Hematopoietic Serine Proteases from the Mast Cell Chymase and Tryptase Loci - a Functional and Evolutionary Analysis

JENNY REIMER
Abstract


Mast cells are key effector cells in allergic and inflammatory diseases. However, their primary role is most likely in host defence against parasitic and bacterial infections. Mast cells are a particularly rich source of serine proteases. These proteases belong to the chymase or the tryptase family, which are encoded from the mast cell chymase and the multigene tryptase loci, respectively. To better understand the biological functions and the molecular evolution of these enzymes we have studied the organisation of these two loci in species ranging from fish to human. We show that the mast cell chymase locus has evolved from a single founder gene to a complex locus during the past 200 Myr of mammalian evolution. Forty-five fish candidate genes for hematopoietic serine proteases were also identified. However, in phylogenetic analyses none of them grouped with individual branches holding mammalian mast cell chymase locus genes, indicating an independent parallel evolution in fish.

Studies of the evolution of the multigene tryptase locus showed that this locus has been highly conserved between marsupials and eutherians. However, no genes belonging to the individual subfamilies identified in eutherians could be identified in fish, amphibians or in birds, which also here indicates parallel evolution.

To study the evolution of specific cleavage specificities associated with these proteases, the extended cleavage specificity of opossum α-chymase was determined and found to be nearly identical to human mast cell chymase and the major mouse mast cell chymase mMCP-4. This indicates a strong pressure to maintain this specificity during mammalian evolution.

Basophils are rare blood cells with functions similar to mast cells that when mature almost completely lack mRNA. To study the proteome and to primarily characterize the granule protein content of basophils, an in vitro purification protocol was developed to obtain transcriptionally active umbilical cord blood-derived basophil precursors.

Keywords: immune system, mast cell, basophil, mast cell chymase locus, multigene tryptase locus, serine protease, evolution

Jenny Reimer, Department of Cell and Molecular Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden

© Jenny Reimer 2008

ISSN 1651-6214
ISBN 978-91-554-7179-8
urn:nbn:se:uu:diva-8676 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-8676)
To my Family
This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:


IV Reimer JM, Samollow PB and Hellman L. High degree of conservation of the multigene tryptase locus over the past 150-200 million years of mammalian evolution. *Manuscript*.


Reprints were made with permission from the publishers.
Contents

Introduction ...................................................................................................11
General overview .....................................................................................11
Mast cells and basophils - introduction .................................................12
  Mast cell and basophil origin and activation ......................................13
  Mast cell heterogeneity ......................................................................13
  Basophil characteristics ....................................................................14
Mast cell and basophil mediators - de novo synthesized .....................15
Mast cell and basophil mediators - preformed ....................................17
Tryptases ..............................................................................................18
Chymases .............................................................................................20
Serine proteases - biochemistry ..........................................................23
  Proteolytic mechanism of serine proteases ......................................24
  Chymotrypsin-like serine proteases; structure and substrate specificity ..25
  Regulation of serine protease activity ..............................................26
  Hematopoietic serine proteases .......................................................27
  The mast cell chymase locus ............................................................28
  The multigene tryptase locus ............................................................31
  The granzyme A/K locus ..................................................................33
  The granzyme M or elastase locus ....................................................33
An evolutionary perspective ..................................................................34
Present investigation ..................................................................................36
Aim .........................................................................................................36
Results and discussion ...........................................................................36
  Emergence and expansion of the mast cell chymase locus during mammalian evolution (paper I) ..........................................................36
  Hematopoietic serine proteases lacking the common disulphide bridge (Cys191-Cys220) are absent in teleosts (paper II) .......................38
  The cleavage specificity for mast cell α-chymases has been conserved during mammalian evolution, from opossum to human (paper III) .39
  Conservation of the multigene tryptase locus, including mast cell tryptases, over the past 185 million years of mammalian evolution (paper IV) .................................................................42
Abbreviations

aa  amino acid(s)
Ang  angiotensin
BLAST  basic local alignment search tool
CCR  chemokine receptor
CLC  charcot-leyden crystal
CPA  carboxypeptidase A
CS  chondroitin sulphate
CTMC  connective tissue mast cell
DPPI  dipeptidyl peptidase I
ECM  extracellular matrix
EST  expressed sequence tag
FceRI  high affinity receptor for IgE
GM-CSF  granulocyte macrophage-colony stimulating factor
Gzm  granzyme
HLA  human leukocyte antigen
IFN-γ  interferon-γ
IL  interleukin
Ig  immunoglobulin
ISP  implantation serine protease
kb  kilobases
kDa  kilodalton
LSMC  lymphatic sinus mast cells
LT  leukotriene
mAb  monoclonal antibody
Mb  megabases
MBP1  major basic protein 1
Mcpt  mast cell protease (gene)
Mcpt  mast cell protease (protein)
MC₇  tryptase-positive mast cell
MC₇C  tryptase- and chymase-positive mast cell
MMC  mucosal mast cell
m/rMCP  mouse/rat mast cell protease (protein)
Myr  million years
NGF  nerve growth factor
NK cell  natural killer cell
PAF  platelet activating factor
PG     prostaglandin
PRSS   protease serine member S
SCF    stem cell factor
Sim.   similar to
TGF-β  transforming growth factor-β
TLR    toll-like receptor
TMT    transmembrane tryptase
TNF-α  tumor necrosis factor-α
TSLP   thymic stromal lymphopoietin
Introduction

General overview

The immune system protects us against infections by viruses, bacteria and other potentially harmful organisms that we encounter in our everyday life. In mammals this system consists of two major branches, the innate and the adaptive immune systems. The innate immune system forms the first line of defence against various intruders. It can recognize common patterns on pathogens, e.g. bacterial cell wall components like muramyl dipeptides, components of the bacterial outer membrane like lipopolysaccharides (LPS), viral double-stranded RNA and viral or bacterial unmethylated CpG DNA. These common structures of microorganisms bind to receptors on various cells and trigger the immune system to respond. Moreover, physical barriers such as skin and mucosal membranes, phagocytic cells, leukocytes and numerous proteins like antibiotic peptides, proteases, lectins and complement components, together form the innate immune system. The innate response is not only important as a first line of defence but also very important for stimulating the adaptive immune system. The adaptive immune response is antigen-specific, i.e. it recognizes specific parts (antigens) of proteins or other macromolecules derived from the foreign invader. Specialized leukocytes; the B and T lymphocytes, mediate the adaptive immune response via production of antibodies and cell-mediated immunity (lysis of infected cells). In addition, innate effector mechanisms use molecules of the adaptive immune system, like immunoglobulins, to increase specificity. Eventually when the pathogen is eliminated, immunological memory remains, which will provide the host with an improved protection at a later reinfection.

Hematopoietic cells like neutrophils, eosinophils, natural killer (NK) cells, T cells, B cells, basophils and mast cells, are in different ways involved in these immune responses. In their cytolytic granules they store important effector molecules such as, histamine, heparin, antibacterial proteins and various enzymes, foremost serine proteases. In fact, serine proteases, like granzymes (Gzm), tryptases and chymases, constitute the major granule protein content of many of these cells. These enzymes are involved in a variety of immunological processes such as, induction of apoptosis in virus-infected cells, degradation of connective tissue components to facilitate im-
mune cell migration and processing of antibacterial peptides. However for most of these very abundant enzymes we have only a vague picture of their various functions in immunity and tissue homeostasis.

In this thesis I will discuss hematopoietic serine proteases, primarily the chymases and tryptases. These enzymes are encoded from two loci the mast cell chymase and the multigene tryptase loci. By investigating the origin of the genes encoded from these loci we try to retrieve important data concerning potentially conserved biological functions and improve our understanding of their molecular evolution. Moreover, focus will be on mast cells and basophils, cells regarded as key effector cells in allergic disorders. The main function of these cells is however most likely in host defense against certain bacterial and parasitic infections. The major protein content in mast cell granules is tryptases and chymases, whereas the granule proteins of basophils remain to be determined. Granule proteins are often used as cellular markers, e.g. chymases and tryptases for mast cells; however, the lack of human (Homo sapiens) basophil-specific granule proteins has for long hampered studies of these cells. Therefore a study characterizing in vitro umbilical cord blood-derived basophils in order to obtain transcriptionally active precursors and analyze their granule constituents is also presented. The focus of this thesis is also on the cleavage specificity of these various enzymes. By elucidating the specificity of the opossum (Monodelphis domestica) α-chymase, we aim at shedding light on the evolution of mast cell chymase cleavage specificity, primarily the α-chymases.

This thesis will begin with a description of mast cells and basophils and their granule content followed by a detailed description of the major focus of this thesis, the hematopoietic serine proteases.

Mast cells and basophils - introduction

Mast cells and basophils are potent inflammatory cells primarily known for their prominent role in allergic disorders. However, their primary function is most likely in protective immune responses against various pathogens. For example, mice lacking mast cells have been shown to have an impaired defense against infections by the intestinal worm, Trichinella spiralis (1) and by Gram-negative bacteria (2, 3). Basophils have been shown to be activated in vitro by viral- and/or parasite-derived antigens e.g. from human immunodeficiency virus (HIV) (4) and Schistosoma mansoni (5).

Mast cells mature and reside in tissue under the influence of stem cell factor (SCF), the ligand for c-kit (CD117), interleukin (IL)-3 and nerve growth factor (NGF, true for human) (6). Mast cells are distributed throughout the “surfaces” of the body, e.g. in the mucosa of airways and intestine, in connective tissue of the skin and around blood vessels and nerves. They are scattered strategically at the sites where pathogens and environmental anti-
gens enter the body. In contrast, basophils circulate in the blood and can be recruited into tissues when needed. The granule protein content of various mast cell populations have been well described, however concerning the granule protein content of basophils very little is known.

Mast cell and basophil origin and activation
Mast cells and basophils originate from a common bone marrow-derived CD34+ progenitor cell (7). Mast cells are relatively long-lived and can increase in number or change phenotypically under certain conditions, e.g. inflammation and immune responses, due to the state of the extracellular environment (8). On the contrary, basophils are short lived and barely proliferate after they have left the bone marrow. Furthermore, mast cells and basophils share morphological features and show similarities in granule contents and surface molecules. Both cell types stain metachromatically with basic dyes, express the high affinity immunoglobulin (Ig) E receptor (FcεRI) and produce and store histamine and the cytokines IL-4 and IL-13. Upon cross-linking of receptor-bound IgE with antigen, mast cells and basophils degranulate and release a variety of inflammatory mediators, either preformed or de novo synthesized (9). Interestingly, it has been shown that human and rodent mast cells are able to reconstitute their granules after degranulation (10,11). Moreover, mast cells and basophils can be activated via binding of the complement components C3a and C5a, the anaphylatoxins, to their respective receptors. Human mast cells under the influence of interferon (INF)-γ can upregulate FcεRI and consequently can be activated via IgG (12). The inhibitory FcεRIIb in coligation with FcεRI has been shown able to inhibit murine mast cell and human basophil activation (13-15). Moreover, mast cells and basophils can be activated in an innate manner via various toll-like receptors (TLRs). For example, basophils secrete IL-4 and IL-13 when activated through TLR-2 (16), whereas mast cells are activated and produce type 1 interferons upon binding of double-stranded RNA to TLR-3 (17).

Mast cell heterogeneity
Mast cells do not represent a homogeneous cell population. Human mast cells are divided into at least two different subtypes based on their protease content. Mast cells expressing both tryptase and chymase are depicted as MCτC cells and mast cells expressing only tryptase have been depicted MCτ cells (see Table 1 for mast cell/basophil characteristics)(18). Both subtypes store the proteoglycans heparin and chondroitin sulphate (19). In addition, MCτC also express a metalloprotease, carboxypeptidase A (CPA) (20). Rodent mast cells are, based on tissue localization and staining properties, divided into connective tissue mast cells (CTMCs), which primarily are found
in the gastrointestinal tract, peritoneum and in the skin, and mucosal mast cells (MMCs), which are localized to the lamina propria of the respiratory tract and in the submucosa of the intestine (21). CTMCs store tryptase, chymase, CPA, histamine and heparin (22). In contrast, MMCs store chymase, low amounts of histamine and chondroitin sulphate (23). Accordingly, human MC<sub>TC</sub> resemble rodent CTMCs considering their granule content and to some extent their tissue distributions.

Interestingly, in marsupials three subtypes of MCs have been described; the well-established MMCs and CTMCs (24), and a third subtype not found in placental mammals termed lymphatic sinus mast cells (LSMCs) (25, 26). The LSMCs can be distinguished from the CTMCs by having three times larger cytoplasmic granules. The three marsupial mast cell subtypes all contain heparin (24). Moreover, heterogenic mast cell populations have been reported to be present in nonmammalian vertebrates such as birds (27), amphibians (28), reptiles (29) and numerous fish species (30). Mast cell-like cells, containing heparin and histamine, have been identified at epithelial surfaces of several invertebrates, including molluscs (31) and tunicates (32). Notably, mast cells and mast cell-like cells are present in distantly related species and thus this indicates that they must fulfil important function in their hosts.

**Basophil characteristics**

In contrast to mast cells, basophils mature in the bone marrow and thereafter circulate in the peripheral blood from where they can migrate into tissues upon proper stimulation (7). Circulating basophils are terminally differentiated; therefore almost no mRNA for their granule content can be obtained. This has long hampered the elucidation of their biochemical functions and properties.

Basophils belong to the granulocyte lineage and represent approximately 0.5% of the leukocytes found in blood. The common basophil/mast cell precursor express the novel mast cell/basophil lineage-specific ectoenzyme 97A6/CD203c (33, 34). CD203c is also expressed on mature mast cells and basophils, although a subset of mast cells is CD203c-negative. In addition to CD203c, a few basophils-specific markers have been identified; BB-1 (CD63) (35, 36) and 2D7 (37), two granule-associated antigens, and Bsp-1 (basophil-specific protein 1) (38), a cell surface antigen. In mice (*Mus musculus*), the serine protease mouse mast cell protease (mMCP) -8 is found exclusively in basophils (39), whereas the rat (*Rattus norvegicus*) counterparts rMCP-8, -9 and -10 are detected in MMCs (40). In a recent study Gallwitz et al. tried to characterize the cleavage specificity of mMCP-8 utilizing chromogenic substrates and substrate phage display (41). However, no proteolytic activity could be detected. They suggested that mMCP-8 may
possess a narrow substrate-binding region due to two extra cysteines, or that it is active as a homodimer.

Basophils express high levels of the low affinity receptor for IL-3 (IL-3Rα, CD123) but not c-kit (42), they also express different chemokine receptors (CCRs), which are important for the recruitment of basophils into inflamed tissue (43). For instance, CCR3 is expressed at high levels and upon binding of the eosinophil-derived ligand eotaxin, migration of basophils is induced (44). The main growth and differentiation factor for basophils is IL-3. Several studies have used IL-3 in trying to derive basophils from bone marrow, cord blood or peripheral blood progenitors (45-47). In most cases small populations of basophils with varying purities were obtained. Surprisingly, the number of basophils is not reduced in IL-3 deficient mice, indicating that other growth factors are also important for basophil development (48). IL-5, granulocyte-macrophage colony-stimulating factor (GM-CSF), transforming growth factor-β (TGF-β) and nerve growth factor (NGF) have been used in attempts trying to culture basophils (reviewed in (43)).

The existence of a common basophil/mast cell progenitor, a basophil/eosinophil (both being granulocytes) progenitor or a basophil/megakaryocyte progenitor has for a long time been a matter of debate. In vitro, basophils and eosinophils are often observed in the same cell colonies. Injection of IL-3 and GM-CSF in humans has been shown to induce basophilia and eosinophilia and increased number of progenitor cells for both cell types (49). Moreover, an eosinophil major basic protein 1 (MBP1) - like molecule has been identified in basophils (50, 51). Recently, a monoclonal antibody (mAb, J175-7D4) targeting the proform of MBP1 was shown to stain basophils but not eosinophils, lymphocytes, neutrophils, monocytes or skin mast cells (52). Mature eosinophils mainly store the active form of MBP1 in their granules (53). Thus, the mAb J175-7D4 provides a novel marker for basophils. Moreover, basophils and eosinophils store Charcot-Leyden crystals (CLC)/galectin-10 in their cytoplasmic granules (54, 55). Until recently, CLC was thought to be exclusively expressed in basophils and eosinophils, however, Kubach et al. recently showed, that CLC is expressed in human CD4+ CD25+ regulatory T cells and is of importance for their suppressive functions (56).

Mast cell and basophil mediators - de novo synthesized

Mast cells and basophils initiate the production of cytokines and lipid mediators such as leukotrienes (LT) and prostaglandins (PG) upon activation. The lipid mediators are metabolites of arachidonic acid, which is modified by enzymes of the cyclooxygenase and lipoxygenase pathways.
Leukotrienes and prostaglandin

Leukotriene C\textsubscript{4} and its degradation products, LTD\textsubscript{4} and LTE\textsubscript{4} (also termed cysteinyl leukotrienes; cys-LTs) are the main lipoxygenase products produced by mast cells and basophils. Cys-LTs binds to two divergent G protein-coupled receptors, CysLT\textsubscript{1} or CysLT\textsubscript{2}, which are unevenly distributed throughout the body. Numerous activities have been ascribed to Cys-LTs, including their contribution to the pathogenesis of asthma e.g. causing bronchoconstriction and stimulating increased mucus secretion and vascular permeability (57). Human mast cells also produce LTB\textsubscript{4} that functions as an important chemoattractant for monocytes, neutrophils, eosinophils and interestingly human cord blood-derived mast cell progenitors (reviewed in (58)) (59). By stimulating production of proinflammatory cytokines and mediators LTB\textsubscript{4} further enhances inflammation.

Prostaglandin D\textsubscript{2} is secreted primarily by CTMCs and is the major product generated via the cyclooxygenase pathway (60). PGD\textsubscript{2} can signal via several different receptors and at different concentrations, resulting in both pro- and anti-inflammatory effects. Binding of PGD\textsubscript{2} to the novel, chemoattractant-receptor homologous molecule expressed on Th2 cells (CRTH2), induce eosinophil (61) and basophil (62) migration and up-regulation of adhesion molecules on these cells. In contrast, signaling via the PGD\textsubscript{2} receptor DP inhibits basophil migration and degranulation (63).

Platelet-activating factor (PAF) is a proinflammatory phospholipid mediator secreted by, among others, mast cells and basophils (reviewed in (64)). PAF retrieved its name when it was first described as responsible for the aggregation of platelets after release from rabbit basophils upon IgE stimulation (65). PAF has been implicated in an extensive range of processes including angiogenesis, inflammation and wound healing. Furthermore it is involved in key-manifestations of allergy and asthma, such as bronchoconstriction and relaxation of vascular smooth muscles. Basophils are thought to be the main source of PAF in late-phase allergic reactions.

Cytokines

Mast cells both store and de novo synthesize cytokines. Tumor necrosis factor (TNF)-\textalpha is a major preformed cytokine in the mast cell. TNF-\textalpha is involved in several pathological and physiological processes e.g. neutrophil recruitment during late-phase reactions (66) and in immunity against bacteria (2, 3). TNF-\textalpha mediates these effects by stimulating endothelial cells to upregulate adhesion molecules and to produce chemokines, which attracts neutrophils, monocytes and lymphocytes to the site of infection. Activated mast cells secrete numerous cytokines including IL-1\textalpha/\textbeta, IL-3, IL-4, IL-5, IL-6, IL-10, IL-13, IL-16 and GM-CSF. IL-3, IL-5, and GM-CSF are important for eosinophil development and survival (67). IL-1 and IL-6 secreted by mast cells are important proinflammatory cytokines.
It was long believed that basophils foremost produced, and secreted, the pro-allergic cytokines IL-4 and IL-13 (68, 69). However, in a recent paper basophils activated by protease allergens were shown to produce, IL-2, IL-31, thymic stromal lymphopoietin (TSLP) and a chemokine, CCL1 (70). Moreover, IL-25 was recently shown to be secreted by human basophils, eosinophils (71) and murine bone marrow-derived mast cells (72). IL-25 is thought to augment allergic reactions by enhancing the expansion of Th2 memory cells and stimulate production of Th2 promoting cytokines. IL-4, IL-13 and TSLP skew the development of T cells into Th2 cells and act as autocrine factors on Th2 cells to enhance the IL-4 production by these (70, 73, 74). Antigen activated Th2 cells can thereafter stimulate B cells to become antigen-specific IgE-producing cells, which in turn can bind to mast cells and basophils that can be activated. Although the Th2 environment is associated with allergy and asthma, it is a desired condition when fighting parasitic infections. Mice deficient in IL-4 and IL-13 show an impaired capacity to expel helminth parasites (75).

Mast cell and basophil mediators - preformed
Mast cells and basophils have large cytoplasmic granules, which contain a number of physiologically potent mediators that are released upon activation. Mast cells store in these granules histamine, serine proteases, CPA and heparin and chondroitin sulphate E proteoglycans. Basophils store histamine and chondroitin sulphate. Serine proteases i.e. chymases and tryptases will be discussed in separate sections below.

Proteoglycans
Proteoglycans contribute to the storage of histamine and proteases. Proteoglycans are negatively charged and consist of a core protein to which different glycosaminoglycan side chains are bound i.e. heparin and chondroitin sulphate A and E (19). In mast cells, serglycin is the main core protein (19, 76).

Histamine
Histamine is a biogenic amine stored in granules together with proteoglycans, and after degranulation it is released into the extracellular milieu. Histamine binds to specific G-protein coupled receptors H1-H4 (77) in various tissues, which can cause increased vascular permeability and contraction of bronchial and intestinal smooth muscle. These manifestations are associated with allergic conditions like asthma, edema and diarrhoea. Basophils are the main source of histamine found in human blood, while mast cells account for the majority of histamine found in tissue (78).
Carboxypeptidase A

Carboxypeptidase A is exclusively produced by human MC<sub>TC</sub> and mouse CTMCs (79, 80). CPA is a zinc-dependent metalloprotease that cleaves substrates at the carboxyl (C), -terminal end, preferentially after bulky aliphatic or aromatic residues. Together with chymase (mMCP-5), CPA is stored in granules as macromolecular complexes associated to heparin proteoglycan (81, 82). The proteases are stored in their active form and remain bound to the proteoglycan after exocytosis. Henningsson <i>et al.</i> recently showed that heparin was necessary for the processing of pro-CPA to its active form (83).

<i>In vitro</i> CPA together with mMCP-4 can convert angiotensin (Ang) I to the Ang II, and then further degrade Ang II to abolish the activity (84).

Tryptases

Tryptases, are serine proteases with trypsin-like cleavage specificity i.e. they preferably cleave after basic amino acids (aa), like Arg and Lys (85). Tryptases have been found in most orders of placental mammals, represented by human (primata), mouse, rat and gerbil (<i>Meriones ungiulatus</i>, rodentia), dog (<i>Canis familiaris</i>, carnivora), sheep (<i>Ovis aries</i>) and pig, (Sus scrufa, artiodactyla) cow (<i>Bos taurus</i>) and horse (<i>Equus caballus</i>, perissodactyla). Human MC<sub>TC</sub> and MC<sub>T</sub>, as well as rodent CTMCs, express and store tryptase, which has been shown to account for up to one fourth of their total protein content (86). The mast cell tryptase genes are more numerous in human than in rodents. These enzymes are encoded by a locus that resides on human chromosome 16 and mouse chromosome 17 (for details see section “The multigene tryptase locus”). In contrast to most hematopoietic serine proteases most tryptase genes are composed of six exons and five introns (87). Tryptases are produced as zymogens, thus they are activated in a two-step procedure and finally stored in granules as fully active enzymes. It has been suggested that dipeptidyl peptidase I (DPPI) may play a role in the amino (N)-terminal processing of tryptase (88). However, mice deficient for DPPI have mature tryptase, although at lower levels (89). Human mast cell tryptases include, α-, β-, γ-, and δ-tryptase. Mice express three mast cell tryptases; mMCP-6, mMCP-7 and γtryptase (also referred to as transmembrane tryptase, TMT). In addition, a novel tryptase, denoted mMCP-11 is present in mouse (90). mMCP-11 has been shown to be expressed in bone marrow-derived mast cells and mast cell-like cell lines (91). mMCP-6 and mMCP-7 is expressed in CTMC and possess different substrate specificity (92, 93).

Several studies have detected significant amounts of α- and β-tryptases in human basophils (94, 95), and further Li <i>et al.</i> found substantial amounts of tryptase and chymase in basophilic cells in asthma and allergy patients (96). In addition, predominately β-tryptase has been found in the basophil leukemia cell line, KU812 (97).
**α and β-tryptases**

Three different β-tryptase cDNAs have been cloned, βI, βII and βIII (87, 98). βII and βIII are allelic variants, and βI is an allelic variant to the human α-tryptase. The β-tryptases display 98-99% sequence identity.

β-tryptases cleaves after P1 Arg and Lys. Phage display has been utilized to determine the extended cleavage specificity; Harris et al. showed that βI- and βII-tryptase had a preference for Pro in P4-, Arg/Lys in P3- and Gln in the P2-position (99). Moreover, βI-tryptase has a general preference for substrates holding Pro residues in complex with heparin proteoglycan (101, 102). The negatively charged heparin interacts with positively charged areas on the surface of adjacent monomers (101). The active site of each monomer faces towards a central pore, contributing to resistance against natural inhibitors. Moreover, the size of the pore restricts which substrates that can be cleaved.

Two highly similar α-tryptases, αI (98) and αII (103) have been identified. Due to an aa substitution in the propeptide α-tryptase is constitutively secreted as a proprotease (88). Although, recombinant α-tryptase has been shown to be processed in a correct manner and to be secreted in an active form, the protease activity is however restricted due to an Asp_{216} (Gly_{216} in β-tryptase) in the substrate-binding region (104). Approximately 25% of the population is deficient for α-tryptase, thus inheriting two copies of βI (105). Occasionally, β-tryptase is also secreted as an inactive precursor, although α-tryptase is the major form found in serum (106, 107) and in the synovial fluid of rheumatoid arthritis patients (108).

Mast cell tryptases are central in the development of inflammation and have a major role in asthma. Tryptase can induce eosinophil and neutrophil infiltration (109), stimulate fibroblasts and the subsequent release of collagen (110) and stimulate an epithelial cell line (H292) to secrete IL-8 (111). However, relatively few natural substrates have been identified, and the enzymes may have a restricted number of natural substrates due to its narrow substrate-binding region. Tryptase inhibitors have proven successful in the treatment of asthma in animal models, although low effects have been shown in humans, reviewed in (112).

**γ-tryptases**

γ-tryptases (TMT) are transmembrane bound proteases encoded from the multigene tryptase locus in both human and mice (113, 114). Two allelic variants have been identified; γI and γII (113). These are expressed in intestine, lung and skin mast cells and after degranulation a C-terminal hydrophobic extension anchors the tryptase to the membrane. There they are strategically positioned to interact with targets in the extracellular milieu. Phylogenetic analyses suggest that γ-tryptases diverged from the soluble tryptase...
ses prior to a common ancestor of human and mice (reviewed in (115)). Considering gene structure and preprosequence, γ-tryptases resemble prostatasin (116) and testisin (117) more than the neighboring α/β-tryptases. Recombinant soluble human γ-tryptase can provoke airway hyperresponsiveness and increased expression of proallergic IL-13 in mice after intratracheal administration (118).

δ-tryptases
When first identified, human δ-tryptase was termed “mMCP-7-like tryptase” due to homology between its fifth exon and mMCP-7 (103, 119). However, to follow the common nomenclature for human mast cell tryptases it was later termed δ-tryptase (120). Two allelic variants exist; δI and δII, which differ in only one aa. Compared to α- and β-tryptases, δ-tryptase is approximately 40 aa shorter, due to a nonsense mutation (103). Interestingly, δ-tryptase also possesses a mutation at aa position -3 (ArgGln) like human α-tryptase, that might influence the activation of the protease. Recombinant δ-tryptase was shown to have trypsin-like cleavage specificity. RT-PCR revealed that δ-tryptase is expressed in several tissues e.g. colon and lung, and immunohistochemistry detected δ-tryptase protein in colon, lung and inflamed synovium into cells morphologically similar to mast cells. In a recent study, Caughey and co-workers showed that the δ-tryptases arose early during primate evolution, and that the propeptide mutation occurred between 6 and 9 Myr ago and that the mutation generating a shortened protease arose 9 and 16 Myr ago (121). Moreover, it was shown that macaques express an active δ-tryptase, which is transcribed (lung) at higher levels than macaque α and β-tryptases.

Chymases
After tryptases, chymases are the most abundant proteins in mast cell granules. As early as in 1959, a protease with activity similar to that of bovine chymotrypsin was described in mast cells (122). Chymases preferentially cleave after P1 aromatic aa such as, Phe, Tyr and Trp (123). Chymases are synthesized as zymogens. After activation by DPPI, they are stored in granules tightly bound to chondroitin sulphate or heparin proteoglycans (89). After degranulation, chymases remain bound to the proteoglycan, and interestingly, in vitro studies have recently shown that cleavage of various substrates is promoted by protease-heparin interactions (124, 125). Proteoglycans further provide a protection against extracellular protease inhibitors, including serpins and α2-macroglobulin (126, 127). Based on phylogenetic differences, chymases are classified as either α- or β-chymases (128). In human, only one chymase (CMA1) gene is present, and it belongs to the α-chymase family (129). Human α-chymase is expressed and stored in granules of human MC_{TC}. In contrast, mouse has one α-chymase, mMCP-5
(130), and four β-chymases, mMCP-1, -2, -4 and -9 (130-132). mMCP-1, -2 and -9 are expressed by MMCs and mMCP-4 and -5 by CTMCs. CTMC and MMC chymases are stored in complex with heparin and chondroitin sulphate proteoglycans, respectively. Three-dimensional models of four mouse chymases revealed two areas of positively charged aa distant to the substrate-binding region of mMCP-4 and -5 (133). In contrast, mMCP-1 and -2 have a lower surface net charge and therefore bind less tightly to proteoglycans. α-chymases has been found in all mammalian species investigated, although β-chymases are most certainly rodent-specific. Nevertheless, Gallwitz et al. speculate that a dog β-chymase once was present, although a stop codon interrupts the aa sequence, and no expression has been detected (for more details, see section “The mast cell chymase locus”) (134).

The rodent α-chymases, mMCP-5 and rMCP-5, and the hamster α-chymase (HAM2) were recently shown to be elastases, thus cleaving substrates C-terminal of Ala, Ile and Val residues (135-137). The altered cleavage specificity is suggested to depend on the substitution of Gly216 to Val216 (135, 138). Furthermore, rMCP-5 was shown to have a high specificity for acidic aa in P2′ (136). The extended substrate specificity partly depends on an extra cavity formed in this subclass of serine proteases, lacking the common (most serine proteases) fourth disulphide bridge (Cys191/Cys220). Overall and interestingly the extended cleavage specificity for rMCP-5 is comparable to that of the human chymase (139) (Andersson unpublished). However, most likely mMCP-1 and mMCP-4 are the functional homologs of human CMA1 (140-142). Moreover, it was recently shown that a guinea pig α-chymase had preference for P1 Leu and did not cleave after aromatic aa residues (143). Clearly, phylogeny and function do not always walk hand in hand.

Several studies have addressed questions concerning potential substrates and physiological roles for chymases. The most well characterized substrate is the decapeptide Ang I, which acts as a vasoconstrictor when properly processed. Ang I can be cleaved at the Phe8-His9 bond, generating Ang II and subsequently Ang II can be degraded by cleavage of the Tyr4-Ile5 bond. In a previous study it was suggested that α-chymases were Ang I-converting enzymes, whereas β-chymases were both Ang I-converting and Ang II-degrading enzymes (128). However, this has been re-evaluated since conflicting data have been presented for several chymases (141, 142, 144, 145). mMCP-4 together with CPA seem to be key-enzymes for Ang I-conversion and degradation in mice (84, 141). Overall, angiotensin-converting enzyme (ACE) is the main Ang I activating enzyme. Furthermore, chymases cleave precursors of the proinflammatory cytokine IL-1β (146) and the potent vasoconstrictor endothelin-1 (147, 148). Chymases are involved in tissue remodelling by degrading proteins of the extracellular matrix (ECM), including collagen (149), fibronectin (124, 150) and vitronectin (151). Heikkilä et al. recently showed that mast cell derived chymase and TNF-α in concert in-
duced apoptosis of endothelial cells \textit{in vitro} (152). Initially chymase degrades fibronectin and vitronectin, thereby interrupting essential cell survival-signals, and subsequently TNF-\(\alpha\) triggers apoptosis. These events are suggested to be involved in plaque erosion associated to atherosclerosis.

Human chymase has been shown \textit{in vivo} to recruit a vast array of inflammatory cells, including eosinophils, lymphocytes, mast cells, macrophages and neutrophils (153-158). These effects are most probably due to activation and secretion of various cytokines and chemokines in response to chymase. Further, chymases are associated with increased epithelial permeability that facilitates the migration of inflammatory cells and effector molecules.

Valuable data regarding the biological functions of chymases have been retrieved from knockout mice. However, the presence of multiple rodent chymase genes and a not fully clarified picture of rodent-human homology of these genes make it difficult to extrapolate animal data to human. To date, mMCP-1, -4 and -5 knockout mice exist; one MMC-chymase, one CTMC-chymase and one CTMC-chymase with elastase-activity. Mice lacking \(\beta\)-chymase mMCP-1 have an impaired defence against the nematode, \textit{T. spiralis} and an increased deposition of larvae in muscles (159). Moreover, mMCP-4 deficient mice were shown to have a reduced capacity to degrade fibronectin and inactivate thrombin (160). It was also shown that mMCP-4 is important for activation of two metalloproteases, MMP-2 and MMP-9 (161). Interestingly, dog mastocytoma cells and canine bone marrow derived mast cells secrete and activate MMP-9 (162-164). These findings further illustrate the central role for mast cells in tissue remodelling and a role for chymases in the degradation of ECM. Finally, it was shown in mice lacking mMCP-5 that skeletal muscle injuries due to ischemia-reperfusion were less severe (165).
Table 1. Characteristics of human mast cells and basophils. Human and mouse, mast cell subclasses and their proteoglycan and protease content are also included. +; expressed, low; expressed at low levels, CS; chondroitin sulphate.

<table>
<thead>
<tr>
<th>BASOPHILS</th>
<th>MAST CELLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major growth factor</td>
<td>SCF</td>
</tr>
<tr>
<td>Surface markers</td>
<td>CD34, CD123, FcεRI, CD203c</td>
</tr>
<tr>
<td>Histamine</td>
<td>+</td>
</tr>
<tr>
<td>Prostaglandins, PGD₂</td>
<td>+</td>
</tr>
<tr>
<td>Leukotrienes, LTC₄</td>
<td>+</td>
</tr>
<tr>
<td>Cytokines</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mast cell subclasses, human and mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCₜ</td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Proteases</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tryptase</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Others</td>
</tr>
</tbody>
</table>

Serine proteases - biochemistry

Approximately one-third of all proteases are classified as serine proteases. Members of this family have been identified across all kingdoms of life, indicating the usefulness of this enzymatic mechanism. The proteolytic activity is based on a nucleophilic Ser residue, being part of the catalytic triad, together with a histidine (His) and an aspartate (Asp) residue (166). The order in which these three residues appear in the peptide chain has been shown to vary among serine proteases and several different variants exist, indicating a convergent evolution of this catalytic mechanism (167). Based on the context of the triad, or the charge-relay system, serine proteases utilizing Ser, Asp and His, are divided into at least four subfamilies, the chymotrypsin, the subtilisin, the carboxypeptidase Y and the Clp protease families. In addition, other protease families exist that use Ser in the catalytic site. However, then in combination with other aa residues than solely Asp and His (167).
Serine proteases of the chymotrypsin family are endopeptidases i.e. they cleave peptide bonds within peptide chains. For an enzyme to be able to cleave a substrate, the substrate must fit into the active site, also termed the substrate-binding pocket (S1). Cleavage occurs at the so-called scissile bond, and residues at the N-terminal side (of the cleaved bond) are termed P1, P2, P3 and so on, and the residues at the C-terminal side are termed P1’ (prime), P2’, P3’, etc. (Figure 1). The corresponding residues of the protease are termed S1/S1’, and so on (168). Depending on the preferred aa at the P1 position, the primary specificity of a protease can be determined. Moreover, residues surrounding the substrate-binding pocket also interact with the substrate; these residues determine the extended cleavage specificity of a protease. The function of a certain protease is also characterized by its preferred substrate(s) and when and where cleavage occurs.

Proteolytic mechanism of serine proteases

The catalytic mechanism of serine proteases is based on the interaction of a nucleophilic Ser residue, an acidic Asp residue and a His residue acting as a general base. The reaction can be divided into two steps, acylation and deacylation. The latter is basically a reverse acylation reaction (169). The catalytic mechanism of serine proteases has been the subject of many extensive reviews, e.g. (167, 169-171).

Acylation; The His residue forms a hydrogen bond with the oxygen of Ser, thereby making Ser more nucleophilic. Hydrogen bonds between His and Asp stabilize the His residue. The active Ser can then attack the carbonyl carbon of the substrate scissile bond. A tetrahedral intermediate is then formed between the Ser (and Gly193) and the substrate, which is stabilized via hydrogen bonds interacting with amide nitrogens, coordinating the so-called oxyanion into the oxyanion hole. The tetrahedral intermediate col-
lapse, an acyl-intermediate is formed and the C-terminal part of the substrate is released. **Deacylation**; A water molecule attacks the ester bond between the substrate and the enzyme. One hydrogen is donated to the His residue and again a tetrahedral intermediate is formed. His returns the hydrogen to the Ser residue and the intermediate decomposes. The substrate is released and the enzyme is restored.

Chymotrypsin-like serine proteases; structure and substrate specificity

The family of chymotrypsin-like serine proteases is one of the most thoroughly investigated. With over 450 identifiers in the MEROPS database it is also the largest protease family and it is constantly growing as new proteases are added (172). The name-giving protease, bovine chymotrypsin A, is composed of 245 aa, with the residues of the catalytic triad at positions, His$_{57}$, Asp$_{102}$ and Ser$_{195}$ (chymotrypsin numbering, according to (173)). Adjacent to the catalytic triad, the S$_1$ primary substrate-binding pocket, formed by residues 189-192, 214-216 and 224-228, are located. Especially residue 189, at the bottom of the pocket and residues 216 and 226, at opposite sides of the pocket, are of utmost importance for substrate recognition (Figure 2) (170). Residues more distant from the substrate P$_1$ aa also interact with the enzyme. For instance, the backbone of aa P$_3$-P$_1$ interacts with the backbone of enzyme residues 214-216, directing the substrate into the right position for cleavage. However, this interaction does not influence the substrate specificity since no aa side chains are involved (170).

**Figure 2.** Schematic illustration of substrate-binding pocket for chymotrypsin, trypsin and elastase. Residues with impact on substrate cleavage at positions 189, 216 and 226 are indicated. Black filled circles denote carbon atoms and white circles denote oxygen atoms.
Proteases of the chymotrypsin family can be further divided into three major subclasses based on the preferred P1 aa, i.e. the primary cleavage specificity; the chymotrypsin, trypsin and elastase subclasses. However, this family is highly diverse and proteases with metase (cleave after Met) and aspase (cleave after Asp) activity are also represented.

Depending on the size and shape of the substrate-binding pocket, (determined by residues 189, 216 and 226, see above), different aa side chains can be accommodated. The residues of the S1 pocket, positions 189, 216 and 226, are for the chymotrypsin subclass; Ser-Gly-Gly, the trypsin subclass Asp-Gly-Gly and for the elastase subclass Ser-Val-Thr. Chymotryptic enzymes hence hold small and uncharged aa in their S1 pocket, resulting in a moderately wide cavity allowing cleavage after aromatic aa like Phe, Tyr, Trp, and large hydrophobic residues Leu, Ile and Met. In contrast, trypsin-like proteases, preferably cleave after the basic aa, Lys and Arg, due to the negatively charged Asp₁₈₉ at the bottom of the S1 pocket. The S1 pocket of elastase-like proteases is lined with aliphatic residues generating a quite narrow pocket, admitting only small hydrophobic aa (Ala, Val). Thus, based on the residues of the specificity-conferring triplet one can hypothesize that a protease prefers to cleave after certain P1 aa residues, i.e. it is possible to speculate on an enzymes primary cleavage specificity.

A group of chymotrypsin-like serine proteases, the hematopoietic granule associated proteases (graspases), differ structurally and also functionally due to the lack of a disulphide bond (174). In most chymotrypsin-like proteases, a bridge is formed between Cys₁₉₁ and Cys₂₂₀ that cross-links the protein chain in the close vicinity of the S1 pocket. The lack of this bridge in graspases contributes to the formation of a pocket where P3 residues of the substrate can interact with the enzyme, generating an extended specificity (175). The graspases specificity extends even more, to the C-terminal side of the cleaved bond, this due to an extra loop inserted between the aa 39 and 40, interacting with the P1´ and P2´ aa of the substrate (176). In addition to the positions mentioned above, there are other sites of importance for the substrate/enzyme interaction, although they will not be discussed further. The graspases are encoded by the mast cell chymase locus, which will be discussed in detail in a later section.

Regulation of serine protease activity

Dysregulation of protease activity has been implicated in several diseases, the regulation of their activity is therefore of great importance. Proteases can be produced constitutively, upon stimulation or both. To avoid self-degradation, most serine proteases are synthesized as inactive pre-pro-enzymes, zymogens, with an N-terminal extension that is proteolytically cleaved to generate the active form. The pre-peptide constitutes a “tag”, a signal peptide, directing the protease to the secretory pathway. This tag is
subsequently removed in the endoplasmatic reticulum (ER) by the signal peptidase. Thereafter the pro-peptide, of varying length, has to be cleaved off to enable the proper folding of the protein to generate an active enzyme.

Most grapseases are activated by DPPI, an enzyme that is capable of removing N-terminal dipeptides (177). When the pro-peptide has been removed, the N-terminal Ile16 (occasionally Val) forms a salt bridge with Asp194 facilitating the formation of the oxyanion hole (described above). Interestingly, it was recently shown that DPPI-null mice only expressed inactive mast cell chymase and that the tryptase activity was significantly reduced (89). Furthermore, Gzm A and B were shown to be produced at normal levels in these mice, although the proteases were inactive (178).

A number of natural and synthetic inhibitors exist for proteases of the chymotrypsin family. Natural protease inhibitors can be found in almost all compartments of the body and bind covalently to their targets. Physiological factors like pH, salt concentration, temperature, etc. also regulates protease activity (179). To some extent the enzyme specificity also regulates its activity, depending on accessibility, localization and modification of substrate motifs.

The activity of proteases also can be modulated by the presence of proteoglycans. Proteoglycans are large proteins with glycosaminoglycan side chains. These macromolecules provide a storage scaffold for many proteases, affecting their activity, retention and polymerization (reviewed in (180)). In the acidic granules (pH 5.5) proteases are positively charged and will bind to the negatively charged proteoglycans. Depending on the aa (Lys, Arg or His) on the protease surface involved in the protease-proteoglycan interaction, dissociation after degranulation occurs at different rates.

**Hematopoietic serine proteases**

Hematopoietic serine proteases are encoded from four different loci, the mast cell chymase, the multigene tryptase, the Gzm A/K and the Gzm M/elastase loci. Such loci holding two or more relatively closely related genes most likely have evolved from one founder gene, which duplicated several times to give rise to a cluster of genes. The encoded proteases could subsequently develop various cleavage specificities. To date, these four loci, which hold closely related genes, have primarily been mapped in human, mouse and rat. In the following four sections a systematic review of these loci and their genes, in both human and mouse will be given. Proteases encoded from the mast cell chymase, the Gzm A/K and the Gzm M/elastase loci feature a common five exon-four intron gene structure (181). Considering the close phylogenetic relationship among the mast cell chymase, the GzmA/K and the GzmM/elastase loci encoded genes, it has been speculated whether these loci arose from two whole genome duplications. Interestingly only three loci has been identified in mouse, rat and human, indicating that a
fourth locus has been deleted or inactivated in mammals. In contrast to genes of the mast cell chymase, the Gzm A/K and the GzmM/elastase loci, several genes encoded by the multigene tryptase locus have six exons separated by five introns (87). Moreover, when exploring gene clusters, genes bordering the regions can be of great value. These flanking genes define the borders for each locus. Furthermore, the cell and tissue distribution and biological function for the respective protease will be discussed.

The mast cell chymase locus

In both human and mouse the mast cell chymase locus (also referred to as the granzyme B cluster) is situated on chromosome 14 (182-184). In mammals, genes encoded at this locus all lack the “active site” disulfide bond formed by Cys191 and Cys220 (discussed above). This feature appears to be unique for the proteases encoded from the mast cell chymase locus (granzymes) and strongly indicates a common evolutionary origin.

The human mast cell chymase locus contains four active protease genes, one α-chymase (CMA1) gene, one cathepsin G (CTSG) gene and two granzyme genes, GZMB and GZMH (Figure 3). The human α-chymase is expressed in MCTC. Cathepsin G is expressed in neutrophils and the two granzymes are expressed and stored in the granules of cytotoxic T lymphocytes (CTLs) and NK cells (185-187). Interestingly, a CTSG-like enzyme has also been detected in MCTC (188, 189), monocytes and neutrophils (190, 191). Due to a negative aa (Glu) in position 226, cathepsin G has both chymotrypsin- and trypsin-like activity (192). Cathepsin G has been suggested to play a role in several physiological and pathological mechanisms, including immune responses towards bacteria, inflammation, wound repair and regulation of blood pressure (reviewed in (193)). Unexpectedly, mice deficient for cathepsin G were shown to be strikingly unaffected by bacterial infections (194). The role of cathepsin G as an anticoagulant or as a procoagulant has been an issue for some time. Contradictory data show that cathepsin G can cleave and activate (195) or inactivate coagulation factor VIII (196, 197) contributing to coagulation/anticoagulation. Moreover cathepsin G activates coagulation factor V (190) and X (198). In addition to its proteolytic activity human cathepsin G possesses antibacterial properties possibly due to an internal antibacterial peptide (199). This is also true for the human granzyme B (199).
The name granzyme refers to the association of an enzyme to the granules. Human granzyme B has aspartic cleavage specificity, i.e. it cleaves particularly after Asp residues. This feature, shared by mouse granzyme B, is unique among eukaryotic serine proteases although it is seen for many caspases (200). The capacity to accommodate and cleave after P1 Asp side chains is assumed to depend on a basic Arg226 in the specificity-conferring triplet (201). Site-directed mutagenesis of this position dramatically alters the P1 preference for mouse granzyme B. The mutant Arg226Gly prefers hydrophobic aa in P1 (202) while the mutant Arg226Glu instead prefers to cleave after basic aa (203). Granzyme B activates caspases-induced cell death after cleavage of caspase-3 and additional caspases, but it can also induce apoptosis of target cells via caspase-independent routes as reviewed in (204). Cytotoxic cells from granzyme B-deficient mice are defective in their ability to induce apoptosis in target cells (205, 206). Interestingly, granzyme B knockout mice also lack the genes for granzyme C and D-G, which may also play an important role in apoptosis (207). Moreover, it was recently shown that Granzyme B together with perforin are important for a subset of regulatory T cells ability to suppress NK and T cell tumour clearance in vivo (208).

Granzyme H, in contrast, has chymotrypsin-like cleavage specificity and cleaves preferentially after Phe, Tyr and a few unbranched hydrophobic residues (209). Recently, it was shown that Gzm H could cleave and neutralize a virus derived Gzm B inhibitor, and thereby restore Gzm B mediated apop-
tosis of virus-infected cells (210). Gzm H is primarily expressed in the granules of NK cells (211).

The human and mouse mast cell chymase loci differ concerning both the size and number of genes encoded within the two loci. However, homologs for human CMA1, CTSG and GZMB are also present in the mouse. At one end of the locus, the mouse Cma1 gene, also termed mast cell protease 5 (Mcpt5), is situated, and next, four genes belonging to the β-chymase family, Mcpt1, -2, -4 and -9 are positioned. Until recently this group of proteases was suggested to be rodent-specific. However, in dog one gene termed CMA2, was shown to either cluster with the β-chymases or the α-chymases depending on if the phylogenetic analyses were based on exon nucleotide or amino acid sequences (41). Dog CMA2 was, however, shown to have a premature stop codon mutation in the open reading frame and no expression could be detected. Sequence comparisons of the β-chymases demonstrate an intimate evolutionary relationship among them (128, 212) (Figure 4). The β-chymases are expressed in both mast cell subtypes present in mouse, CTMCs and MMCs, summarized in (21, 41).

Next to the β-chymases resides another rodent-specific gene, Mcpt8. This gene is only expressed in basophils and serves as an important differentiation marker for these rare blood cells (39, 213). Moreover, in rat, as many as seven homologs to Mcpt8 has been identified (40, 134). Interestingly, in contrast to mMCP-8, three of rat homologous, rMCP-8, -9 and -10, have been shown to be expressed in MMCs (40). The specificity-conferring triplet of mMCP-8 resem-
bles the one seen for mouse granzyme B, therefore it has been suggested that mMCP-8 also has a preference for negatively charged P1 amino acids, i.e. has aspartase-like cleavage specificity (213). However, despite rigorous trials using phage-display and chromogenic substrates, Gallwitz et al. could not detect any proteolytic activity for mMCP-8 (41).

Neutrophil CTSG/Ctsg localizes to the center of both the mouse and human loci, and both loci are flanked by GZMB/Gzmb. In between these two genes seven granzyme genes, Gzmc-g, l and n are situated in the mouse. Analyses of the primary sequences reveals that at least six of these genes have a proper reading frame, Gzml however is predicted to be a pseudogene due to a premature termination site (214). This massive expansion of granzyme genes seems to be mouse-specific, the corresponding region in rat holds, besides Gzmb, only two additional functional genes; Gzmc and natural killer cell protease 7 (Nkpt7) (134). Many of these granzymes, including mouse Gzm C-G and N, and human Gzm H have been poorly characterized and are therefore referred to as “orphan” granzymes (215). Mouse Gzmb and c seem to be homologous to human Gzmb and h. Grossman et al. suggest that a duplication of an ancestral Gzmb gene occurred before the split of primates from rodents, and that separate gene duplication events subsequently took place in the rodent lineage (214).

Evidently, species-specific duplications of genes encoded from the mast cell chymase locus have occurred in the mouse. It is tempting to speculate that multiple granzymes present in mouse are of importance for their survival. Is it possible that similar events have taken place in other species, adapted to live in certain pathogen rich environments?

The multigene tryptase locus

Mast cell tryptases are encoded by the multigene tryptase locus. The human and mouse multigene tryptase loci map to syntenic regions on chromosome 16p13.3 and 17A3.3, respectively (91, 103, 113, 216). Both loci are composed of two gene clusters separated by more than 1 Mb (Figure 5A). Based on conserved residues in the primary sequence, the genes have been divided into “group 1 tryptases” and “group 2 tryptases” (91). The tryptases of group 1 all have an unpaired cysteine in their propeptide which is suggested to participate in the dimerization of two protease molecules. In the C-terminal part of the propeptide, group 1 tryptases possess an Arg or a Lys, while group 2 tryptases hold a conserved Gly. The size of the group 1 and 2 genes is approximately 5 and 2 kb, respectively. The nomenclature of the genes encoded from the multigene tryptase locus is somewhat confusing. For clarity, the most frequently used names will be mentioned in the following section. In this thesis focus will be on the group 2 tryptases and on one of the group 1 tryptases, the tryptase-γ, which is expressed in mast cells.

In human, genes belonging to the group 1 tryptases include TPSG1/γ-tryptase/transmembrane tryptase (TMT, γI and γII), protease serine
member S (PRSS) 27/marapsin/pancreasin, PRSS33/EOS, testis serine protease 1 (TESSP1), PRSS21/eosinophil serine protease-1 (ESP-1)/testisin, Tmprss8 (transmembrane protease serine 8, pseudogene) and PRSS22/tryptase ε/brain-specific serine protease 4 (BSSP-4). The group 1 tryptase genes, except TPSG1, cluster tightly together, whereas TPSG1 resides on the opposite outermost part of the locus next to the group 2 tryptase genes. The human group 2 tryptases, which are primarily expressed in mast cells, comprise the genes TPSB2 (encoding βII and βIII tryptases), TPSAB1 (encoding α and βI tryptases), TPSD1/tryptase δ (δI and δII) and the two pseudogenes PRSS29P/implantation serine protease 2 (ISP2) and Similar to (Sim.) Mastin.

As for the group 2 tryptases, the mouse locus resembles the human locus, encoding Mcpt6, Tpsab1/Mcpt7, Sim. Isp2/Spl1 (pseudogene), Prss29/Isp2, Prss28/Isp1 and Prss34/Mcp11. Eight group 1 tryptase genes are present in mouse, Tpsg1, Prss27, Prss22, Tmprss8/distal intestinal serine protease (mDISP), Prss21, Prss32, mTESSP-1 and Prss33.

All but two of the group 1 tryptases in mouse, Prss22 and Prss33, are attached to the cell membrane in different ways. Tryptase-γ has a C-terminal transmembrane domain with a short cytoplasmic tail whereas Prss27, Tmprss8, Prss21, Prss32 and mTessp1 have been shown to be glycosylphosphatidylinositol (GPI) anchored proteases (90, 113, 114, 217). Although tryptase-γ clusters with the mast cell tryptases, and certainly is their closest neighbour, phylogenetic analyses indicate a more distant relationship suggesting that the ancestor of the γ–tryptases predate the split between human and rodents (115) (Figure 5B).

**Figure 5.** The multigene tryptase locus. Panel A shows a map of the human multigene tryptase locus at chromosome 16. Group 1 and 2 tryptases (shaded in grey) are indicated, arrows depict transcriptional orientation and Ψ denotes pseudogenes. Panel B shows a dendrogram including human (Hs), mouse (Mm) and rat (Rn) mast cell tryptases. Amino acid sequences for mature proteases were aligned using Clustal X and the derived consensus tree (1000 replicates) using neighbor-joining is shown. Gene names are indicated in brackets.
The granzyme A/K locus

The granzyme A and K genes are organized in a locus which in human and mouse resides on chromosome 5q11-q12 and 13, respectively (218-221). These granzymes are associated with the cytotoxic granules of NK and T cells. In 1986 Pasternack et al. detected protease activity in the granules of T cells. The most abundant enzyme in part responsible for this activity was later denoted granzyme A. Gzm A was shown to be a disulphide-linked homodimer with trypsin-like activity that preferentially cleaves after P1 Arg or Lys (200, 222). Gzm A is a pro-apoptotic protease. However, Gzm A does not activate caspases (223) like Gzm B, but instead cleaves proteins associated with the nucleus of the target cell, i.e. lamins of the nuclei envelope, histone H1 and DNA (summarized in (224)). Gzm A-induced cell death is associated with the occurrence of large DNA fragments in contrast to the oligonucleosomal fragments seen after Gzm B induced caspase-dependent cell death. Gzm A has been shown to activate an ER-associated nuclease, SET, which induces single-stranded DNA nicks (225). Interestingly, Gzm A-deficient mice retain their potential to kill target cells in vitro (226, 227). On the other hand, the capacity to fight viral infections, i.e. ectromelia and herpes simplex, is reduced in these mice (228, 229).

Granzyme K, one of the orphan granzymes, possesses trypase activity (230). It has been suggested that Gzm A and K fulfil similar functions in the host defence against virus-infected and transformed cells. As described for Gzm A, Gzm K also induces caspase-independent cell death and the formation of reactive oxygen species (ROS) (231).

Perforin is a pore-forming protein found exclusively in the cytotoxic granules of NK and T cells (232, 233). Polymerized perforin facilitate the uptake of granzymes and promotes the release of granzymes from endosomal vesicles into the cytosol of target cells. Perforin-deficient mice have an impaired capacity to induce cell death in target cells and are consequently abnormally susceptible to a variety of viruses (111, 234-236). Perforin-deficiency seems to erase the capacity of a cytolytic cell to maintain the general antiviral defence, whereas deficiency in one granzyme seems to result in an increased susceptibility against one or a few viruses. This strongly implies that perforin has a central role in the process of granzyme delivery and granzyme-induced cell death.

The granzyme M or elastase locus

Granzyme M, azurocidin 1, proteinase 3, neutrophil elastase 2 and complement factor D/adipsin, cluster tightly on human chromosome 19p13.3 (237, 238). This gene cluster is termed the Gzm M/elastase locus. In mouse, homologs to the human genes, resides on chromosome 10q21.2. However, no azurocidin gene has been identified in the mouse.
Human and mouse granzyme M have met-ase activity cleaving after P1 Met, Leu and Nle (239, 240). Granzyme M is expressed almost exclusively by NK cells although low amounts have been found in subsets of T cells (241, 242). Interestingly, inactive alternative transcripts have been identified in cells of rodent hippocampus and mouse retina (243, 244). Granzyme M has been shown both to initiate caspase-independent (245) and caspase-dependent apoptosis (246). Entry of Granzyme M into cells seems to be perforin-dependent (245, 246).

Neutrophil elastase, azurocidin and proteinase 3 are stored in the azurophilic granules of neutrophils together with cathepsin G, which is encoded by the mast cell chymase locus. Neutrophils are phagocytic granulocytes rapidly recruited to the site of inflammation. Engulfed microorganisms end up in the phagolysosome where they are degraded by serine proteases among others. (Reviewed in (193)).

Azurocidin is a serine protease homologue, which has lost its proteolytic activity (247). A Ser57 and a Gly195 have replaced residues His57 and Ser195 of the catalytic triad. However, azurocidin have antimicrobial activity, primarily against Gram-positive bacteria (248, 249). Several candidate regions of the protease have been suggested to possess antibacterial properties (reviewed in (250)). However, the exact mechanism has not been determined. It has been proposed that azurocidin simply binds to bacterial cells and thereby acts as an opsonin facilitating phagocytosis by monocytes (251).

Neutrophil elastase and proteinase 3 are elastases with a similar mode of action; they cleave numerous matrix proteins like fibronectin, laminin, collagen type IV, elastin etc (summarized in (252)). Consequently, inhibition of these neutrophil-associated proteases results in less severe symptoms in models for degenerative and inflammatory diseases e.g. ischemia (253) and collagen induced arthritis (a mouse model that resembles human rheumatoid arthritis) (254). Proteinase 3 has been shown to cleave the human cathelicidin hCAP-18, and thereby generate the antibacterial peptide LL-37 (255). Moreover, mice deficient in neutrophil elastase were shown to have impaired capacity to clear Gram-negative bacterial infections (256). Mice lacking neutrophil elastase and/or cathepsin G were further shown to be more susceptible to fungal infections than wild-type mice (257).

A complement factor protein, adipsin or complement factor D, is also located in the Gzm M locus. Adipsin cleaves factor B in complex with complement component C3 (reviewed in (258)). This is one important step in the alternative pathway of the complement system.

An evolutionary perspective

Evolution is the process by which present species of bacteria, plants and animals arose. Earth is suggested to be more than 4.5 billion years old and
several studies imply that the diversification of animals began over one billion years ago. Evolution is not a straight line of improvements but an ornate route of changes and diversification of species. Mutations, deletions and insertions in the germline DNA are processes essential for evolution, but not always successful at the individual level. Important knowledge about life is today retrieved from studies of organism genomes. Since the first genome of *Haemophilus influenza* (259) was sequenced in 1995, over 180 additional genomes has been fully or partly annotated, including mouse (260) and human (261, 262). Today, many studies in the field of immunology are conducted in human and rodents, however comparative studies in more distant species are necessary. Knowledge about toll-like receptors of the innate immune system, gene conversion in birds generating antibody diversity and the discovery of phagocytes in echinoderms are just examples of studies that have contributed to the present understanding of the immune system. Further, the complexity of the immune systems is driven by challenges of varying kind, depending partly on an organism’s “lifestyle”, therefore it is of great interest to explore the genomes of organisms other than human and rodents. Genomic studies can help to uncover what makes a species unique, at which rate species evolve, when a gene arose or was deleted, etc. Moreover, studies of the appearance of a gene family during evolution can help to clarify their biological functions and importance. In conclusion, comparisons of genomes, and particularly genes of special interest, can improve the usefulness of animal models and consequently assist to improve human health.

We have in our studies of the different mast cell and basophil proteases used an evolutionary perspective to increase our understanding of their importance in mast cell/basophil biology and in immunity in large.
Present investigation

Aim
The general aim of this thesis was to investigate hematopoietic granule serine proteases expressed by mast cells and basophils, and the genes encoding these proteases. The focus has been on evolutionary aspects of these proteases to identify key functional components that have been maintained during evolution. Furthermore, as mast cell chymases and tryptases are encoded by different loci, it was of great interest to map these loci in varying species in order to trace their origin. It was also of particular interest to investigate the cleavage specificity of the opossum chymase, a protease closely related to the most ancient chymase identified. Finally, special interest was given to in vitro cultured cord blood-derived basophils as a source of mRNA, to further characterize their granule protein contents, particularly serine proteases.

Results and discussion

Emergence and expansion of the mast cell chymase locus during mammalian evolution (paper I)
In a recent study Gallwitz et al. showed that the mouse and rat mast cell chymase loci have expanded quite dramatically in regard to size and number of genes during recent mammalian evolution (134). Evidently, a species-specific expansion of granzyme genes in mouse and of the Mcpt2- and the Mcpt8-families in rat seems to have taken place. In addition and as discussed in paper II, no direct homologs to the mast cell chymase locus genes have been identified in teleosts.
Therefore, in this paper, our aim was to study the origin of the mast cell chymase locus genes in evolutionary diverse mammalian species. The genomes for cattle, opossum and the partial genome for platypus (*Ornithorhynchus anatinus*) were therefore analyzed for the presence of homologs to the different chymase locus genes in mice and human. All three species are mammals, cattle belong to the group of cetartiodactyls (ruminants), opossum belong to the marsupials, while platypus is an egg-laying mammal, a monotreme.
The overall organization for the cattle mast cell chymase locus resembles the one seen in human and rodents, i.e. it has an α-chymase gene on one flank, a granzyme B gene on the opposite flank and a cathepsin G gene in the centre of the locus. Interestingly, two α-chymase genes (CMA1a, CMA1b), linked in a tail-to-tail orientation, and two cathepsin G genes (GTSG1, CTSG2) linked in a head-to-head manner were identified. Moreover, at least four duodenase genes were identified, BDMD1, BDMD2, BDMD3 and BDMD4. However, these genes do not cluster together; BDMD1 and BDMD2 are situated between the CMA1b and the CTSG1 genes, and BDMD3 and BDMD4 are situated between granzyme H and granzyme B. An additional duodenase gene, BDMD5, was also identified, however the chromosomal localization could not be determined.

In contrast to the other mast cell chymase locus genes, which are involved in different immune functions, duodenases are food digestive enzymes. Since duodenases are not present in human and rodents, we wanted to assess whether duodenases could be found in other ruminants. Interestingly, two ESTs were identified in sheep; sMCP-1/BDMD1 and sMCP-3/BDMD2. These enzymes are closely phylogenetically related to the cattle duodenases. Hence, we propose that duodenases are ruminant-specific. In accordance to what was observed in cattle, two ESTs for sheep α-chymases, sMCP-2/Cma1a and sMCP-4/Cma1b, were also identified. Thus, double α-chymases seem to be ruminant-specific.

In opossum, only two mast cell chymase locus genes were identified. Phylogenetic analyses clearly showed that one of the genes was an α-chymase, whereas the other gene neither clustered with granzyme B/H, cathepsin G or the duodenases, but between these groups. Therefore this gene was termed grathepsodenase (granzyme, cathepsin G, duodenase). The two opossum genes did not cluster together but were separated by more than 44 Mb. However, the flanking genes were homologous to the genes found adjacent to the human and dog mast cell chymase loci. This indicated that a chromosomal rearrangement most probably occurred in marsupials, which lead to the separation of these two genes.

In a previous study performed in our laboratory, a platypus “granzyme” had been identified using degenerated primers directed against conserved motifs of the catalytic triad (263). At the time of this study, only a partial genome assembly for platypus was available. Despite rigorous BLAST (basic local alignment search tool) searches, no additional mast cell chymase locus genes could be identified.

Similar to previously identified mast cell chymase locus genes, all genes identified in this study lack the disulphide bridge formed by Cys191-Cys220. Moreover, the novel cattle, opossum and the sheep α-chymases hold a specificity-conferring triplet identical to the one found in the human and dog α-chymases. This indicates a conservative evolutionary pressure to maintain the specificity of α-chymases. As double α-chymases have been reported in
ruminants only, we suggest that a duplication of an ancestral α-chymase gene occurred when they shared a common ancestor. The sheep and cattle duodenases primarily have two different specificity-conferring triplets, N-G-D or D-G-N. It was recently shown that BDMD1 and sMCP-1/BDMD1, holding the N-G-D triplet, have dual specificity due to the negatively charged Asp226. This dual specificity is also seen for human cathepsin G holding Glu226 (192, 264, 265). In contrast, the two cattle cathepsin G genes possess a S-G-A triplet (seen in most α-chymases) that does not allow for dual specificity. The presence of double cathepsin G genes seems to be specific for the cetartiodactyls. The specificity-conferring triplet (A-G-R) of the opossum grathepsodenase is equal to the distinct triplets of mouse and rat granzyme B. Rodent granzyme B displays aspartase specificity, cleaving after acidic aa like Asp and Glu (202, 266). The preference for negatively charged aa is due to the basic Arg226 (201).

Taken together, we suggest that an ancestral mast cell chymase locus gene was present 215 Myr ago. A duplication of this gene took place before the marsupial/eutherian divergence, approximately 185 Myr ago. This is reflected by the presence of two mast cell chymase locus genes in the opossum; the α-chymase and the grathepsodenase. Finally, around 80-100 Myr ago, before the separation of the major groups of placental mammals, the ancestral gene related to the grathepsodenase, duplicated again and gave rise to the ancestral genes of granzyme B/H, cathepsin G, the Mcpt8-family (rodents) and the duodenases (ruminants).

Hematopoietic serine proteases lacking the common disulphide bridge (Cys191-Cys220) are absent in teleosts (paper II)

Here we studied the potential presence of mast cell chymase locus genes in different teleosts. Initially we screened a cod (Gadus morhua) head kidney cDNA library using a PCR-based strategy. Degenerated primers against conserved motifs surrounding the catalytic triad of serine proteases were utilized. These primers have successfully been used previously to isolate a number of novel hematopoietic serine proteases (40, 263, 267, 268). Using these primers, two granzyme-like cDNA clones were isolated; one homologous to the T cell granzymes A and K, termed cod Gzm A/K, and the other distantly related to the proteases encoded from the mast cell chymase locus, termed Gzm-like I. Bony fishes or teleost constitute a large and extremely diverse group of vertebrates. Therefore, we also screened two channel catfish (Ictalurus punctatus) cDNA libraries, made from either a mixed leukocyte culture, primarily containing NK-like cells, or a macrophage cell line. Three novel serine proteases were isolated. However, none of these were closely related to any of the well-established groups of hematopoietic serine proteases. They were termed catfish Gzm-like I, II and III. As more fish genomes became available we performed BLAST searches for candidate
genes in GenBank, Ensembl and the TIGR databases. In total, 40 novel sequences from seven different teleost species were identified. Again, none of the fish sequences grouped with the individual branches for serine proteases encoded by the mast cell chymase locus in mammals. However, several of these enzymes clustered with the branch leading to the individual chymase locus genes, thereby representing early ancestors of this family. In addition, several of the other newly identified proteases clustered with proteases encoded from the granzyme M locus, the granzyme A and K locus, or within groups including exclusively fish proteases.

All sequences identified showed the common five exon-four intron gene organization seen in hematopoietic serine proteases (212). Moreover, hematopoietic serine proteases encoded from the mast cell chymase locus all lack a disulphide bridge, formed by Cys191 and Cys220, which is highly conserved in most other serine proteases (174). This disulphide bridge span over the substrate-binding region and consequently proteases lacking this bridge possess a broader substrate-binding cleft. Interestingly, the Cys191-Cys220 bridge were present in all but two bony fish sequences, including the cod Gzm-like I and a zebrafish (Danio rerio) sequence. Cod Gzm-like I lack Cys191 and the zebrafish sequence lack both cysteines. Thus, it is appealing to speculate that these genes are related to a common ancestral gene lacking the Cys191-Cys220 bridge, that was present before the divergence of teleosts and tetrapods that gave rise to the mast cell chymase locus genes.

By analyzing residues of the substrate-binding region one can retrieve information about protease cleavage specificity. Three amino acids are of special interest, residues 189, 216 and 226 (chymotrypsin numbering), referred to as the specificity-conferring triplet. We aligned the sequences for bony fish proteases that assembled as separate groups. Two different groups of specificity-conferring triplets were observed, however none resemble triplets previously seen in mammalian hematopoietic serine proteases.

Taken together, these findings suggest that serine proteases of the mast cell chymase locus evolved after the separation of tetrapods from teleosts. However, serine proteases encoded from the granzyme A/K and granzyme M loci evolved as separate subfamilies prior to the divergence of teleosts and tetrapods approximately 420 million years ago.

The cleavage specificity for mast cell α-chymases has been conserved during mammalian evolution, from opossum to human (paper III)

In paper I we identified a number of novel genes encoded from the mast cell chymase locus in cattle and opossum. The opossum chymase is of special interest to us, being the first clear homolog to any of the mast cell chymase locus genes hitherto identified. Moreover, it is interestingly to notify that
both rodents and ruminants hold more than one chymase gene in contrast to human and opossum. In rodents, it has been shown that the different chymases also have distinct cleavage specificities.

In this study we wanted to determine both the primary, and the extended, cleavage specificity of the opossum chymase. In phylogenetic analyses this enzyme clearly clusters with the α-chymases. However, as stated above, the finding that rodent α-chymases (135, 136) possess elastase activity, it is important to characterize the opossum chymase extended cleavage specificity to truly understand the evolution and origin of mast cell chymases.

Thus, recombinant opossum chymase was produced in eukaryotic cells. The primary specificity was determined using a panel of chromogenic substrates. The opossum chymase was demonstrated to cleave two chymase substrates with aromatic residues in P1 (Phe, Tyr) and a Met-ase substrate with P1 Met. Cleavage after hydrophobic aa like Met and Leu, has been shown previously for a number of chymotrypsin-like proteases (123).

The extended substrate specificity of the opossum chymase was then determined by phage display methodology (Figure 6). Accordingly, positions surrounding the P1 aa residue, both at the C- and N-terminal side, can be analyzed. This technique has been used previously in our laboratory to determine the extended specificity for proteases encoded from the mast cell chymase locus (136, 140, 269, 270). T7 phages are modified to express a random nonamer and a His6-tag on their surface. Approximately $5 \times 10^7$ different recombinants are present in the library. Thus, the protease has a vast array of possible peptides to cleave. The phages are immobilized by binding of the His6-tag to a matrix of Ni2+-NTA. The protease of interest, in this study the opossum chymase, is then added and susceptible nonamers can be cleaved. The phages released are then amplified in Escherichia coli bacteria and the recovered sublibrary is exposed to the protease and an additional selection round is initiated. After five rounds of selection approximately 100 individual phage clones are isolated and the sequence encoding the nonapeptide (in each phage) is determined.
The retrieved aa sequences were subsequently analyzed. The total over- or under representation of any aa in the positions P4-P3’ were calculated. There was a high overrepresentation of the aromatic aa Trp and Phe and a moderate overrepresentation of Tyr. A weak overrepresentation of the negatively charged aa, Asp and Glu, was also observed. In contrast, the positively charged aa Lys and Arg were underrepresented. The retrieved sequences were aligned, and the preference for a specific aa in each position P4-P3’ was evaluated. A clear preference for the aromatic aa Trp (54%), Phe (27%) and Tyr (19%) were observed in the P1 position. The strong preference for Trp has not been shown previously for any other α-chymase. Moreover, a preference for negatively charged aa Asp and Glu was observed in the P2’ position. Acidic aa in the substrate P2’ position most probably interact with positively charged aa in position 143 and 192 of the enzyme (170, 272). The same pattern/preference has been seen for mMCP-4 and in natural substrates cleaved by the human chymase. In positions surrounding the scissile bond, P4, P3, P2 and P1’, opossum chymase favours aliphatic aa. This is in accordance with aa preferred in the corresponding positions by the human chymase and mMCP-4. Moreover, a preference for Trp and Leu are seen in P1’ position. Tandem aromatic aa at the cleaved bond, i.e. in the P1 and P1’ po-
sitions, have also been observed in sequences cleaved by rMCP-1 and mMCP-4 (140). The retrieved consensus sequence, of seven aa, can subsequently be used in order to identify possible in vivo substrates for opossum CMA1 and homologous enzymes. As discussed in the introduction, chymases are implicated in several pathophysiological conditions. By determining chymases extended cleavage specificities it would be possible to develop effective and enzyme-specific inhibitors.

The observed similarities in cleavage specificities between human CMA1, mMCP-4 and opossum CMA1, clearly indicate that these enzymes might have coevolved with an in vivo substrate conserved in certain positions involving cleavage. Moreover, the results presented in this paper indicate that there has been a strong selective pressure to maintain the specificity for mast cell α-chymases for more than 185 Myr of evolution.

Conservation of the multigene tryptase locus, including mast cell tryptases, over the past 185 million years of mammalian evolution (paper IV)

In papers I-III we have studied the evolution of the mast cell chymase locus. In paper IV we instead changed the focus to the locus encoding the mast cell tryptases; the multigene tryptase locus. This locus in human, mouse and rat can be divided into two separate clusters, one encoding group 1 tryptase genes and one encoding group 2 tryptase genes. The group 1 tryptases consist of membrane-bound trypsic proteases (most of them), whereas group 2 tryptases primarily include mast cell granule proteases. In more detail, the human group 1 tryptase region holds the following genes; PRSS31/TPSG1, PRSS27/pancreasin, PRSS33/EOS, TESSP1, PRSS21/testisin, Tmprss8 (pseudogene) and PRSS22/tryptase ε and the group 2 tryptase region encodes genes for; TPSB2, TPSAB1, TPSD1 and two pseudogenes; PRSS29P/ISP2 and Sim. mastin.

In this paper we mapped the multigene tryptase loci in dog, cattle and opossum. The overall arrangement for the three loci resembles the order of the previously characterized loci in human and mouse. The newly mapped loci are composed by two clusters separated by more than 1 Mb, a cluster encoding group 1 tryptase genes and a cluster holding group 2 tryptase genes. Moreover, as in human and mouse, homologous genes (CACNA1H/UBE2I) flank the region encoding group 2 tryptase genes in all three species.

The dog locus holds five presumably active group 1 tryptase genes and only two active genes in the group 2 tryptase region. Moreover, a tryptase-like gene, situated between dog TPSAB1 (53% aa identity) and dog mastin, was identified. The protease encoded by this gene is most probably inactive, as the catalytic triad residue Asn102 is replaced by an Asp102.
A chromosomal rearrangement has placed the cattle group 1 tryptase region on the other side of the group 2 tryptase genes, as compared to the human, mouse and dog loci. The cattle multigene tryptase locus holds five group 1 tryptase genes. Double PRSS21 genes are present, and one of these, Sim. PRSS21b, has a premature stop codon. In the group 2 tryptase region cattle hold only one active mast cell tryptase (273) gene and a highly similar pseudogene. Moreover, a mastin gene and a gene denoted, Sim. MCP-11 (mastin-like protease previously identified in mouse and rat) resides in this region. Cattle mastin shares approximately 60% aa sequence identity with cattle MCP-11 and dog mastin. A partial EST and genomic DNA have previously been identified for cattle mastin (274). Cattle MCP-11 is most homologous to pig mastin (TC30) (66.1% aa identity) and rat MCP-11 (65% aa identity).

The opossum multigene tryptase locus extends over a much larger region than in the other investigated species. The cluster holding group 1 tryptase genes is situated 5.9 Mb 5´ to the cluster of group 2 tryptase genes. The opossum group 1 tryptase region is inverted compared to the corresponding regions in human, mouse, dog and cattle. Interestingly, a Tmprss8 gene is present in the opossum locus; the gene probably encodes an active protease, which clusters with mouse Tmprss8 in phylogenetic analyses. Tmprss8 is a transmembrane protease expressed in mouse distal gut (275). A pseudogene for Tmprss8 is located in the human multigene tryptase locus (113). However in dog and cattle Tmprss8 genes are absent. Moreover, in the present opossum genome assembly (MonDom5), five genes can be mapped to the group 2 tryptase region. Four of these are mast cell tryptase-like genes (SP13, SP1, SP2, Sim SP2), of which two most likely are inactive (SP2 and SP13). Two additional tryptase-like genes (SP3 and SP4) were identified in a prior assembly of the opossum genome, but these two are absent in the current assembly. Despite this, they are included in the analyses. cDNA for opossum, SP1 and notably opossum SP3 and SP4 were isolated. Interestingly these three resemble both human α- and β-tryptases and mouse mMCP-6 and -7 in different aspects. Two of them, SP1 and SP3 hold a negatively charged Asp189 at the bottom of the specificity-conferring triplet, allowing for cleavage after positively charged P1 aa residues (Figure 2). However, opossum SP4 holds a Gly189. These three tryptase-like opossum proteases also possess all the typical mast cell tryptase features, e.g. the majority of the conserved tryptophans (human and mouse holds nine), eight conserved cysteines etc. The opossum SP1 cDNA sequence was isolated from skin and spleen, SP3 from skin, spleen and intestine, and SP4 from skin and intestine. This indicates that SP1 is restricted to CTMCs, whereas the other two, SP3 and SP4, are potentially expressed in both CTMCs and MMCs. Additionally, in the opossum group 2 tryptase region a gene denoted Sim. ISP2 (SP8) was identified. Functional implantation serine proteases (ISP) have so far only been discovered in mouse. However, recently an intact ISP2
sequence was identified in a rhesus macaque (121). Opossum Sim. ISP2 seems to encode an active protease.

As the organization of the multigene tryptase locus appeared to be preserved over time, we extended the screening into monotremes, amphibians and teleosts. In the partial platypus (Ornithorhynchus anatinus) genome, five group 1 tryptase genes and two group 2 tryptase genes were identified. The two platypus proteases related to the group 2 tryptases, Sim. Tryptase β and Sim. Tryptase, fall on the same branch as opossum Sim. ISP2.

The screening of nonmammalian vertebrates, zebrafish (Danio rerio) frog (Xenopus laevis and Xenopus tropicalis) resulted in the identification of tryptases-like proteases related to the group 1 tryptase genes. However, none of these group with mammalian multigene tryptase loci proteases, but instead on separate species-specific branches.

In conclusion, mammalian genes encoded from the group 1 tryptase region cluster in protease-specific subgroups deeply rooted in the phylogenetic tree. In contrast, the genes encoded from the group 2 tryptase gene region, the only ones likely to be expressed in mast cells, cluster in species-specific groups. These genes seem to have diversified within species. The group 2 tryptases Mastin/mMCP-11 and ISP on the other hand, cluster in protease-specific groups.

In conclusion, the overall organization of the multigene tryptase locus seems to have been conserved throughout mammalian evolution. However, until data can be retrieved about the organization of a possible platypus multigene tryptase locus it is difficult to speculate about how a more ancient multigene tryptase locus was organized and when it appeared.

Purification and characterization of immature umbilical cord blood-derived basophils (paper V)

Basophils are of major importance in acute allergic inflammatory reactions by their release of immediate mediators such as histamine, LTC₄ and cytokines, mainly IL-4 and IL-13. They originate from a CD34+ progenitor cell and terminally differentiate in the bone marrow before they enter the blood circulation. Basophils constitute approximately 1% of the blood leukocytes and can be recruited to tissues upon proper stimuli. However, the investigation of their true biological role has been long hampered by the lack of basophil-specific markers e.g. surface antigens and granule components. Furthermore, the low amount of basophils in the blood and the lack of mRNA in mature basophils make it even more complicated. To overcome these problems, immature, transcriptionally active basophils are required.

To address this problem, umbilical cord blood cells were differentiated into basophils in the presence of IL-3. A previous study by Kepley et al.
showed that pulsing the cells with IL-3 generated high numbers of mature basophils. However we wanted transcriptionally active immature cells (276). In our hands the continuous presence of IL-3 in the cultures gave a much better result, higher amount and more immature transcriptionally active cells.

By flow cytometry analyses, these precursor basophils were shown to co-express FceRI and the novel mast cell/basophil marker 97A6 (CD203c) (34). Interestingly a small fraction of FceRI and CD203c+ cells were observed by day 10, possibly representing basophils with an even more immature phenotype.

To confirm the presence of immature basophils in the cultures, cells from day 9 and 12 were purified by negative selection using magnetically labeled antibodies targeting non-basophil surface antigens (MACS Basophil Isolation kit, Miltenyi Biotec). The purity of the isolated cells was 97-98%, as determined by Alcian blue staining. However, the yield was only 14%, thus we assumed that some of the surface markers targeted by the Basophil Isolation kit were present on immature basophils. Flow cytometry of these cells revealed that a mature population of basophils was isolated by this purification procedure. Thus, we conclude that different surface markers are present on immature compared to mature basophils. To isolate more immature basophils we modified the cocktail of antibodies used in the negative selection process, and we analyzed the surface marker expression on the purified cell population.

We could show that immature cultured basophils express the human leukocyte antigen (HLA) -DR, which was down regulated on more mature cultured basophils. It has previously been shown that HLA-DR is virtually absent on circulating peripheral blood basophils (277). At day 10 of culture, 59% of the cells were HLA-DR+ CD203c+ compared to day 20, when less than 25% were HLA-DR+ CD203c+. Furthermore, the following markers included in the Basophil Isolation kit, CD3, CD14, CD15, CD16, CD19 and CD21 were absent on these cells.

On the basis of these findings two different cocktails of antibodies, a one-step cocktail, targeting CD3, CD14, CD15, CD16 and CD19, and a two-step cocktail targeting CD3, CD16, CD36, CD45RA, CD14 and CD15 were developed and used to purify immature in vitro cultured basophils. Neither of the cocktails included antibodies towards HLA-DR. The highest purity (95%) and recovery (59%) of cells was achieved using the two-step cocktail.

Finally, 34x10^6 immature umbilical cord blood-derived basophils were isolated using the two-step cocktail. These cells were used to construct a cDNA library for large-scale sequencing. The library contained approximately 600 000 recombinants. Significant levels of the FceRI α-chain and two common basophil/eosinophil proteins, CLC protein (2%) and MBP (0.6%) were isolated.

To summarize, the surface antigen expression of IL-3 derived cord blood basophils was determined in order to establish a purification protocol for
these cells. Additionally, we show that the retrieved basophil precursors are transcriptionally active, and thus can be used as a source for further biochemical analyses to improve the understanding of these rare blood cells.

Concluding remarks

Granule stored serine proteases are major effector molecules of mast cells. Accumulating data regarding human and rodents serine proteases has resulted in a deeper understanding of their roles in a vast array of biological processes. However, to gain a broader insight into mast cell serine proteases and their biological function it is of great importance to expand the studies into more distantly related species, that is, to look at it from an evolutionary perspective.

In this thesis detailed maps for the cattle and opossum mast cell chymase loci, and the dog, cattle and opossum multigene tryptase loci, are presented. The studies of the mast cell chymase locus revealed a dramatic expansion of this locus over time, considering both locus size and number of genes encoded. We identified a gene in opossum as being the first clearly “committed” mast cell chymase locus gene, an α-chymase. We showed that the protease encoded by this gene, opossum CMA1, exhibits cleavage specificity similar the human α-chymase. Remarkably, this specificity has been conserved during more than 185 million years of mammalian evolution! Moreover, 45 novel teleost serine proteases related to mammalian mast cell chymase locus genes are presented. Interestingly, these teleost proteases all hold a disulphide bridge, which is consistently absent in mammalian proteases encoded from this locus. In phylogenetic analyses none of the teleost proteases branch with mammalian mast cell chymase locus genes. However they cluster with the branch leading to the different subfamilies of mast cell chymase locus genes. Moreover, several of the teleost enzymes cluster with serine proteases encoded from the granzyme A/K and the granzyme M/elastase loci. In contrast to the mast cell chymase locus, the multigene tryptase locus including mast cell tryptase genes feature a conserved arrangement and hold similar numbers of genes in opossum, cattle and dog as compared to human and mouse. Moreover, in the monotreme platypus, tryptase genes most probably encoded by the same locus were identified. With a good picture of how the loci including mast cell chymases and tryptases are arranged in mammals, we turned the focus to basophils. The granule proteins stored in basophils have been poorly characterized, however in mouse basophils a serine protease, mMCP-8, has been identified. To be able to clone human basophil-specific granule proteins, umbilical cord blood-derived basophils were cultured and harvested, when still immature and transcriptionally active. The surface antigens of these cells were analyzed and based on these data we could set up a protocol to purify immature basophils. A high-
quality cDNA library from 34x10^6 purified immature purified cord blood-derived basophils was constructed. Initial analyses identified typical basophil transcript. This study represents the first step towards the characterization and understanding of *in vitro* cultured umbilical cord blood-derived basophils proteome. Hopefully this can help us to gain more knowledge on the granule content and cell surface markers of this illusive cell, that recently have been shown to have such a profound effect on early events in an inflammatory response (70, 278).
Sammanfattning på svenska


I min forskning är två typer av immunceller centrala, mastceller och basofil granulocyer (basofiler). Dessa två cellytter associeras starkt med allergiska och inflammatoriska sjukdomar, men de är också viktiga för en organismens försvar mot parasiter och bakterier. Jag har framför allt studerat en grupp av enzymer, serinproteaser, som lagras inuti mastcells och basofilers granula. Allergier och inflammatoriska sjukdomar ökar ständig, framför allt i västvärlden. Det är därför av stor vikt att öka kunskapen om bakomliggande mekanismer och om vilka celler, molekyler etc. som är involverade t.ex. i allergiers uppkomst och förlopp.

Mastcellers effekter medieras till stor del av ovan nämnda serinproteaser och då framförallt av kymaser och tryptaser. Dessa två enzym utgör den största mängden av de proteiner som finns i mastcells granula. Kymaser och tryptaser är därför användbara markörer för mastceller, då man vill detektera mastceller i ett prov taget från t.ex. en sjuk person. Vad gäller basofiler så vet man fortfarande inte vilka proteiner som lagras i deras granula. Det innebär att det inte finns några bra verktyg för att detektera och mäta aktivering av basofiler.

Kymaser och tryptaser är s.k. endopeptidaser, enzymer som kan klyva peptidbindningar inuti i ett protein. Det är ännu ej helt kartlagt vilka naturliga substrat som dessa enzym föredrar att klyva. Enzymer kan delas in i olika familjer utifrån vilka aminosyror de föredrar att klyva efter. Kymaserna tillhör kymotryptinfamiljen och klyver främst efter stora aromatiska aminosyror, medan tryptaserna tillhör trypsinfamiljen och föredrar att klyva efter positivt laddade aminosyror. I många djurarter finns det flera olika gener för kymaser och tryptaser. Gener som ger upphov till liknande proteiner hährör...


Som tidigare nämnts, hittade vi en kymasgen i opossum. Denna gen var en första tydliga kymasgen och fylogenetiska analyser visar att den tillhör en subfamilj av kymaser som kallas för α-kymaser. Vi ville förstå hur viktiga dessa α-kymaser är och även hur bevarat deras sätt att klyva på är (artikel III). Vi producerade därför opossum α-kymaset och gjorde analyser på dess klyvningsspecificitet. Vi undersökte också vilka aminosyrer, som fanns på båda sidor om den aminosyra som enzymet klöv efter, för att på
detta vis undersöka enzymets utvidgade klyvningsspecificitet. Utifrån de data vi fick fram så tycks det som om opossums α-kymas klyver på ett sätt som liknar α-kymaset i människa. Det verkar alltså som om denna specificitet har bevarats under marsupial/placental däggdjurs-evolution, vilket tyder på att denna funktion är central i pungdjur och placentala däggdjur.

Vi har också undersökt multigentryptaslokusetet i hund (Canis familiaris), ko och opossum (artikel IV). Tvärtom vad vi såg för mastcellskymaslokkusetet, så verkar detta lokus vara relativt lika i pungdjur och placentala däggdjur. Detta tyder på att även mastceller hos pungdjur kan uttrycka och utsöndra tryptaser.

Som nämnts ovan, så finns det inga bra markörer för att detektera basofiler. Tidigare studier har visat att denna celltyp till största delen stänger av sin proteinproduktion innan de lämnar benmärgen där de mognar. Det vill säga, mogna basofiler som återfinns cirkulerande i blodet innehåller inget mRNA, som vanligen kan användas för att indirekt undersöka t.ex. vilka proteiner som en cell lagrat i sina granula.

Acknowledgements


I would like to express my sincere gratitude to the people who have encouraged me and contributed to the work presented in this thesis, especially I would like to thank:

**Lasse Hellman**, my supervisor, for the inspiration, cooperation and guidance in science. I value your great knowledge in immunology, paleontology, how to reconstruct a house, gardening and so on and so forth. It has been great fun to work with you!

Co-supervisor **Gunnar Nilsson**, many thanks for your constructive input into the almost “never-ending” basophil project. It was really great that you believed in it!

Co-supervisor **Sara Wernersson**, for taking me under your wings when I first came to Lasses lab. I sometimes miss the late nights in the lab…I think we were a fantastic team.

Past and present members of the “HELLMAN GROUP”; **Maike**, a tall American (!?), who showed to be a sweet, special, little German girl! Thank You for all the moments we shared, in the “psychology corner”, in the gym, in a jeep driving through Israel’s desert or at a mountain in Rio de Janeiro. I hope we will get to see a lot of each other in the future. Thank you for sharing with me your scientific and personal competence. **Mattias A**, for making me realize that the pace of life can vary. You have been a fantastic support in this “process”, Good Luck! **Sayran**, I hope things will go your way.

**Maria Ave**, for being reliable, warm and kind and for your “Puss-Puss”, **Parvin**, for your black humor and for your joyful attitude! The girls from the old days, **Anna L, Camilla, Ulrika, Lotta, Molly, Jeannette**, for giving me a smooth start in the world of science and for showing me that there is something out there…Thank You! All project students that contributed greatly to science and atmosphere. Especially; **Tina**, for our talks and for being so cheerful, **Yue**, for your energy and for being so jolly, **Mattias E**, for being such a nice guy, for your fantastic input into the opossum project, and for your patience with our endless “girl talks”, **Per H**, for being friendly and amusing,

Past and present members of the “KLEINAU GROUP”; **Sandra** for being such a bundle of energy, it is nice to see the fire that you have within you, **Sofia**, my soulmate in
so many ways. You have really been a shoulder to lean on. Let’s start that “immuno-
company” that we’ve been talking about! **Kajsa N**, for being such a positive person,
showing me that if you want something - you’ll get it – I will try to mimic you!
**Cecilia** for your special attitude, I like your personality, **Maria Andrén**, for your
pleasant and open manner. All nice project students that popped by in your lab;
**Sinnisky, Erik, Mia, Anders** etc.

**Lars Pilström, Niklas and Siv** for being so honest and friendly “neighbors”. **Siv** for
our “morning talks”. **Eva-Stina** for contributing to wild days and crazy nights.

All friendly colleagues at ICM who make the days enjoyable! **Per N** for invaluable
support during the last painful weeks, **Linda M** for being so considerate! **Nicole** for
our chats, **Erika E**, for sparkling discussions during lunches, you are special. All
girls being part of the “ONSDAXÖL” and Immunologists included in “PAFIBB”, it’s
so pleasant and convenient to be among like-minded souls! Other people at BMC,
thank you all!

**THANK YOU!** All wonderful friends that I have missed so much…”Old” friends from
Vänersborg, **Anna B, Cissi, Lina, Frida G, Kajsa**, Thank You, for making the time
outside the lab so enjoyable, “New” and “Old” friends that I earned during my Upp-
sala-time, **Malin** (my first!), **Hanna, Turid, Frida H-W** (Jag hoppas att våra
lördagstraditioner håller i sig!), **Dragos, Maria S, Charlotte, Helene, Tone** and of
course your kids and better half’s as well. You all make the world a better place!

Min Uppsalafamilj; **Nadja, Patrik och Daniel**. Tack för allt fantastiskt stöd i form
av barnpassning, “sambopassning” å goda middagar! Det är härligt att ni finns så
nära.

Tack **mormor å morfar**. Ni är bäst! Tack mormor för ditt uppriktiga intresse för det
som jag pysslat med under dessa år.

Bror **Jonas**, du har alltid varit min förebild…jag läste till och med tyska! Du är den
bästa bror som någon kan ha! **Kajsa**, tänk att du blev min svägerska! Tack för att ni
gett mig världens finaste syskonbarn; **Emil och Gabriel**…nu har jag tid för bus
igen.

**Mamma å Pappa**, trots avstånd är ni mig närmre än någonsin. Jag beundrar er op-
timism i svåra stunder! Vad vore jag utan ert oändliga stöd?!**HENRIK**, min älskling, min prins! Din kylskåpsoesi gör mig stark, min favorit är;
“**kär le k ömhet och beundra n**”. Tack för ditt obegränsade tålamod, tack för att du
föstår, du är fantastisk!

**ALVA** mitt sanna solsken å min åska. Du få mig att le varje dag!

---

**Jenny 🌸**
References

phosphatase SHIP by phosphorylated Fc gammaRIIB during negative regulation of IgE-dependent mouse mast cell activation. *Immunol Lett* 54:83.


61


Baker, E., T. J. Sayers, G. R. Sutherland, and M. J. Smyth. 1994. 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


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 427

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)