *Mcpt1 like* (XP_001079049.1) show mutation in one of the three residues of the catalytic triad indicating that they have lost their proteolytic activity and possibly also become pseudogenes.

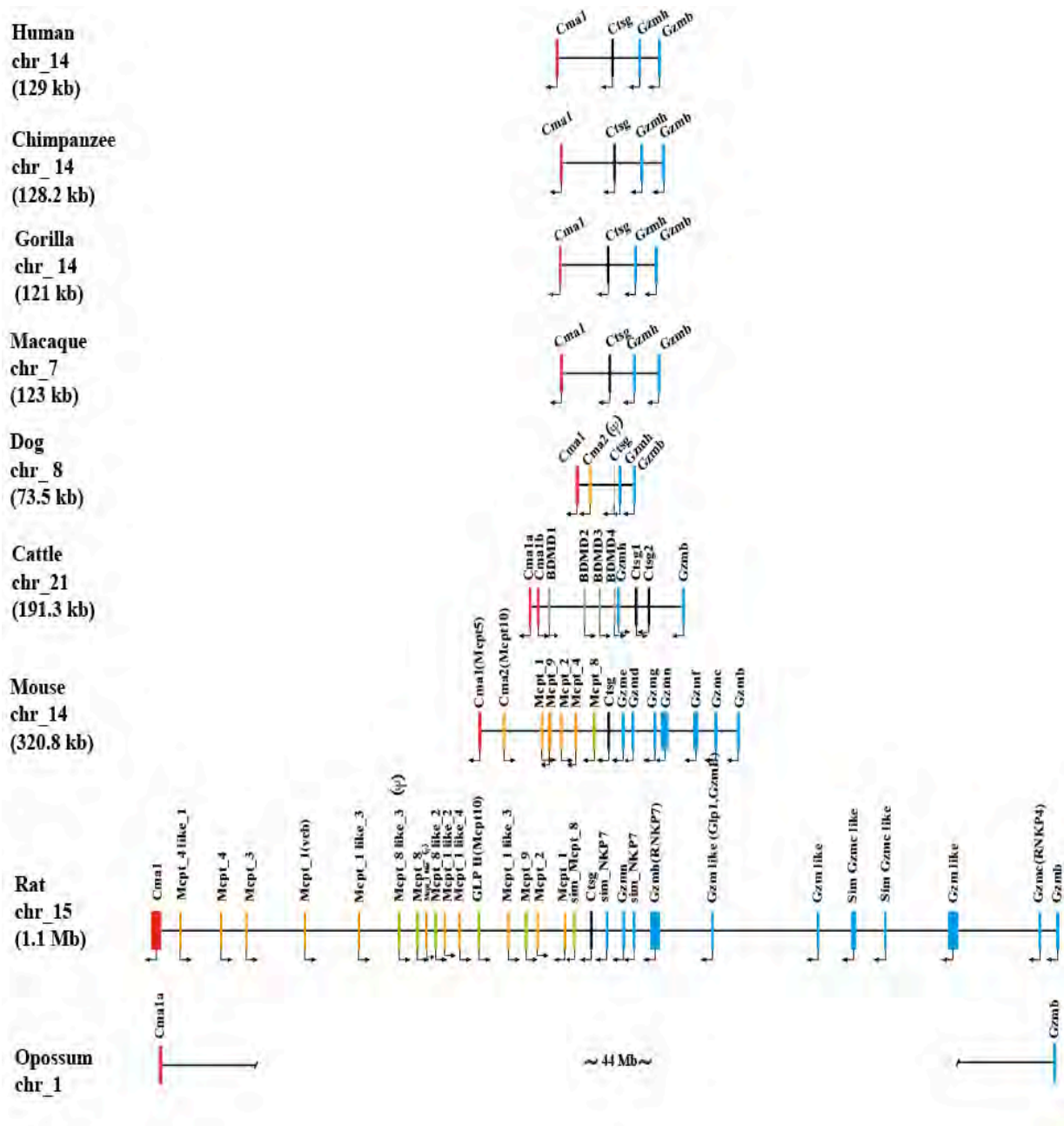


Figure 2. Chromosomal organization of the chymase locus genes in a number of different mammals.

Species name, to the right, gene position in scale and their names in the center of the figure and the chromosome number to the left side are shown. Different colours of the gene boxes indicate the gene family they belong to. The location of the genes on chromosome is indicated by a horizontal line and the gap between opossum chromosome indicate the distance between the genes on the same chromosome. Pseudogenes are indicated by a ψ mark.

The analysis of a panel of mammals concerning the structure of their chymase loci has shown that a remarkable expansion of this locus has occurred in rodents, from 4 genes in humans to 15 and 28 functional genes in mouse and rats, respectively. Interestingly, a marked expansion has also occurred in cattle with duplicated cathepsin G and α -chymases and the appearance of a completely new subfamily, the duodenases with a new tissue specificity and function. From being immune proteases they are now expressed in the digestive system with food digestive functions.

Interestingly three fish proteases included in the phylogenetic tree in Figure 6 also cluster with the chymase locus genes of mammals indicating that chymase locus related genes also are present in fishes. However, these proteases do not group with any of the various subfamilies within the chymase locus of mammals like chymases, cathepsin G or granzymes but are instead distantly related to all of them.

3.1.1 Analysis of the granzymeA-K locus

A study by a former project student in the lab (Yue Li) has shown that the GzmA/K locus has been relatively well conserved from lower vertebrates to placental mammals. However, due to incomplete genome sequences, chromosomal localization and the structure of the locus has been very fragmentary in many species. I have therefore performed a more complete analysis of this locus (Fig. 3). In humans the genes for gzm A and K are located together on chromosome 5 along with other protein coding genes like ESM1, CDC20B, CCNO, GPX8, DHX29, SKIV2L2 and PPAP2A (Fig. 3). A very similar chromosomal arrangement is observed in chimpanzee, macaque, mouse and rat (Fig. 3). In a phylogenetic analysis gzm A and gzm K genes of these species were clearly grouped into separate gzm A and gzm K branches (Fig. 4). In the dog gzm K is located on chromosome 4 and gzm A on chromosome 2, which clearly indicated a chromosomal rearrangement, possibly by a crossing over event that has occurred during canine evolution (Fig. 3). Two gzm A genes have been identified in cow and elephant, which probably are the result of a recent gene duplication (Fig. 3). In opossum only gzm A was identified, which indicates that gzm K may have been lost in the marsupial lineage. In a monotreme, the platypus, neither gzm A nor K genes have been identified. However, several of the border genes have been detected. Interestingly, the border genes DHX29, SKIV2L, PPAP2A are here present with another gene HCN1. In placental mammals HCN1 is located outside these border genes. It is most likely that gzm A/K genes have not yet been annotated in this species. In birds like chicken, turkey and zebra finch only gzm A has been identified. Gzm K could possibly have been lost in the reptile lineage, or alternatively the gene duplication giving rise to granzyme K may have occurred during early mammalian evolution. In a phylogenetic analysis of these species gzm A genes clearly cluster with gzm A genes of mammals, which may indicate that gzm K has been lost. In amphibians like the frogs several gzm A genes are observed but no gzm K genes. This finding favors the second alternative, that the appearance of gzm K took place during early mammalian evolution. So the question of the appearance of gzm K is still open. The presence of three relatively distantly related gzm A genes also indicates that two gene duplications have occurred very early in amphibian evolution possibly at the time or just after the separation from early reptiles and mammals. In various fish species like fugu, medaka, stickleback, catfish, salmon and zebrafish we find homologues to either gzm A or K (Fig. 4). However, although all the fish gzm A-K genes appear in the granzyme K branch, these proteases are located far out on this branch in the phylogenetic tree, which make it difficult to assign them as either A or K. They are most likely almost equally related to A or K, which indicates that the separation into A and K had not occurred at the time that fishes separated from tetrapods. The timing of that separation is now assigned to a period around 390-420 million years ago [35]. In fugu three gzm A-K genes have been identified, medaka has one gene, stickleback one, catfish one, salmon one and zebrafish has two (Figs 3 and 4). This analysis show that all species investigated, excluding platypus where the genome still is very incomplete, have one or several gzm A-K genes which clearly indicate that these genes are very important for survival. They have appeared more than 450 million years ago and they have been maintained in all vertebrates that have been studied in detail.

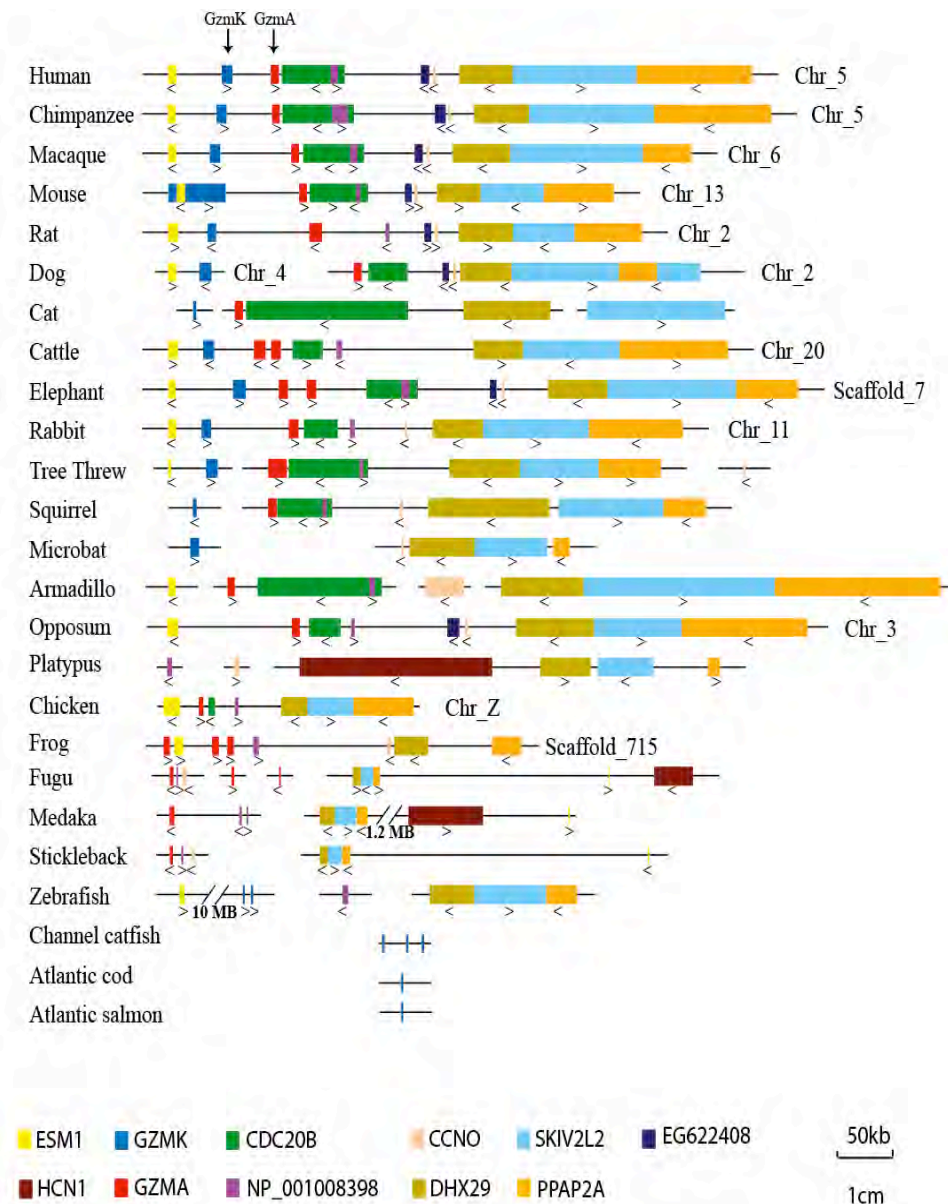


Figure 3. Chromosomal organization of the gzm A-K locus in vertebrates. The chromosome number where the locus is found is indicated at the right side of the figure, when the number is known. Boxes and lines with different color indicate the gene family they belong to. The location of the genes on chromosome or scaffold indicates by one single horizontal line and multiple horizontal lines indicate different contigs or not fully sequenced regions. Dashed line mark large gaps in the figure (a large sequenced region that is too large to be shown in the figure) with the size of the gap indicated by numbers below.

3.2 Phage display analysis of the extended cleavage specificity of the macaque mast cell chymase.

The macaque chymase is an interesting enzyme from an industrial point of view, to get a model enzyme and an organism to study the effect of chymase inhibitors for human use. In order to evaluate the macaque as such a model we have studied the extended cleavage specificity of this enzyme and compared it with the corresponding human and dog enzymes. The three enzymes were obtained from Johnson and Johnson in the USA in an active and pure form. They had been produced in a baculovirus system in insect cells and were after activation and purification tested for proteolytic activity against synthetic substrates. The active enzymes were then used to study their extended specificity by phage display. A library with approximately 50 million different phages were used and each phage had a nine amino acid long random peptide inserted in one of the capsid proteins followed by a six histidine long affinity tag. In first selection round using this library and active macaque chymase similar numbers of released phages were observed for both PBS control and enzyme treated phages. Upon continuing the selection the number of phages released by the enzyme increased and after five rounds the difference in titers between PBS control and macaque chymase was 174 times. Hundred individual plaques were then selected for PCR amplification of the region containing the nine amino acid long random region. The amplified DNA samples were analyzed on 1.2% agarose gels (data not shown). Clones that gave nice PCR bands were selected for sequencing.

Of the 100 sequences 70 were of good quality and represented sequences cleaved by the macaque chymase. Thirty of the sequences contained either stop codons in the nonamer region or were background phages.

3.2.1 Alignment of nonapeptide sequences.

The nonamer regions of the phage DNA sequences were translated into amino acids sequences. Out of 100 sequences 78 sequences were aligned manually. In rest some of them lost his tag and some are repeated sequences. These sequences were aligned manually in P4-P3' direction in order to obtain a consensus sequence, by considering the aromatic amino acids in P1 position. Where MC preferred to hydrolyze peptide bond after P1 amino acid. Sequences with one aromatic amino acid in P1 position aligned first and this pattern was used to align the sequences with multiple aromatic amino acids. When comparing with results from previous studies of the human chymase (HC) and dog chymase (DC), macaque chymase (MC) seemed to have very similar or identical amino acid preference in P1 position. MC preferred aromatic amino acids (tyrosine, phenylalanine and tryptophan) in the P1 position. In the P2' position MC, as HC prefer negatively charged amino acids (glutamic acid and aspartic acid) (Fig. 5). Dog chymase also prefer aromatic amino acids in P1 position. However, DC differ from the two other enzymes in that it prefer positively charged amino acids (arginine) in the P2 position. In the P2' position the DC also showed a less pronounced preference for negatively charged amino acids compared to MC and HC.

Human chymase Macaque chymase Dog chymase

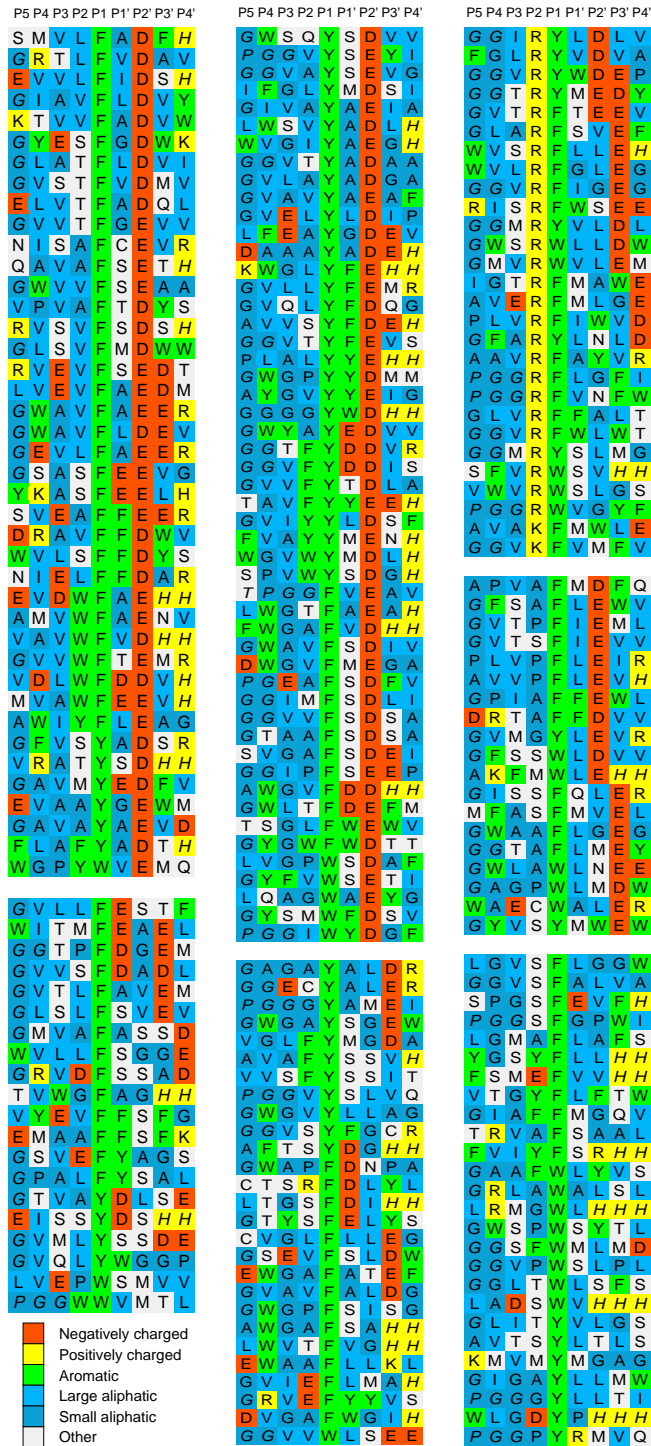


Figure 5. Alignment of nonamer sequences from the phages obtained after phage display with HC, MC and DC. Amino acids were arranged in P5-P4' direction, where the cleavage site is located between the P1 and P1' position. Amino acids were colored according to their characteristics, like positively charged, negatively charged, aromatic, aliphatic etc. Sequences were aligned by considering the cleavage site in P1 position.

3.3 Cloning of gzm A, K into a mammalian expression vector:

Production of human granzymes A and K and catfish A/K was first tested in bacteria. However, the protein was not active why the coding sequences of the three enzymes had to be recloned into a mammalian expression vector. The inserts from the bacterial expression vector pET21a-d (+) vector was excised by restriction digestion and the mammalian expression vector pCEP-Pu2 was cleaved with the same enzymes and the cleavage products were separated on agarose gels. Vector and fragments were extracted from the gel (Fig. 6). The granzyme A, K and A/K fragments were ligated into the pCEP-Pu2 vector, for mammalian expression.

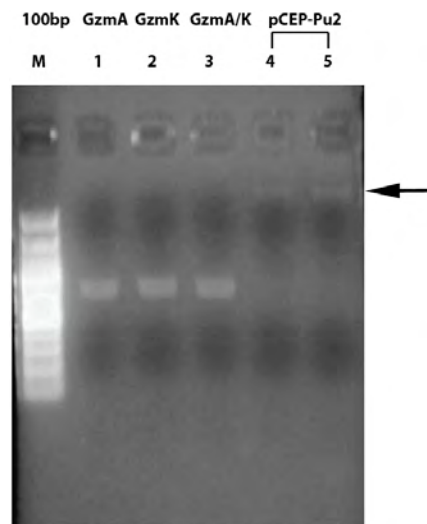


Figure 6. An aliquot of the extracted granzyme fragments and the vector was separated on a 1.2% agarose gel. As seen from figure bands in lane 1, 2 and 3 shows purified gzm A, K, and A/K inserts. Arrow indicates the vector bands in lanes 4 and 5, the pCEP-Pu2 vector.

3.3.1. Restriction enzyme analysis to confirm correct inserts in clones of granzyme A, K, A/K in the pCEP-Pu2 vector

The DNA fragments encoding the granzymes were successfully ligated into the pCEP-Pu2 vector. To confirm the correct inserts in the vector, different restriction sites were selected based on nucleotide sequence of the three different granzymes. For human gzm A Pst1 (CTGCAG) gave a characteristic banding pattern and for gzm K EcoRV (GATATC) was chosen. Nco1 (CCATGG) was selected for catfish gzm A/K. Xho1 (CTCGAG) site, which is found in all clones was chosen to perform the double digest. Based on the restriction analysis we can conclude that the right sequences had been inserted in the various clones (Fig. 7).

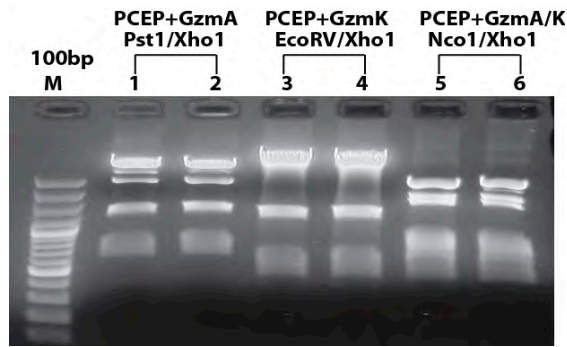


Figure 7. Restriction analysis for confirmation of positive clones. Separation on a 1.2% agarose gel was performed on the double digested samples. Two different samples from each clone were analyzed. In lane 1 and 2 the pCEP vector with gzm A had been digested with PstI/XhoI, in lane 3 and 4 the pCEP vector with gzm K inserts digested with EcoRV/XhoI. In lanes 5 and 6 pCEP with gzm A/K were digested with NcoI/XhoI. Result from the double digest confirms correct inserts.

3.3.2 Purification of recombinant proteins:

Recombinant gzm A and K proteins were successfully purified from conditioned media of the transfected HEK293 EBNA cells by Ni-NTA affinity chromatography. The bound proteins were eluted with 100 mM imidazole. Relatively good yields and high purity was observed for human granzymes A and K. However for the fish enzyme, catfish gzm A/K we obtained very low amounts of protein (Fig. 8)

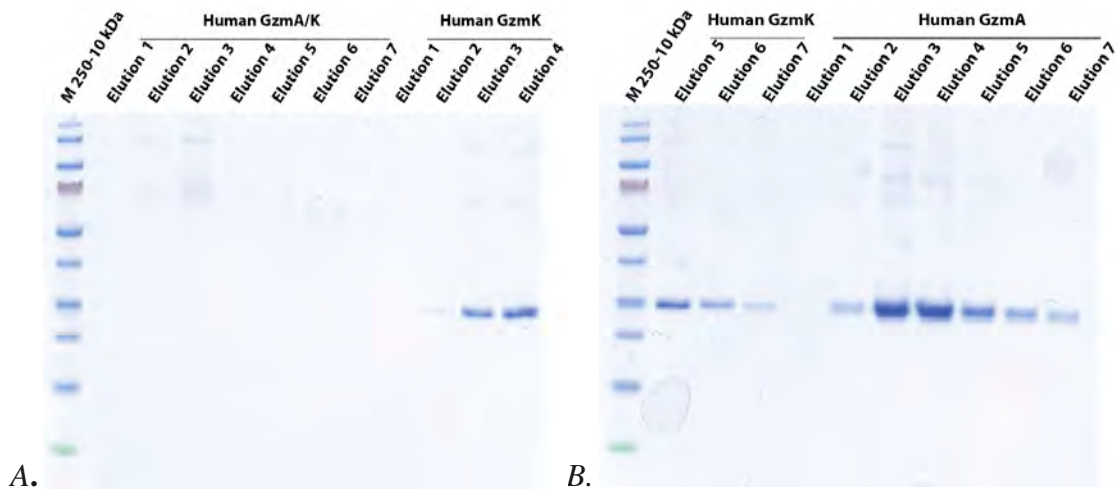


Figure 8. SDS-PAGE analysis of purified recombinant proteins. Catfish gzm A/K samples are seen in lanes 1-7 of gel (A). No detectable bands were seen in the correct position of the gel. Fractions from the gzm K purification are shown in lanes 3 and 4 of gel (A), and lanes 5-7 in gel (B). Fractions 3, 4, 5 and 6 show good levels of protein with high purity. Fractions from the purification of gzm A is seen on gel B, where all fractions except the first contain the recombinant enzyme and the enzyme has a relatively high purity.

3.3.3 Activation of human Granzyme A, K and catfish Granzyme A/K

Eluted protein from the purifications of human granzyme A and K as well as fractions from the catfish A/K purification were treated with enterokinase to make the protease active by removing N-terminal signal peptide and the enterokinase site (Fig. 9). These activated proteases will be used to study the extended cleavage specificity by substrate phage display

method. The enterokinase cleavage removes an approximately 1.5 kDa fragment from the N-terminal of the protein.

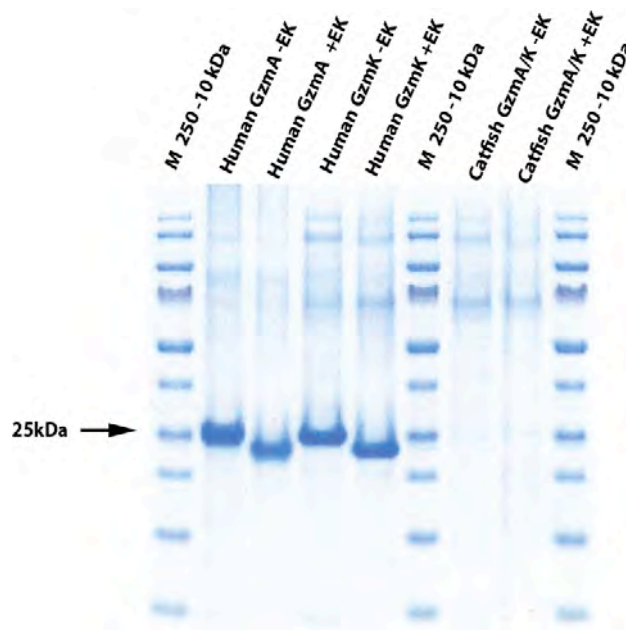


Figure 9. SDS-PAGE analysis of enterokinase cleaved granzymes A, K, A/K

SDS-PAGE (4-12% gradient) gel analysis was performed to confirm the activation of human granzymes A, K and catfish granzyme A/K by enterokinase. Bands at 25 kDa show the proteases without enterokinase activation. After activation the size of the bands are reduced by approximately 1.5 kDa. The lanes of catfish granzyme A/K shows no bands in the region where we expect to see the protease, which indicates the catfish granzyme A/K protease was not produced in the mammalian cells or that it is being degraded before leaving the cells.

4. 0 Discussion

Mast cells play an important role during the induction phase of an allergic inflammation. However, they do not only have negative properties, they are also involved in the protection of the host against various pathogens [13]. The chymase is a serine protease stored in mast cells granules. This chymotryptic enzyme has been implicated in a number of quite diverse biological functions like remodeling of tissues, promotion of inflammation and in angiotensin conversion [41]. The studies presented in this communication has focused on the functional diversification during vertebrate evolution of the mast cell chymase and other related hematopoietic serine proteases and also tried to identify a novel model organism to perform preclinical studies in a relevant animal model to assess the potency of chymase inhibitors to be used as new anti-inflammatory drugs in humans. This latter project has been focused on determining the extended cleavage specificity of the macaque counterpart of the human chymase by phage display analysis.

Mouse and rat are often used in preclinical studies. However, the human and mouse chymase locus show very significant differences. The human chymase locus have 4 genes (Fig. 2), the mast cell α -chymase, neutrophil cathepsin G and T cell granzymes B and H [9], whereas the corresponding loci in mouse and rat have expanded quite dramatically and have 15 and 28 protein coding genes, respectively [42]. The human chymase has chymotrypsin like activity, which means that it prefer aromatic amino acids like phenylalanine and tyrosine in the P1 position [43]. According to the chymotrypsin numbering [21] the human chymase hold 189 Ser, 216 Gly and 226 Ala in its specificity determining positions [9, 23]. A phylogenetic analysis shows that mouse and rat mMCP5 and rMCP5 genes cluster with α -chymases (Fig.

4). However mMCP5 and rMCP5 preferred aliphatic amino acids, like alanine and valine, in the P1 position and thereby have elastase like activity and not chymotrypsin like activity as the primate α -chymases [44, 45]. Mouse has also 5 β - chymase like genes and rat has 11. However, the mouse and rat β - chymases mMCP4 and rMCP1 prefer aromatic amino acid phenylalanine in P1 position like the human chymase [43]. In the P2' position mMCP4 and the human chymase also both prefer negatively charged amino acids [43]. However, rMCP1 prefer serine in this position [43, 46]. In addition and in contrast to humans rat and mouse also have multiple Mcpt8 and Gzm genes. In the human chymase locus no member of the Mcpt8 family have been detected. In contrast to humans mouse and rat hold 7 and 11 gzm genes, respectively. Humans have only two, gzm B and H. For the above listed reasons rodents are not suitable models to examine effects of inhibitors against the human chymase. Dog is also not a good model as the dog chymase has a slightly different extended specificity than the human chymase.

In contrast to rodents macaque has only one chymase gene and this protein is highly homologous, to the human chymase, showing 98% identity. For this reasons I chose the macaque chymase to perform substrate phage display method to characterize its specificity. The macaque chymase was found to prefer aromatic (Phe, Tyr and Try) in the P1 position and acidic amino acids (Glu and Asp) in the P2 position as the human chymase (Fig. 4). This result suggests that macaque is the suitable animal model to study the inflammatory response during allergic reactions and to examine protease inhibitors in vivo, aimed for human use. The great similarity between the chymase locus in these two species is also a major advantage.

Cytotoxic T cell and nature killer cells are immune cells involved in the defense against viral infected cells and tumors [47, 48]. The cytotoxic T cells carry large numbers of granules that contain serine proteases, so called granzymes, and a pore-forming protein named perforin. The cytotoxic T cells are involved in elimination of cells infected with virus or tumors by forming pores with perforin to allow calcium and the different granzymes to enter the cell. Recent studies suggests that granzymes induce the cell death, apoptosis, by several different mechanisms [19]. However, the enzymatic function of granzymes A and K have not been studied in detail. To identify natural substrates for these two granzymes I have performed an evolutionary study of these granzymes, from bony fishes to placental mammals, followed by expression of these two proteases in a mammalian expression system to study their extended cleavage specificity.

The granzyme A, K locus have been relatively well conserved during vertebrate evolution. During phylogenetic analyses granzyme A and K genes also clearly cluster in separate groups. The fish homologues seemed to be distantly related to both granzyme A and K. However, slightly more related to gzm K (Fig 4). Three other catfish gzm-like proteins were also found to be distantly related with mast cell chymase locus genes. The extended cleavage specificity and the in vivo targets for granzyme A and K are not yet known and we do not know the specificity of any of the fish homologues to the chymase locus genes described above. This was one of the aims of this project.

In my work I have selected the human granzyme A and K and catfish granzyme A/K. I first expressed the human gzm A and K, and catfish gzm A/K proteins in a bacterial expression system the *E. coli* strain Rosetta-gami. The proteins were produced in high yields. However, we were not able to get them in an enzymatically active form after cleavage with enterokinase. The reason could be that the high production at 37°C resulted in the formation of inclusion bodies. The aggregation of the proteins in the inclusion bodies makes it difficult to recover them in a native conformation. According to previously published studies on gzm

K this protease can be produced in *E. coli* and the protease could in their hands be obtained in an active form by refolding techniques starting from inclusion bodies [47]. However I tried to refold the protein by standard refolding protocol but did not succeed. I therefore decided to test the eukaryotic expression system HEK293 EBNA (Human embryonic kidney cells) to produce human gzm A and K and catfish gzm A/K. Of these three proteases, I have so far succeeded to produce human gzm A and K in an active form at high purity and at a good yield, to analyze primary and extended specificity. However, codfish granzyme A/K was much more difficult to produce and we could not detect any protein from almost a liter of conditioned media. The reason why we get so low yields of the catfish granzyme is not known but could be due to its high positive charge. The same phenomenon has previously been observed with a few other hematopoietic proteases like mMCP-4 and both mouse and human cathepsin G (unpublished data). The focus is therefore now to test an alternative expression system, the baculovirus expression system, which has been used successfully to produce a number of related proteases with high yields.

In summary we have succeeded in getting a good view of the evolution of the chymase locus during the past 150 million years of mammalian evolution as well as the evolution of the granzyme A-K locus during 450 million years of vertebrate evolution. The phage display analysis of the macaque chymase has led to the identification of a new animal model to study the effects of chymase inhibitors aimed for human clinical use.

The work on granzyme A and K has also passed an important first level, the production of pure enzymatically active protease to be used in phage display analysis of their extended cleavage specificities, which in the future will serve as tools to identify novel in vivo targets for these important immune proteins.

6.0 References

1. Branden, C. and Tooze, J., 1999. *Introduction to protein structure*. Vol. 2nd Edition.
2. Rawlings, N.D. and A.J. Barrett, *Evolutionary families of peptidases*. Biochem J, 1993. 290 (Pt 1): p. 205-18.
3. Carter, P. and J.A. Wells, *Dissecting the catalytic triad of a serine protease*. Nature, 1988. 332(6164): p. 564-8.
4. Neurath, H., *Evolution of proteolytic enzymes*. Science, 1984. 224(4647): p. 350-7.
5. Davie, E.W. and H. Neurath, *Identification of a peptide released during autocatalytic activation of trypsinogen*. J Biol Chem, 1955. 212(2): p. 515-29.
6. Davie, E.W., K. Fujikawa, and W. Kisiel, *The coagulation cascade: initiation, maintenance, and regulation*. Biochemistry, 1991. 30(43): p. 10363-70.
7. Meyer-Hoffert, U., *Neutrophil-derived serine proteases modulate innate immune responses*. Front Biosci, 2009. 14: p. 3409-18.
8. Caughey, G.H., et al., *The human mast cell chymase gene (CMA1): mapping to the cathepsin G/granzyme gene cluster and lineage-restricted expression*. Genomics, 1993. 15(3): p. 614-20.
9. Gallwitz, M., J.M. Reimer, and L. Hellman, *Expansion of the mast cell chymase locus over the past 200 million years of mammalian evolution*. Immunogenetics, 2006. 58(8): p. 655-69.

10. Irani, A.A., et al., *Two types of human mast cells that have distinct neutral protease compositions*. Proc Natl Acad Sci U S A, 1986. 83(12): p. 4464-8.
11. Weidner, N. and K.F. Austen, *Heterogeneity of mast cells at multiple body sites. Fluorescent determination of avidin binding and immunofluorescent determination of chymase, tryptase, and carboxypeptidase content*. Pathol Res Pract, 1993. 189(2): p. 156-62.
12. Caughey, G.H., *Mast cell proteases as protective and inflammatory mediators*. Adv Exp Med Biol, 2011. 716: p. 212-34.
13. Amin, K., *The role of mast cells in allergic inflammation*. Respir Med, 2012. 106(1): p. 9-14.
14. Pham, C.T., *Neutrophil serine proteases: specific regulators of inflammation*. Nat Rev Immunol, 2006. 6(7): p. 541-50.
15. Reeves, E.P., et al., *Killing activity of neutrophils is mediated through activation of proteases by K⁺ flux*. Nature, 2002. 416(6878): p. 291-7.
16. Faurschou, M. and N. Borregaard, *Neutrophil granules and secretory vesicles in inflammation*. Microbes Infect, 2003. 5(14): p. 1317-27.
17. Tkalcevic, J., et al., *Impaired immunity and enhanced resistance to endotoxin in the absence of neutrophil elastase and cathepsin G*. Immunity, 2000. 12(2): p. 201-10.
18. Kobayashi, S.D., et al., *Neutrophils in the innate immune response*. Arch Immunol Ther Exp (Warsz), 2005. 53(6): p. 505-17.
19. Chowdhury, D. and J. Lieberman, *Death by a thousand cuts: granzyme pathways of programmed cell death*. Annu Rev Immunol, 2008. 26: p. 389-420.
20. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain*. Biochem Biophys Res Commun, 1967. 27(2): p. 157-62.
21. Greer, J., *Comparative modeling methods: application to the family of the mammalian serine proteases*. Proteins, 1990. 7(4): p. 317-34.
22. Czapinska, H. and J. Otlewski, *Structural and energetic determinants of the S1-site specificity in serine proteases*. Eur J Biochem, 1999. 260(3): p. 571-95.
23. Perona, J.J. and C.S. Craik, *Structural basis of substrate specificity in the serine proteases*. Protein Sci, 1995. 4(3): p. 337-60.
24. Perona, J.J. and C.S. Craik, *Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold*. J Biol Chem, 1997. 272(48): p. 29987-90.
25. Gallwitz, M. and L. Hellman, *Rapid lineage-specific diversification of the mast cell chymase locus during mammalian evolution*. Immunogenetics, 2006. 58(8): p. 641-54.
26. Reimer, J.M., P.B. Samollow, and L. Hellman, *High degree of conservation of the multigene tryptase locus over the past 150-200 million years of mammalian evolution*. Immunogenetics, 2010. 62(6): p. 369-82.
27. Baker, E., et al., *The genes encoding NK cell granule serine proteases, human tryptase-2 (TRYP2) and human granzyme A (HFSP), both map to chromosome 5q11-q12 and define a new locus for cytotoxic lymphocyte granule tryptases*. Immunogenetics, 1994. 40(3): p. 235-7.
28. Fink, T.M., et al., *The human granzyme A (HFSP, CTLA3) gene maps to 5q11-q12 and defines a new locus of the serine protease superfamily*. Genomics, 1993. 18(2): p. 401-3.
29. Masson, D., M. Zamai, and J. Tschopp, *Identification of granzyme A isolated from cytotoxic T-lymphocyte-granules as one of the proteases encoded by CTL-specific genes*. FEBS Lett, 1986. 208(1): p. 84-8.
30. Shresta, S., et al., *Residual cytotoxicity and granzyme K expression in granzyme A-deficient cytotoxic lymphocytes*. J Biol Chem, 1997. 272(32): p. 20236-44.
31. Grossman, W.J., et al., *The orphan granzymes of humans and mice*. Curr Opin Immunol, 2003. 15(5): p. 544-52.

32. Kam, C.M., D. Hudig, and J.C. Powers, *Granzymes (lymphocyte serine proteases): characterization with natural and synthetic substrates and inhibitors*. *Biochim Biophys Acta*, 2000. 1477(1-2): p. 307-23.
33. Quesada, V., et al., *Comparative genomic analysis of the zebra finch degradome provides new insights into evolution of proteases in birds and mammals*. *BMC Genomics*, 2010. 11: p. 220.
34. Praveen, K., D.L. Evans, and L. Jaso-Friedmann, *Evidence for the existence of granzyme-like serine proteases in teleost cytotoxic cells*. *J Mol Evol*, 2004. 58(4): p. 449-59.
35. Wernersson, S., et al., *Granzyme-like sequences in bony fish shed light on the emergence of hematopoietic serine proteases during vertebrate evolution*. *Dev Comp Immunol*, 2006. 30(10): p. 901-18.
36. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. *Bioinformatics*, 2007. 23(21): p. 2947-8.
37. Perriere, G. and M. Gouy, *WWW-query: an on-line retrieval system for biological sequence banks*. *Biochimie*, 1996. 78(5): p. 364-9.
38. Huang, R.Y., T. Blom, and L. Hellman, *Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three mouse mast cell-specific serine proteases*. *Eur J Immunol*, 1991. 21(7): p. 1611-21.
39. Gallwitz, M., M. Enoksson, and L. Hellman, *Expression profile of novel members of the rat mast cell protease (rMCP)-2 and (rMCP)-8 families, and functional analyses of mouse mast cell protease (mMCP)-8*. *Immunogenetics*, 2007. 59(5): p. 391-405.
40. Lutzelschwab, C., et al., *Characterization of mouse mast cell protease-8, the first member of a novel subfamily of mouse mast cell serine proteases, distinct from both the classical chymases and tryptases*. *Eur J Immunol*, 1998. 28(3): p. 1022-33.
41. Caughey, G.H., *Mast cell tryptases and chymases in inflammation and host defense*. *Immunol Rev*, 2007. 217: p. 141-54.
42. Puente, X.S. and C. Lopez-Otin, *A genomic analysis of rat proteases and protease inhibitors*. *Genome Res*, 2004. 14(4): p. 609-22.
43. Andersson, M.K., et al., *The extended substrate specificity of the human mast cell chymase reveals a serine protease with well-defined substrate recognition profile*. *Int Immunol*, 2009. 21(1): p. 95-104.
44. Kunori, Y., et al., *Rodent alpha-chymases are elastase-like proteases*. *Eur J Biochem*, 2002. 269(23): p. 5921-30.
45. Karlson, U., et al., *Extended substrate specificity of rat mast cell protease 5, a rodent alpha-chymase with elastase-like primary specificity*. *J Biol Chem*, 2003. 278(41): p. 39625-31.
46. Andersson, M.K., U. Karlson, and L. Hellman, *The extended cleavage specificity of the rodent beta-chymases rMCP-1 and mMCP-4 reveal major functional similarities to the human mast cell chymase*. *Mol Immunol*, 2008. 45(3): p. 766-75.
47. Wilharm, E., et al., *Generation of catalytically active granzyme K from Escherichia coli inclusion bodies and identification of efficient granzyme K inhibitors in human plasma*. *J Biol Chem*, 1999. 274(38): p. 27331-7.
48. Sattar, R., S.A. Ali, and A. Abbasi, *Bioinformatics of granzymes: sequence comparison and structural studies on granzyme family by homology modeling*. *Biochem Biophys Res Commun*, 2003. 308(4): p. 726-35.

2.1.1 Genes analyzed in the chymase locus.

In total 10 vertebrates have been analyzed for chymase locus genes and their homologous genes and other protein coding genes located in the chymase locus by using tblastn with known human chymase locus genes as bait.

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Human (<i>Homo sapiens</i>)	1	<i>CMAI</i>	Chymase	NP_001827
	2	<i>CTSG</i>	Cathepsin G	NP_001902
	3	<i>GZMH</i>	Granzyme H	NP_219491
	4	<i>GZMB</i>	Granzyme B	NP_004122

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Chimpanzee (<i>Pan troglodytes</i>)	1	<i>CMAI</i>	Chymase	XP_001170224
	2	<i>CTSG</i>	Cathepsin G	XP_522810
	3	<i>GZMH</i>	Granzyme H	XP_522811
	4	<i>GZMB</i>	Granzyme B	XP_509879

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Macaque (<i>Macaca mulatta</i>)	1	<i>CMAI</i>	Chymase	NP_001181562
	2	<i>CTSG</i>	Cathepsin G	XP_001114339
	3	<i>GZMH</i>	Granzyme H	XR_013129
	4	<i>GZMB</i>	Granzyme B	XP_001114420

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Dog (<i>Canis familiaris</i>):	1	<i>CMAI</i>	Chymase	NP_001013442
	2			
	3	<i>CTSG</i>	Cathepsin G	XP_850688
	4	<i>GZMH</i>	Granzyme H	XP_547751
	5	<i>GZMB</i>	Granzyme B	XP_547752

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Cattle (<i>Bos taurus</i>)	1	<i>CMA1a</i>	Mast cell protease 2 isoform 1	XP_593156
	2	<i>CMA1b</i>	Mast cell protease 2	XP_001790127
	3	<i>BDMD1</i>	Duodenase 1 precursor	NP_776721
	4	<i>BDMD2</i>	Duodenase 1 predicted	XP_002696733
	5	<i>BDMD3</i>	Duodenase 1 like	XP_001249717
	6	<i>BDMD4</i>	Duodenase 1	XP_585694
	7	<i>GZMH</i>	Granzyme H	XP_874639
	8	<i>CTSG1</i>	Cathepsin G1	XP_581980
	9	<i>CTSG2</i>	Cathepsin G2	XP_587026
	10	<i>GZMB</i>	Granzyme B	XP_585453

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Mouse (<i>Mus musculus</i>):	1	<i>CMA1/Mcp5</i>	Alpha Chymase	NP_034910
	2	<i>CMA2/Mcp10</i>	Mast cell protease 10 precursor	NP_001019885
	3	<i>MMCP-1</i>	Mast cell protease 1	NP_032596
	4	<i>MMCP-9</i>	Mast cell protease 9	NP_034912
	5	<i>MMCP-2</i>	Mast cell protease 2	NP_032597
	6	<i>MMCP-4</i>	Mast cell protease 4	NP_034909
	7	<i>MMCP-8</i>	Mast cell protease 8	NP_032598
	8	<i>CTSG</i>	Cathepsin G	NP_031826
	9	<i>GZME/CCP3</i>	Granzyme E	NP_034503
	10	<i>GZMD/CCP5</i>	Granzyme D	NP_034502
	11	<i>GZMG/Ctla-7</i>	Granzyme G	NP_034505
	12	<i>GZMN</i>	Granzyme N	NP_694692
	13	<i>GZMf/CCP4</i>	Granzyme F	NP_034504
	14	<i>GZMC/CCP2</i>	Granzyme C	NP_034501
	15	<i>GZMB/CCP1</i>	Granzyme B	NP_038570

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
---------	--------------	-----------	--------------	-----------------------

Rat (<i>Rattus norvegicus</i>):	1	<i>CMA1/Mcpt5</i>	Chymase	NP_037224
	2	<i>Mcpt4l1</i>	Mast cell protease 4 like 1	XP_573779
	3	<i>Mcpt4</i>	Mast cell Protease 4	NP_062194
	4	<i>Mcpt3/RMCP-3</i>	Mast cell Protease 3	NP_001163937
	5	<i>Mcpt1</i>	Mast cell Protease 1 precursor (chymotrypsin-like)	NP_058841
	6	<i>RGD1562035</i>	Mast cell Protease 1 like	XP_224209
	7	<i>Mcpt8/RMCP8</i>	Mast cell protease 8	NP_067609
	8	<i>Mcpt8l2</i>	Mast cell protease 8 like 2 precursor	NP_001128482
	9	<i>Mcpt1l2</i>	Mast cell protease 1 like 2	XP_573783
	10	<i>Mcpt1l4, CAE48391,</i>	Mast cell protease 1 like 4(TPA)	XP_573784
	11	<i>Mcpt10/GLP2</i>	Granzyme-like protein 2	NP_058842
	12	<i>Mcpt1l3</i>	Mast cell protease 1 like 3	XP_573785
	13	<i>Mcpt9 (Rat_rCG23443, isoform CRA_a)</i>	Mast cell protease 9	NP_062196
	14	<i>Mcpt2</i>	Mast cell protease 2	NP_742041
	15	<i>LOC100360872</i>	Protease 1 like	XP_002728330
	16	<i>RGD1565970</i>	Mast cell protease 8 like	XP_573787
	17	<i>CTSG</i>	Cathepsin G	NP_001099511
	18	<i>RGD1564878 (similar to NKP7)</i>	Granzyme G like	XP_577429
	19	<i>GZMN</i>	Granzyme N	NP_001178045
	20	<i>LOC691670 (Similar to NKP7)</i>	Uncharacterized protein LOC691670	NP_001103116
	21	<i>RNKP7 (GZMF/GZMG/RNKP7)</i>	Granzyme H	NP_703196
	22	<i>RGD1561819/NK cell protease 1 precursor/(GZMB)</i>	Granzyme like protein 1 like	XP_577431
	23	<i>LOC100361261</i>	rCG23560-like	XP_002725132
	24	<i>LOC691695</i>	Granzyme C like	XP_001079306
	25	<i>LOC691701</i>	Granzyme C like	XP_001079321
	26	<i>LOC100361468</i>	Granzyme-like protein 1-like	XP_002728303
	27	<i>GZMC/RNPK4</i>	Granzyme C	NP_599159
	28	<i>GZMB/GLP</i>	Granzyme-like	NP_612526

		<i>III/RNKP-1/ GLP-1/fragmentin</i>	protein 1 precursor	
--	--	-------------------------------------	---------------------	--

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Opossum (<i>Monodelphis domestica</i>)	1	<i>CMAI</i>	Chymase	XP_001369716.1
	2	<i>GZMB</i>	Granzyme B	XP_001369757.1

2.1.2 Genes analyzed in the granzyme A-K locus

Endothelial cell-specific molecule 1 (ESM1), granzyme K (GZMK), granzyme A (GZMA), cell division cycle protein 20 homolog B isoform 1 (CDC20B), probable glutathione peroxidase 8 GPX8, Geminin-like/IDAS, cyclin-O (CCNO), ATP-dependent RNA helicase DHX29 (DHX29), superkiller viralicidic activity 2-like 2 (SKIV2L2), lipid phosphate phosphohydrolase 1 isoform 2 (PPAP2A) potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 1 (HCN1).

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Human (<i>Homo sapiens</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	NP_008967.1
	2	<i>GZMK</i>	granzyme K	NP_002095
	3	<i>GZMA</i>	granzyme A	CAG33249.1
	4	<i>CDC20B</i>	cell division cycle protein 20 homolog B isoform 1	NP_689836.2
	5	<i>GPX8</i>	probable glutathione peroxidase 8	NP_001008398.2
	6	<i>EG622408/MCI/IDAS</i>	Geminin-like	NP_001177716.1
	7	<i>CCNO</i>	cyclin-O	NP_066970.3
	8	<i>DHX29</i>	ATP-dependent RNA helicase	NP_061903.2
	9	<i>SKIV2L2</i>	superkiller viralicidic activity 2-like 2	NP_056175.3
	10	<i>PPAP2A</i>	lipid phosphate phosphohydrolase 1 isoform 2	NP_795714.1
	11	<i>HCN1</i>	potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 1	NP_066550.2

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
---------	--------------	-----------	--------------	-----------------------

Chimpanzee (<i>Pan troglodytes</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	XP_518187.3
	2	<i>GZMK</i>	Granzyme K	XP_527196.1
	3	<i>GZMA</i>	Granzyme A	XP_001148261
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_001147385.1
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	NP_001230800.1
	6	<i>EG622408</i> <i>MCI/IDAS</i>	Geminin-like	XP_001147303.2
	7	<i>CCNO</i>	Cyclin-O	XP_001147153.1
	8	<i>DHX29</i>	ATP-dependent RNA helicase	XP_001147019.1
	9	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_001145794.1
	10	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_001146001.1
	11	<i>HCN1</i>	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 1	XP_001144667

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Macaque (<i>Macaca mulatta</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	ENSMMUP00000026559
	2	<i>GZMK</i>	Granzyme K	XP_001097338.1
	3	<i>GZMA</i>	Granzyme A	XP_001097639
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_001097843.2
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSMMUP00000002145
	6	<i>EG622408/MCI/IDAS</i>	Geminin-like	ENSMMUP00000034715
	7	<i>CCNO</i>	cyclin-O	XP_001098228
	8	<i>DHX29</i>	ATP-dependent RNA helicase	XP_001099143
	9	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSMMUP00000013605
	10	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSMMUP00000034702

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Mouse (<i>Mus musculus</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	NP_076101
	2	<i>GZMK</i>	Granzyme K	NP_032222.1
	3	<i>GZMA</i>	Granzyme A	NP_034500.1
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_138861.6
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	NP_081403
	6	<i>Geminin-like</i>	Geminin-like	ENSMUSP00000089721
	7	<i>CCNO</i>	Cyclin-O	NP_001074531.1
	8	<i>DHX29</i>	ATP-dependent RNA helicase	NP_766182.2
	9	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	NP_082427.1
	10	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	(NP_032273.1

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Rat (<i>Rattus norvegicus</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	NP_072126.1
	2	<i>GZMK</i>	Granzyme K	NP_058815.1
	3	<i>GZMA</i>	Granzyme A	NP_703198.1
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	NP_001008398
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	NP_001099881
	6	<i>Geminin-like</i>	Geminin-like	XP_002725936
	7	<i>CCNO</i>	Cyclin-O	ENSRNOP00000013871
	8	<i>DHX29</i>	ATP-dependent RNA helicase	XP_002725935.1
	9	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	NP_001029265.1
	10	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	NP_071983

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Dog (<i>Canis familiaris</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	XP_546319.2
	2	<i>GZMK</i>	Granzyme K	XP_546318.2
	3	<i>GZMA</i>	Granzyme A	XP_544335.2
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_535236.3
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	NP_001239252
	6	<i>Geminin-like</i>	Geminin-like	XP_003433839.1
	7	<i>CCNO</i>	Cyclin-O	XP_544336.2
	8	<i>DHX29</i>	ATP-dependent RNA helicase	XP_535238.2
	9	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_849676.1
	10	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_849687.2

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Cat (<i>Felis catus</i>)	1	<i>GZMK</i>	Granzyme K	ENSFCAP00000013032
	2	<i>GZMA</i>	Granzyme A	ENSFCAP00000007580
	3	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	ENSFCAP00000007581
	4	<i>DHX29</i>	ATP-dependent RNA helicase	ENSFCAP00000007582
	5	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSFCAP00000009820

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	NP_001091570.1
	2	<i>GZMK</i>	Granzyme K	NP_001192888.1
	3	<i>GZMA_1</i>	Granzyme A 1	NP_001092565.1

Cattle (<i>Bos taurus</i>)	4	<i>GZMA_2</i>	Granzyme A 2	NP_001001142.1
	5	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	NP_001193911
	6	<i>GPX8</i>	Probable glutathione peroxidase 8	NP_001039553.1
	7	<i>Geminin-like</i>	Geminin-like	DAA17916.1
	8	<i>CCNO</i>	Cyclin-O	XP_002696375
	9	<i>DHX29</i>	ATP-dependent RNA helicase	NP_001193063.1
	10	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	NP_001092482.1
	11	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	NP_001073798.1

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Elephant (<i>Loxodonta africana</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	XP_003408081
	2	<i>GZMK</i>	Granzyme K	XP_003407962
	3	<i>GzmA_like</i>	Granzyme A 1	XP_003407963
	4	<i>GZMA low quality</i>	granzyme A 2	XP_003407964
	5	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_003407965
	6	<i>GPX8</i>	Probable glutathione peroxidase 8	XP_003408083
	7	<i>Geminin-like</i>	Geminin-like	XP_003408085
	8	<i>CCNO</i>	Cyclin - O	XP_003407966
	9	<i>DHX29</i>	ATP-dependent RNA helicase	XP_003407967
	10	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_003407969
	11	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_003408086

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
---------	--------------	-----------	--------------	-----------------------

Rabbit (<i>Oryctolagus cuniculus</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	XP_002714077.1
	2	<i>GZMK</i>	Granzyme K	XP_002714022.1
	3	<i>GzmA_like</i>	Granzyme A 1	XP_002714023.1
	4	<i>GZMA_like 2</i>	Granzyme A 2	XP_002714079.1
	5	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_002714024.1
	6	<i>GPX8</i>	Probable glutathione peroxidase 8	XP_002714080.1
	8	<i>CCNO</i>	Cyclin-O	XP_002714025.1
	9	<i>DHX29</i>	ATP-dependent RNA helicase	XP_002714081.1
	10	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_002714082.1
	11	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_002714083.1

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Tree threw (<i>Tupaia belangeri</i>):	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	ENSTBEG00000001630
	2	<i>GZMK</i>	Granzyme K	ENSTBEG00000001611
	3	<i>GZMA</i>	Granzyme A	ENSTBEG00000002701
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	ENSTBEG000090003317
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSTBEG00000004779
	6	<i>DHX29</i>	ATP-dependent RNA helicase	ENSTBEG00000005039
	7	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSTBEG00000007769
	8	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSTBEG00000008812

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
---------	--------------	-----------	--------------	-----------------------

Squirrel (<i>Spermophilus tridecemlineatus</i>)	1	<i>GZMK</i>	Granzyme K	ENSSTOG00000015201
	2	<i>GZMA</i>	Granzyme A	ENSSTOG00000010768
	3	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	ENSSTOP00000009666
	4	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSSTOP00000009668
	5	<i>CCNO</i>	Cyclin - O	ENSSTOP00000009671
	6	<i>DHX29</i>	ATP-dependent RNA helicase	ENSSTOP00000009687
	7	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSSTOP00000009702
	8	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSSTOG00000010825

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Microbat (<i>Myotis lucifugus</i>)	1	<i>GZMK</i>	Granzyme K	ENSMLUP00000009254
	2	<i>CCNO</i>	Cyclin - O	ENSMLUP00000001023
	3	<i>DHX29</i>	ATP-dependent RNA helicase	ENSMLUP00000001045
	4	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSMLUP00000001067
	5	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSMLUP00000001070

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Armadillo (<i>Dasyus novemcinctus</i>):	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	ENSDNOP00000009559
	2	<i>GZMK</i>	Granzyme K	ENSDNOP00000010100
	3	<i>GZMA</i>	Granzyme A	ENSDNOP00000012397
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	ENSDNOP00000012398
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSDNOP00000012401
	6	<i>DHX29</i>	ATP-dependent RNA helicase	ENSDNOP00000011903
	7	<i>SKIV2L2</i>	Superkiller	ENSDNOP00000011927

			viralicidic activity 2-like 2	
	8	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSDNOP00000013780

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Opossum (<i>Monodelphis domestica</i>)	1	<i>ESM1</i>	Granzyme K	XP_001365467
	2	<i>Gzm A</i>	Granzyme A	XP_001380969.2
	3	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_001380974
	4	<i>GPX8</i>	Probable glutathione peroxidase 8	XP_001365538.1
	5	<i>Geminin</i>	Geminin-like	XP_001380992.2
	6	<i>CCNO</i>	Cyclin - O	XP_001381005.1
	7	<i>DHX29</i>	ATP-dependent RNA helicase	XP_001381009.1
	8	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_001365599.1
	9	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_001381015

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Platypus (<i>Ornithorhynchus anatinus</i>)	1	<i>GPX8</i>	Probable glutathione peroxidase 8	XP_001512035.1
	2	<i>CCNO</i>	Cyclin - O	XP_001507946.2
	3	<i>DHX29</i>	ATP-dependent RNA helicase	XP_001513873.2
	4	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_001513852.2
	5	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_001513815.2

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	XP_428357.3

Chicken (<i>Gallus gallus</i>)	2	<i>GZMK</i>	Granzyme K	XP_423832.2
	3	<i>GZMA</i>	Granzyme A	NP_989788.1
	4	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_423833.3
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	XP_423834
	6	<i>DHX29</i>	ATP-dependent RNA helicase	XP_424728
	7	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	NP_001012962.1
	8	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	XP_424730.3

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Western clawed frog (<i>Xenopus tropicalis</i>):	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	NP_001120192
	2	<i>GZMA1</i>	Granzyme A1	NP_001107158.1
	3	<i>GZMA2</i>	Granzyme A2	XP_002941155.1
	4	<i>GZMA3</i>	Granzyme A3	XP_002941159.1
	5	<i>CDC20B</i>	Cell division cycle protein 20 homolog B isoform 1	XP_002941160.1
	6	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSXETG00000028092
	7	<i>CCNO</i>	Cyclin - O	NP_998837.1
	8	<i>DHX29</i>	ATP-dependent RNA helicase	ENSXETP00000000784
	9	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	XP_002941157.1
	10	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	NP_001011077.1

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
---------	--------------	-----------	--------------	-----------------------

Fugu (<i>Takifugu rubripes</i>):	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	ENSTRUG00000012890
	2	<i>GZMA1</i>	Granzyme A1	ENSTRUP00000001550
	3	<i>GZMA2</i>	Granzyme A2	ENSTRUP000000043129
	4	<i>GZMA3</i>	Granzyme A3	ENSTRUP000000042784
	5	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSTRUG00000000697
	6	<i>CCNO</i>	Cyclin - O	ENSTRUG00000000724
	7	<i>DHX29</i>	ATP-dependent RNA helicase	ENSTRUG000000014726
	8	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSTRUG000000014571
	9	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSTRUG000000014543

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Medaka (<i>Oryzias latipes</i>):	1	<i>GZMA</i>	Granzyme A	ENSORLGG00000001790
	2	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSORLGG00000001726
	3	<i>EG622408</i>	Geminin-like	ENSORLGG00000001722
	4	<i>CCNO</i>	Cyclin - O	ENSORLGG000000014140
	5	<i>DHX29</i>	ATP-dependent RNA helicase	ENSORLGG00000005364
	6	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSORLGG00000005443
	7	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSORLGG00000005464

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Stickleback (<i>Gasterosteus aculeatus</i>)	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	ENSGACG000000016408
	2	<i>GZMA</i>	Granzyme A	ENSGACG000000004719

	3	<i>GPX8</i>	Probable glutathione peroxidase 8	ENSGACG00000004723
	4	<i>CCNO</i>	Cyclin - O	ENSGACG00000004733
	5	<i>DHX29</i>	ATP-dependent RNA helicase	ENSGACG00000016244
	6	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	ENSGACG00000016260
	7	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSGACG00000016270

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Zebrafish (<i>Danio rerio</i>):	1	<i>ESM1</i>	Endothelial cell-specific molecule 1	NP_001070209.1
	2	<i>GZMK</i>	Granzyme K	XP_001335166.1
	3	<i>GZMK</i>	Granzyme K	XP_001335076.2
	4	<i>GPX8</i>	Probable glutathione peroxidase 8	NP_956516.1
	5	<i>DHX29</i>	ATP-dependent RNA helicase	XP_697933.5
	6	<i>SKIV2L2</i>	Superkiller viralicidic activity 2-like 2	NP_999933.1
	7	<i>PPAP2A</i>	Lipid phosphate phosphohydrolase 1 isoform 2	ENSDARG00000053381

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Channel catfish (<i>Ictalurus punctatus</i>):	1	<i>GZM I</i>	Gzm_like_I	NP_001187138.1
	2	<i>GZM II</i>	Gzm_like_II	AAX36078.1
	3	<i>GZM III</i>	GZM_like_III	NP_001187139.1
	4	<i>GZM</i>	Gzm_precursor	NP_001187108.1

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Atlantic cod (<i>Gadus morhua</i>):	1	<i>GZMA/K</i>	granzyme A/K	AAV73910.1
	2	<i>GZMA putative</i>	putative granzyme A	AV73911.1

--	--	--	--	--

Species	No. of genes	Gene Name	Protein Name	Protein Accession No.
Atlantic salmon (<i>Salmo salar</i>):	1	<i>GZMA</i>	GzmA_precursor	NP_001134509.1

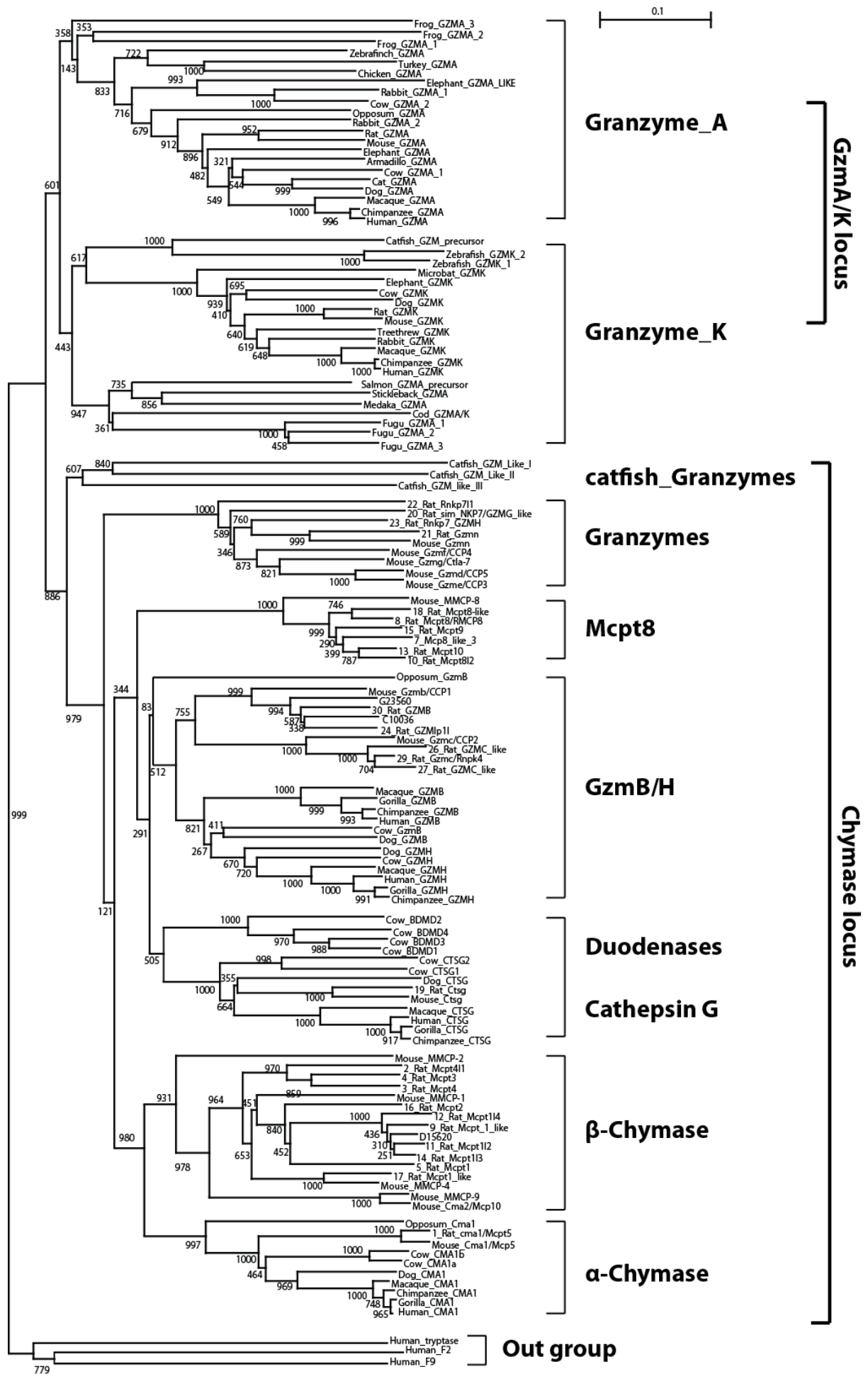


Figure 4. Phylogenetic analysis of protein sequences originating from genes within the chymase and the gzmA/K locus of various vertebrates. The phylogram is based on the alignment of sequences of the mature proteolytically active serine proteases. Numbers indicate 1000 independent bootstrap values analyzed by Neighbour-Joining method. Symbol “] ” with bold sentences indicate the family of genes they belong to.