Developmental and Comparative Immunology 127 (2022) 104273

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Analysis of the mast cell expressed carboxypeptidase A3 and its structural and evolutionary relationship to other vertebrate carboxypeptidases

Srinivas Akula [a,b,*], Lars Hellman [b], Francesc Xavier Avilés [c], Sara Wernersson [a]

[a] Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, BOX 7011, SE-75007, Uppsala, Sweden
[b] Department of Cell and Molecular Biology, Uppsala University, Uppsala, The Biomedical Center, Box 596, SE-751 24, Uppsala, Sweden
[c] Institut de Biotecnologia i de Biomedicina (IBB) and Departament de Bioquímica i de Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

ARTICLE INFO

Keywords:
Mast cells
Carboxypeptidase
Granule proteases
Evolution

ABSTRACT

Metallo-carboxypeptidases are exopeptidases with diverse expression and function, found in all kingdoms of life from bacteria to mammals. One of them, the carboxypeptidase A3 (CPA3), has become an important component of the mammalian immune system by its expression in mast cells. Mast cells (MCs) are highly specialized sentinel cells, which store large amounts of bioactive mediators, including CPA3, in very abundant cytoplasmic granules. Clinical studies have found an increased CPA3 expression in asthma but the physiological role as well as the evolutionary origin of CPA3 remains largely unexplored. CPA3 belongs to the M14A subfamily of metallo-carboxypeptidases, which among others also includes the digestive enzymes CPA1, CPA2, CPB1 and CPO. To study the appearance of CPA3 during vertebrate evolution, we here performed bioinformatic analyses of homologous genes and gene loci from a broad panel of metazoan animals from invertebrates to mammals. The phylogenetic analysis indicated that CPA3 appeared at the base of tetrapod evolution in a branch closer to CPB1 than to other CPAs. Indeed, CPA3 and CPB1 are also located in the same locus, on chromosome 3 in humans. The presence of CPA3 only in tetrapods and not in fishes, suggested that CPA3 could have appeared by a gene duplication from CPB1 during early tetrapod evolution. However, the apparent loss of CPA3 in several tetrapod lineages, e.g. in birds and monotremes, indicates a complex evolution of the CPA3 gene. Interestingly, in the lack of CPA3 in fishes, zebrafish MCs express instead CPA5 for which the most closely related human carboxypeptidase is CPA1, which has a similar cleavage specificity as CPA3. Collectively, these findings clarify and add to our understanding of the evolution of hematopoietic proteases expressed by mast cells.

1. Introduction

Mast cells (MCs) are hematopoietic cells primarily located at the interphase between body and environment where they act as sentinel cells of the innate immune system (Abraham and St John, 2010; Galli et al., 2008; Galli and Tsai, 2010). The majority of the MCs originate from the fetal yolk sac (Li et al., 2018; Gentek et al., 2018). However, in adults, during inflammatory conditions MC precursors can migrate from the bone marrow to various tissues and differentiate into mature MCs (Metcalfe et al., 1997; Hallgren et al., 2007). MCs are densely packed with cytoplasmic granules, which upon activation can fuse with the plasma membrane and release a number of potent physiologically acting substances including bioactive amines, proteoglycans and proteases (Hellman and Thorpe, 2014; Pejler et al., 2007; Wernersson and Pejler, 2014). These proteases are the major protein component of the granule and can account for up to 35% of the total protein of the cell (Schwartz et al., 1987). The well-known serine proteases, tryptase and chymase, are the dominating granule-stored proteases in human MCs. Human MCs also express the serine protease cathepsin G, which is primarily found in neutrophils. In addition to these three serine proteases, which all are endopeptidases, MCs can also express high amounts of a metallo-carboxypeptidase, the carboxypeptidase A3 (CPA3), which is an exopeptidase. Together with chymase and cathepsin G, CPA3 is primarily found in one of the major MC subpopulations, the connective tissue MCs (Irani et al., 1991; Reynolds et al., 1989; Lutzelschwab et al., 1997). However, expression of CPA3 mRNA in mucosal type MCs of

* Corresponding author. Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, BOX 7011, SE-75007 Uppsala, Sweden.
E-mail address: Srinivas.Akula@slu.se (S. Akula).

https://doi.org/10.1016/j.dci.2021.104273
Received 24 August 2021; Received in revised form 28 September 2021; Accepted 28 September 2021
Available online 4 October 2021

human lungs have also been reported (Siddhuraj et al., 2021). Notably, findings from several clinical studies have indicated a link between asthma and lung expression of CPA3 which has increased the interest in this less well-characterized MC protease (Dougherty et al., 2010; Balzar et al., 2011; Fricker et al., 2019; Pejler, 2019). The mentioned serine proteases of MCs (chymase, tryptase and cathepsin G), belong to the large family of trypsin-related serine proteases. Members of this protease family are involved in a wide variety of physiological processes including, food digestion, blood clotting, angiotensin conversion and immune regulation (Caughey, 2011; Pejler et al., 2010). The role and evolution of these hematopoietic serine proteases have been studied quite extensively (Hellman and Thorpe, 2014; Huang and Hellman, 1994; Gallwitz and Hellman, 2006; Gallwitz et al., 2006; Akula et al., 2015; Hellman et al., 2017; Reimer et al., 2010). However, the evolution of the carboxypeptidase (CP) that is specifically expressed by hematopoietic cells, the MC CPA3, has not been studied in detail. To close this gap in our understanding of the evolution of hematopoietic granule proteases, we here explored with particular focus, the appearance of CPA3 during vertebrate evolution.

The MC expressed CPA3 is an exopeptidase that cleaves at the C-terminal end of proteins or peptides by releasing single amino acids (Natsuaki et al., 1992). CPA3 belongs to the M14 zinc-binding metallo-carboxypeptidases (Skidgel and Erdos, 1998; Rawlings, 2020). Based on structural similarity within the M14 family, four subfamilies have been identified, 1) the A/B subfamily (M14A), 2) the N/E subfamily (M14B), 3) the bacterial peptidoglycan hydrolyzing enzymes subfamily (M14C) and 4) the complex cytosolic carboxypeptidases CCPs/Nna1-like subfamily (M14D) (MEROPS (Kalinina et al., 2007),). CPA3 belongs to the M14A subfamily which also includes the digestive enzymes CPA1, CPA2, CPB1 and CPO, the regulatory plasma enzyme CPB2 (also called thrombin-activatable fibrinolysis inhibitor, TAFI, or CPU), and the less studied CPs with diverse expression and function CPA4, CPA5, and CPA6. The close M14B subfamily, including CPD, CPE, CPM, CPN and CPZ, has been found to exhibit regulatory functions e.g. processing of neuroendocrine peptides (Petrera et al., 2014; Sapio and Fricker, 2014; Fernandez et al., 2010; Garcia-Pardo et al., 2020). The cleavage preference of the different M14 CPs may differ quite substantially. Their primary specificity, i.e., their preference for the residue in the P1' site, are either aliphatic, hydrophobic residues, basic residues or acidic residues (Tanco et al., 2013, 2017). CPs are classified based on their preference for C-terminal aromatic/hydrophobic residues (A-type/CPA-like), basic residues (B-type/CPB-like) or acidic residues (O-type/CPO-like). The M14 CPs are present in all kingdoms of life including eubacteria, archaea, fungi, plants and metazoan animals including mammals and also in viruses, with distinct representations. CPs are, like the serine proteases, involved in a number of physiological processes, like food digestion, cleavage of hormones and neuropeptides, tubulin processing and immunity. Some studies have also shown the involvement of CPs in disease, including epilepsy, febrile seizures, enhanced fibrinolysis, reduced wound healing and increased inflammatory response to lipopolysaccharide (Sapio and Fricker, 2014; Quesada et al., 2009). Moreover, we found that inhibition of CP activity linked to members of the M14A subfamily could reduce symptoms in a murine asthma model, suggesting a pathogenic role for these CPs in asthma (Waern et al., 2021).

In this paper, we mainly focus on the evolution of the MC-specific CPA3, which belongs to the M14A subfamily of CPs (Reynolds et al., 1989). CPA3 performs important roles in degrading the endothelin-1, in the protection against injected venom, cleavage of a few neuropeptides and potentially also in regulating immunity (Piliponsky et al., 2008; Schneider et al., 2007). The cleavage specificity of CPA3 has been studied extensively, however, the full physiological role of CPA3 is far from fully elucidated (Tanco et al., 2013; Pejler et al., 2009). To dig deeper into such roles, and particularly on the potentially conserved functions of CPA3 in immunity and tissue homeostasis we here used bioinformatics and comparative genomics to trace the origin of the mast cell CPA3 during vertebrate evolution.

2. Results

2.1. Five gene loci harbor the nine M14A carboxypeptidases found in vertebrates

Human and mouse M14A carboxypeptidase sequences were used as query sequences to identify similar sequences in a large panel of vertebrate genomes in the NCBI database using the TBLASTN algorithm. The ensemble database was also later screened for related sequences to obtain the best coverage of the various genomes included in this study. Five different gene loci were identified, the CPB1-CPA3, the CPO, the CPA6, CPA1-A2-A4-A5, and the CPB2 locus. The size of the genes and the distances between genes were calculated and used to produce maps drawn to scale of these chromosomal regions as presented in Figs. 1–3. The large differences in gene sizes that can be found and that originate from differences in intron sizes is exemplified in Fig. 4. Related CP sequences from a few selected non-vertebrate species were also used for the phylogenetic analysis to give a better picture of the evolution of the large family of M14 CPs.

2.2. The CPB1-CPA3 locus

The CPA3 gene is located on human chromosome 3 along with pancreatic CPB1, a locus here referred to as the CPB1-CPA3 locus. At one end this locus is bordered by AGRT1, the gene for the angiotensin II receptor type 1, which is the best characterized Ang II receptor (Fig. 1A). At the other end the locus is bordered by GYG1 and HLTF (Fig. 1A). The GYG1 gene is encoding glycogenin-1, an enzyme involved in glycogen synthesis, and the HLTF gene is encoding Helicase-like transcription factor, a double-stranded DNA translocase. In all studied mammals, we found a similar pattern except for the platypus, which apparently has lost the CPA3 gene. Reptiles and amphibians seem in general to have a similar pattern as in mammals with both CPB1 and CPA3 (Fig. 1A). However, in chicken, we only found CPB1 and no CPA3 indicating that CPA3 secondarily have been lost in both monotremes and birds. The NCBI database was also screened for the potential presence of CPA3 in other bird species (52 species) and in the echidna, another monotreme, but no signs of CPA3 was found in these species. In general, mammals have only one copy of CPB1 and one of CPA3. However, in both the green anole (a reptile) and the Western clawed frog (an amphibian) we observed a duplication of the CPA3 gene (Fig. 1A). When studying the corresponding locus in fishes, we only found the CPB1 but not the CPA3 gene in the species analyzed, gar and zebrafish. Similarly, CPB1 but not CPA3 was found in the elephant shark and the lamprey, a jawless fish (Fig. 1A). Six other shark species showed a similar situation as for the elephant shark (data not shown).

2.3. The CPO locus

The CPO locus has a very similar overall organization in almost all animal species analyzed with a single CPO gene and with the bordering genes MDH1B at one end and KLF7 at the other end. However, we could observe a few differences to this pattern in a few species. A loss of the CPO gene was found in both mice and platypus while three additional gene copies of CPO was found in Western Clawed frogs. Interestingly, these three new CPO copies in the Western Clawed frogs were all positioned outside one of the bordering genes, the MDH1B (Fig. 1B). CPO could not be detected in any of the shark genomes analyzed or in the lamprey (Fig. 1B).

2.4. The CPA1, -A2, -A4, -A5 locus

All of the other CPA genes, except CPA3 and CPA6, are in the human genome located in one locus on chromosome 7. This locus harbors the

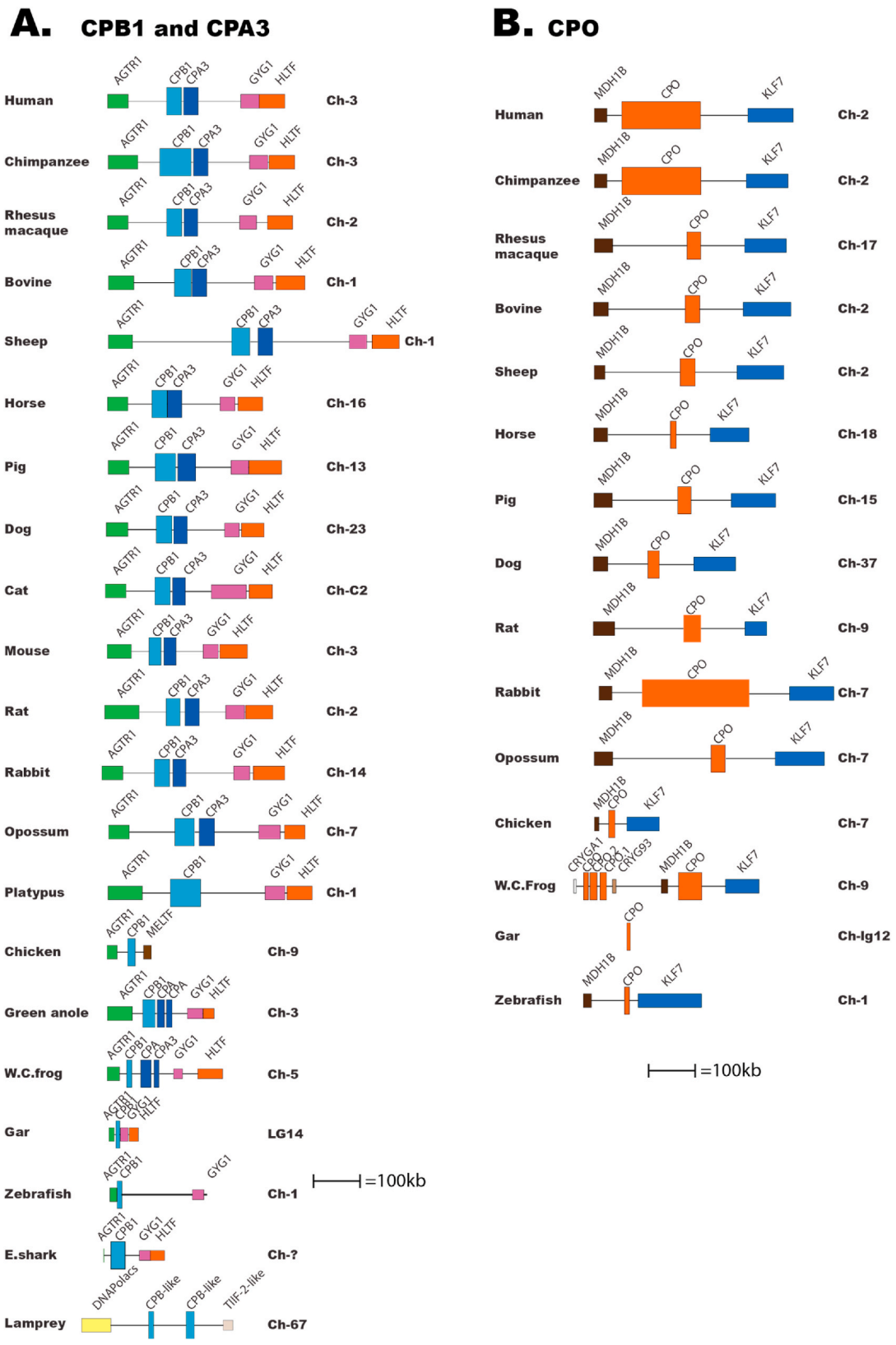

Fig. 1. The two loci for CPB1 and CPA3 and for CPO. The genes encoding carboxypeptidases are depicted in double-height to easily locate them in the maps. The bordering genes are included to trace the origin of the locus and to also define changes that have occurred upstream or downstream of the locus. Gene names for the carboxypeptidases, as given in the database. Panel **A;** shows the CPB1-CPA3 locus. The CPB1 genes are depicted in light blue and the CPA3 gene is a slightly darker blue. The bordering genes are shown in different colors to easily distinguish them from each other and from the CPs. Panel **B** shows the CPO locus. The CPO gene is depicted in Orange.

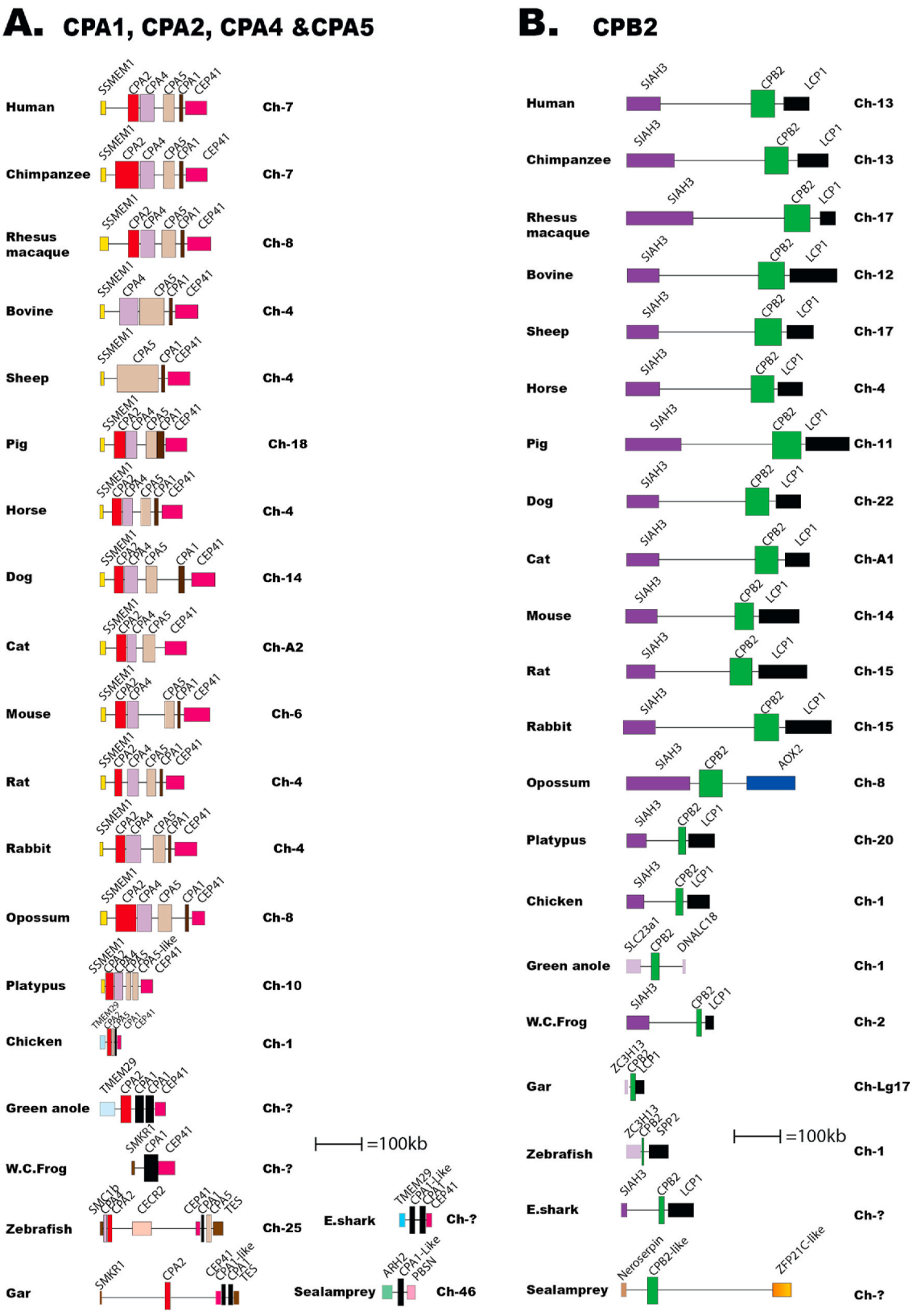

Fig. 2. The two loci for CPA1, CPA2, CPA4 and CPA5 and for CPB2. The genes encoding carboxypeptidases are depicted in double-height to easily locate them in the maps. The bordering genes are included to trace the origin of the locus and to also define changes that have occurred upstream or downstream of the locus. Gene names for the carboxypeptidases, as given in the database. Panel **A** shows the locus encoding the CPA cluster with the genes CPA1, CPA2, CPA4 and CPA5. The CPA2 genes are depicted in red, the CPA4 in purple, the CPA5 in beige and the CPA1 in black. The bordering genes are depicted in different colors to easily distinguish them from each other and from the CPs. Panel **B;** shows the CPB2 locus. The CPB2 gene is depicted in light green.

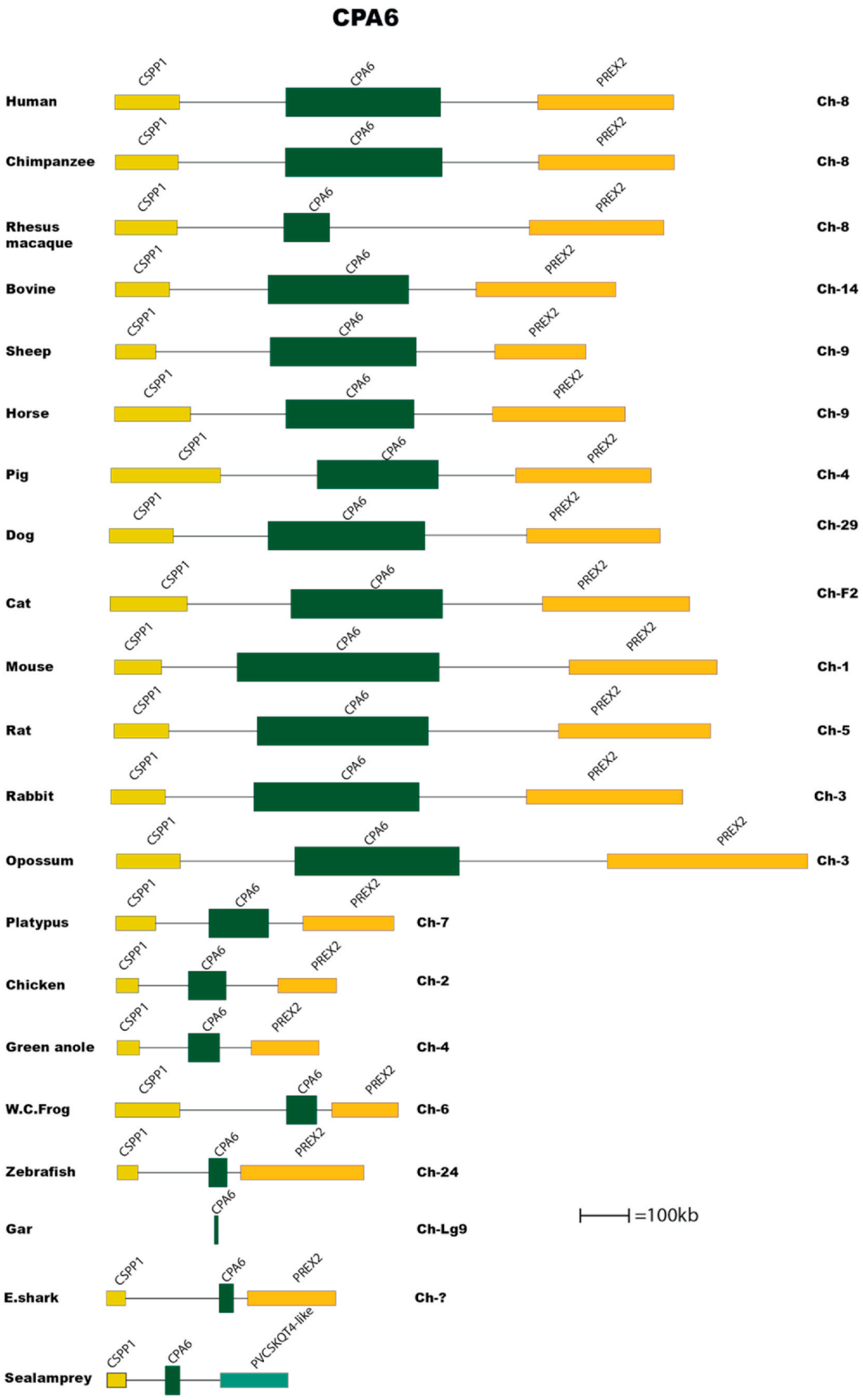

Fig. 3. The CPA6 gene locus. The genes encoding CPA6 are depicted in double-height to easily locate them in the maps. The bordering genes are included to trace the origin of the locus and to also define changes that have occurred upstream or downstream of the locus. Gene names for the carboxypeptidases, as given in the database. The CPA6 gene is depicted in dark green. The bordering genes deposited in different colors to easily distinguish.

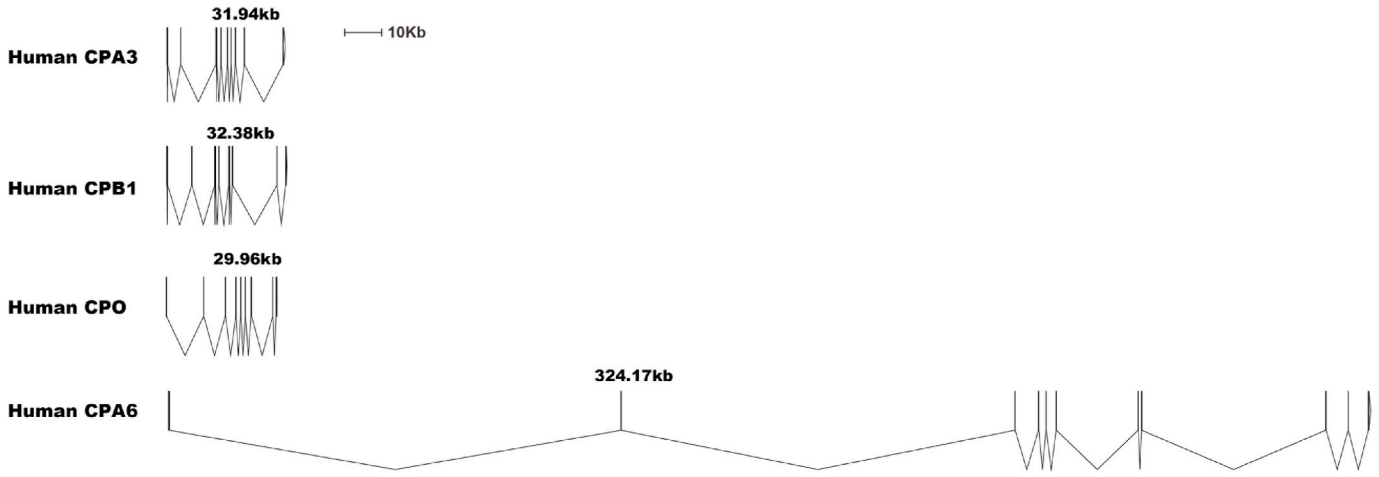

Fig. 4. Exon intron structure of human CPA3, CPO, CPB1 and CPA6. This figure is included to show the marked difference in size of these genes and that the size difference depends on difference in size of several of the introns and not on exon size differences. The exons are shown as vertical lines.

genes for CPA1, CPA2, CPA4 and CPA5, and is bordered on one end by SSMEM1, a serine rich membrane protein of unknown function, and at the other end by CEP41, a centrosomal and microtubule binding protein. The topology of this locus is very similar in all the studied mammalian species, except a loss of CPA2 gene in cow and sheep and a loss of CPA1 gene in the platypus. In the platypus we also find a duplication of CPA5. In the chicken and the green anole the locus organization is also very similar, except for a loss of CPA4 gene in the chicken and a loss of both A4 and A5 and a duplication of the CPA1 gene in the green anole (Fig. 2A). In both the chicken and the green anole one of the bordering genes have changed, i.e., the SSMEM1 gene is replaced by TMEM29, a protein of unknown function (Fig. 2A). In amphibians as represented by the Western clawed frog we only find CPA1 and also a change at one end of the neighboring genes indicating a major loss of genes in this locus (Fig. 2A). Interestingly, here the SSMEM1 gene has been exchanged for Smkr1, a small lysine rich protein expressed almost exclusively in human testis (Fig. 2A). The same bordering gene, the Smkr1, is found in the gar (Fig. 1B). In the zebrafish this bordering gene has been changed to Smc1b, a chromosomal maintenance protein involved in meiosis and mitosis (Fig. 2A). In both gar and zebrafish all the CPA genes of the human locus are present and one of the neighboring genes, CEP41, is the same as in the mammalian locus (Fig. 2A). However, the position of the individual CP genes differs from the mammalian locus (Fig. 2A). At one end of the locus we find CPA4 and CPA2 and on the other side of the common bordering genes with mammals, the CEP41, we find CPA1 and CPA5 in the zebrafish. In the gar the names of the genes differ somewhat and the CPA4 gene is lost (Fig. 2A). However, it is sometimes difficult to define the exact closest homologue why one should be careful with comparing genes only based on their given names, as they sometimes not directly match the closest neighbor in a phylogenetic tree. Interestingly, in the elephant shark there are two CPA1 like genes and the bordering genes are the same as for the green anole and the chicken (TMEM29 and CEP41), whereas in the lamprey there is only one CPA gene in this locus and both bordering genes are different from all other vertebrates analyzed (Fig. 2A).

2.5. The CPB2 locus

In most mammals, the *CPB2* locus consists of a single *CPB2* gene with the neighboring genes SIAH3 at one end and LCP1 at the other end (Fig. 2B). SIAH3 is a gene encoding an E3 ubiquitin protein ligase family member and the LCP1 gene encodes Plastin-2, an actin binding protein. The few differences observed to this pattern are in the bordering genes and no additional copies and no loss of the *CPB2* gene has been observed in any of the species analyzed. In the gar and the opossum the neighboring genes at one end of the locus has changed and in the green anole and zebrafish the neighboring genes at both ends of the locus have changed indicating active rearrangements in the region of the *CPB2* gene. In the opossum, the neighboring gene LCP1 at one end has been replaced with AOX2, the aldehyde oxidase 2, and in the gar, the neighboring gene SIAH3 has been replaced by the zinc finger gene ZC3H13. In the green anole, both bordering genes have been changed, at one end the SIAH3 has been replaced by Slc23a1 and at the other end LCP1 has been replaced by dnalc18 (Fig. 2B). The Slc23a1 gene encode one of two sodium dependent vitamin transporters and the dnalc18 gene encode the dynein light chain 1, an axonemal protein. In the zebrafish, both bordering genes have been replaced, at one end with the same gene as gar, the zinc finger protein ZC3H13, and at the other end by Spp2, the secreted phosphoprotein 2. The elephant shark has also only one CPB2 gene and the bordering genes are the same as for the amphibians and all mammals (Fig. 2B). In the lamprey both bordering genes are different from all other species analyzed, with neuroserpin on one end and ZEP21C-like on the other end (Fig. 2B).

2.6. The CPA6 locus

A single CPA6 gene is present in the human genome with the neighboring gene CSPP1 at one end and with PREX2 at the other end of the locus (Fig. 3). The CSPP1 gene encodes a methionine rich storage protein and PREX2 the phosphatidyl −3,4,5-triphosphate-dependent Rac exchange factor 2. This locus is highly conserved and has an identical organization from cartilaginous fishes to mammals. However, size differences can be observed, which originate from differences in the size of introns and inter-gene regions (Figs. 3 and 4). It is primarily several of the introns that have increased in size in human CPA6 when comparing human CPA3, CPO and CPB1 with human CPA6 (Fig. 4). In the lamprey one of the bordering genes has been changed, the PREX2 has been replaced by the gene PVCSQT4-like (Fig. 3).

2.7. Phylogenetic analysis

Phylogenetic analyses were performed with the amino acid sequences of 163 different M14A carboxypeptidases ranging from lace coral, purple sea urchin, oyster, lamprey, sharks and amphibians, to humans. The more distantly related M14B CPs, CPD and CPM, were used as an outgroup to obtain a more robust tree. Only the amino acid sequences of the active CPs were used in this analysis to obtain the most reliable estimate of sequence relatedness. All the protein sequences were obtained from NCBI database and the fasta formatted sequences were run using MAFFT. The aligned sequences were then used in different

programs to run the phylogenetic analysis. All of the trees gave a very similar pattern why we here show one of these trees, the MrBayes tree.

The CPD and CPM, belonging to the M14B clade of CPs together with the well known CPE, CPN and CPZ, formed a reference and clear outgroup in these trees (Fig. 5) (Garcia-Pardo et al., 2020). All other CPs, belonging to the M14A clade formed a big branch with several subfamilies (Fig. 5). Interestingly the CPO ends up in a small branch in between the CPA3 and CPB1 indicating that they are very closely related (Fig. 5). The CPA1, CPA2, CPA4 and CPA5 family members formed a distinct subfamily and the CPA3, CPO and CPB1 formed another distinct subfamily (Fig. 5). The CPA6 formed a clearly distinct subfamily between these two larger subfamilies and the CPB2 formed another small branch outside of all of these other subfamilies, thereby being the most distantly related members of this large family of M14A type CPs (Fig. 5). From the phylogenetic analysis, both CPA3 and CPO were found to be very closely related to CPB1, which indicates that both CPA3 and CPO may have appeared by duplications of the CPB1 gene. However, the position of CPO is then slightly problematic as it appears in between CPA3 and CPB1 in the phylogenetic tree but is located in a completely different chromosomal region. Interestingly also was the presence of a lamprey CPB, marked by a red arrow in Fig. 5, that seems to be positioned at the base of the CPB1, CPA3 and CPO branch of the tree indicating that it is representing an early ancestral form of this branch leading to the separation into the now three members of this branch. Fig. 5 is also available in an enlarged format in Supplementary Figs. S1 and S2.

3. Discussion

The phylogenetic analysis of the different mammalian carboxypeptidases clearly showed that the mast cell expressed CPA3 is more closely related to CPB1 (and CPO) than to other A-type CPs (CPAs), including the M14A carboxypeptidases CPA1, CPA2, CPA4, CPA5, and CPA6. We found a close colocalization of CPA3 and CPB1 in the same locus in all tetrapods and with the same bordering genes as in the basal ray finned fish species, the gar, which only express CPB1 (Fig. 1A). These findings strongly indicate that CPA3 is a result of a relatively recent, in evolutionary terms, gene duplication of CPB1. However, as indicated above, the evolutionary relationship between CPB1, CPA3 and CPO is then somewhat elusive (Fig. 5). CPO ended up in between CPA3 and CPB1 in the phylogenetic tree, demonstrating that CPO is very closely related in primary sequence to both CPA3 and CPB1 (Fig. 5). CPA3 was found in all mammals analyzed except for the platypus and was also found in both reptiles and amphibians, but was absent in all fish species analyzed. Interestingly, both green anole and the Western clawed frog have two copies of CPA3. In contrast, birds, as represented by the chicken, have lost the CPA3 gene, which is similar to what we found in the platypus (Fig. 1A). The analysis of the gene loci encoding CPA3 and CPB1 in Fig. 1A thereby indicates that CPA3 did appear at the base of the tetrapods possibly around 400 million years ago.

The mystery is then the appearance of CPO. Has an extra copy been moved to another chromosomal region and in that case by what mechanism? CPO is present in both gar and zebrafish indicating that CPO did appear much earlier than CPA3, which makes the tracing of their evolutionary history even more problematic. These results indicate that an early duplication of CPO or CPB1 did result in the transfer of the new copy to a new chromosomal location and that a second duplication during early tetrapod evolution resulted in CPA3. However, the high sequence similarity between CPO and CPB1, and that all three are almost equally related then contradict such a scenario. A second alternative, which is even more unlikely, is that CPA3 was lost in all fish species similar to what we found in the platypus and in chicken. We are therefore left with a mystery of how CPO, or possibly CPB1, has moved to another chromosomal location. However, such a scenario is not impossible as other similar examples can be found, e.g. in our recent study on granzyme A/K genes in cichlids (Akula et al., 2015). We found that one of the multiple copies of granzymes A and K like genes present in cichlids had moved to a completely different location in the middle of another locus encoding hematopoietic serine proteases, and there is surrounded by no other closely related granzyme A/K related members (Akula et al., 2015). Similar to the situation for CPO the mechanism may also here be difficult to trace.

CPA3 is in mammals tightly associated with MCs and basophils. Interestingly, MCs in the zebrafish have also been found to express a carboxypeptidase although CPA3 does not exist in fish (Fig. 1A). After closer analysis of the fish MC carboxypeptidase it was found to be CPA5, a relatively distantly related M14A CP encoded from another locus (Fig. 2A) (Dobson et al., 2008). The zebrafish CPA5 shows the highest homology with human CPA1 (64%) and only 38% identity with human CPA3 (Dobson et al., 2008). The question therefore arises how the CPA5 lost its MC expression in tetrapods and instead CPA3 became the dominant CP in tetrapod MCs. This is one of the intriguing questions concerning the emergence of CPs as an important part of the MC protease repertoire from early vertebrates to humans. Interesting also is the potential similarities and differences in cleavage specificity of fish CPA5 and tetrapod CPA3 and thereby their potential biological targets.

A comparison of the cleavage specificity and tissue expression of CPA3 with that of the other members of the M14A subfamily could reveal important evolutionary aspects. Relatively detailed studies have been performed on the majority of the nine human M14A CPs (CPA1-6, CPB1-2 and CPO). Human MC CPA3 prefers the aromatic amino acids Phe and Tyr, but cleaves also after Leu but not after Trp and Ala (Natsuaki et al., 1992). A peptide library based study of mouse CPA3 has shown that this enzyme prefers Leu, Phe, Tyr and Trp at P1' position of a substrate, which is the most important specificity determining site for the carboxypeptidases, and Leu/Ile at the P1 site (Tanco et al., 2010). As described in the introduction, CPA3 has a protective role by inactivating harmful substances like toxins and endotoxins (Piliponsky et al., 2008; Schneider et al., 2007). The closely related CPB1 is one of the pancreatic CPs released into the intestinal lumen taking part in food digestion. In marked contrast to CPA3, CPB1 does not cleave after aromatic amino acids but instead after the basic amino acids Arg and Lys (Avilés et al., 2013). Both CPA1 and CPA2 are pancreatic CPs with preference for bulky aromatic amino acids (Barber and Fisher, 1972; Bukrinsky et al., 1998; Gardell et al., 1988). Unlike the other digestive CPs, CPO is a membrane bound brush border enzyme expressed by epithelial cells and is specific for acidic amino acids such as Asp and Glu (Garcia-Guerrero et al., 2018; Lyons and Fricker, 2011). The plasma carboxypeptidase CPB2, also known as TAFI, function similar to the pancreatic CPB1, which prefers lysine and arginine at the C-terminal region. However, CPB2 circulates in plasma and regulates fibrinolysis by removing the fibrin C-terminal residue (Claesen et al., 2021). The remaining M14A members, the CPA4, 5 and 6, are all non-pancreatic A-type CPs with diverse expression and relatively unknown physiological functions. CPA4 is located in nucleoplasm and cytosol and is expressed e.g. in esophagus and skin (human protein atlas) (Fricker and Gomis-Rüth, 2013). Interestingly, human CPA5, which is highly expressed in testis, primary in tubules, has a very similar specificity to CPA3 with preference for aromatic amino acids and Leu (Fricker and Rawlings NDS, 2013). CPA6 is expressed in several tissues, including olfactory bulb and other nerve tissues, and is present in the extracellular matrix, presumably involved in neuropeptide cleavage (Claesen et al., 2021; Fricker et al., 2013; Mao et al., 1999; Heylen et al., 2011).

When looking at the very diverse functions of the M14A CPs that are closely related to the MC expressed CPA3, we can conclude that the origin of the MC expressed CPs seems as a relatively complex evolutionary journey. The two closest related CPs with respect to primary sequence are CPB1 and CPO, which both are digestive enzymes. CPB1 is secreted from the pancreas together with CPA1 and CPA2 whereas CPO is a membrane bound brush border enzyme. Both of them also have completely different primary specificities compared to CPA3. CPA3 seems to appear at the base of tetrapods around 400 million years ago

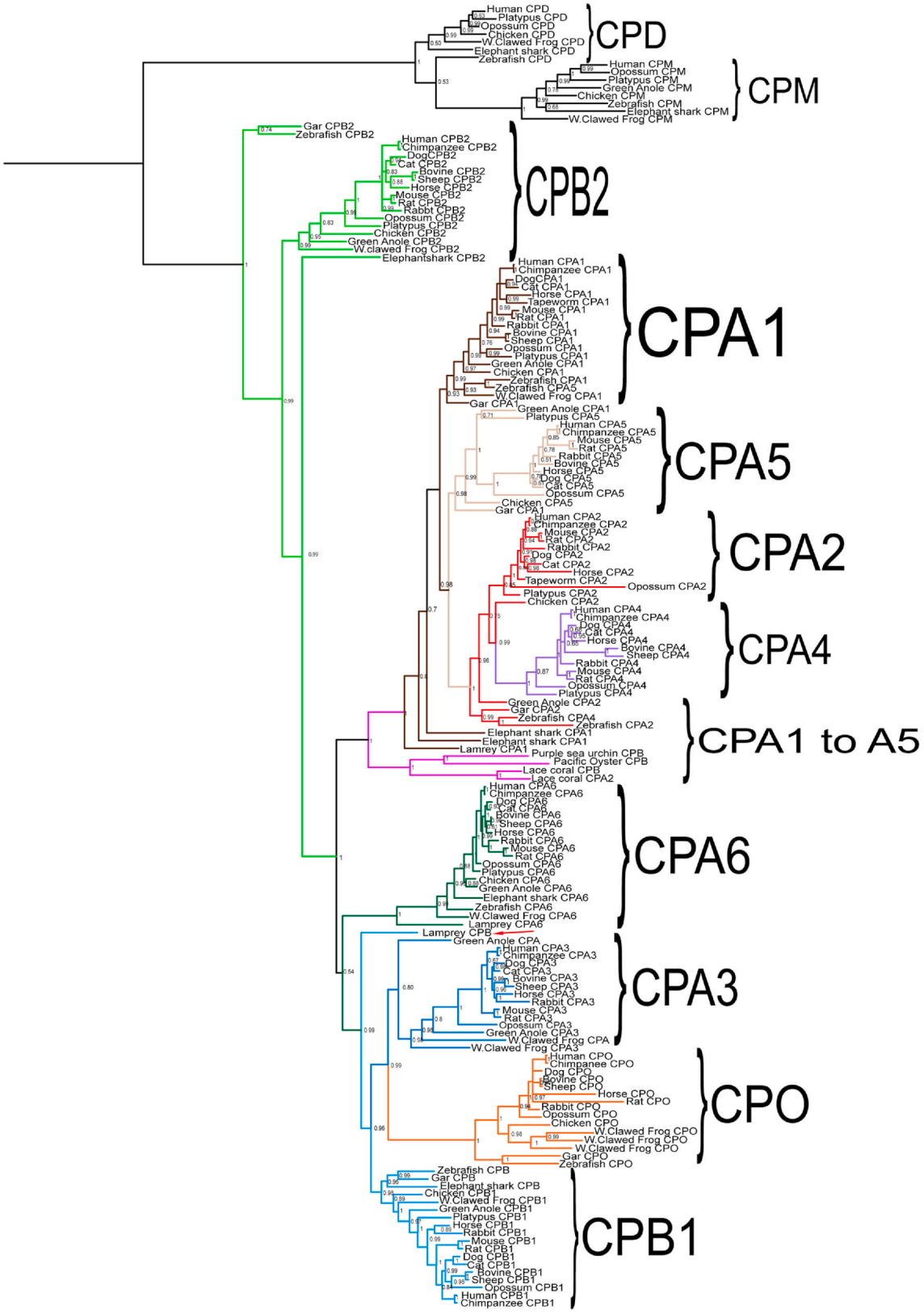

Fig. 5. A phylogenetic tree of the protein sequences of M14A carboxypeptidases CPA1, CPA2, CPA3, CPA4, CPA5, CPA6, CPO, CPD and CPM. The tree was constructed using MrBayes with the MCMC method and the sequences were obtained from NCBI tblastn by homology search. Phylogenetic tree of carboxypeptidase relationships based on Bayesian inference with the MrBayes program of CPs genes. Numbers on nodes indicate posterior probabilities. The branches representing the different CPs are depicted in the same colors as in the loci Figs. 1–3 and with CPA1 in brown, CPA2 in red, CPA3 in dark blue, CPA4 in purple, CPA5 in beige, CPA6 dark green, CPO in orange, CPB1 in light blue and CPB2 in light green. Noteworthy is the inclusion of the related M14B CPs, CPD and CPM in the analysis, belonging to the related M14B subfamily, appearing in black in the upper branch of the tree, as an outgroup to obtain a more robust tree.

whereas the closest homologue, CPO, is present in early ray finned fishes with an origin 50–100 million years earlier. Interestingly, lamprey has a gene that was positioned at the base of the CPB1/CPA3/CPO branch of the evolutionary tree (Fig. 5). This indicates that an early ancestor of this branch was present before the appearance of the jawed vertebrates and that the three closely related members did appear by gene duplications during early stages of the evolution of jawed vertebrates possibly at the base of ray finned fishes. The later appearance of the CPA3 at the base of tetrapods is then a major problem as indicated earlier. However, we see no easy explanation to this evolutionary riddle at the moment. To this mystery can then be added that zebrafish lack CPA3 but zebrafish MCs do instead express fish CPA5. To our knowledge no detailed analysis of the cleavage specificity of zebrafish CPA5 has been performed, but it may have similar specificity as human CPA3 as the closest human homologue CPA1 has a similar specificity as human CPA3. The fact that fish MCs also express a carboxypeptidase, similar to mammalian MCs, indicates that carboxypeptidases have an important function in MC biology. This also indicates a parallel convergent evolution in fish and tetrapods where two initially independent genes have by convergent evolution obtained similar biological functions. The origin of CPA3 and CPO may remain a riddle. However, further analysis of the cleavage specificity of zebrafish CPA5 and/or ample interactomics screening will hopefully in a near future shed light on the conserved key targets for these enzymes and shed new light into their role in MC biology.

4. Materials and methods

4.1. Gene loci

We used the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and the Ensemble (http://www.ensembl.or/index.html) databases for identifying the gene loci and the particular genes for the different M14 carboxypeptidases. Human and mouse M14 carboxypeptidases protein sequences were initially used as query sequences in Translate Basic Local Alignment Search Tool (TBLASTN) in the nucleotide collection (nr/nt) database. The identified genes in fish, frogs, reptiles and birds were then used to screen for potential additional homologous genes to complete the picture. The gene loci were compiled from the resulting data, and the sizes of the genes, including the distance between the neighboring genes, were calculated and used to construct scale maps of the loci. The figures were created in Adobe Illustrator (CC24.3). The Gene bank accession numbers used in constructing the gene figures and the phylogenetic tree are listed in the supporting information (S1).

4.2. Alignment and phylogenetic analyses

Alignments: The sequences of 163 selected M14 carboxypeptidases were aligned in version 7 of MAFFT (https://mafft.cbrc.jp/alignment/server//), using the BLOSUM62 as the scoring matrix and as an option G-INS_I strategy for optimal results for sequences with global alignment, with default parameters (Katoh et al., 2019). To check the alignment conservation and confidence the GUIDANCE2 server (http://guidance.tau.ac.il/ver2//) was used (Sela et al., 2015). To verify the multiple sequence alignment from MAFFT, another alignment algorithm (T-coffee) was used to verify that both alignments were similar.

Phylogenetic analyses: For all 163 proteases, the entire sequence of the active form, not including the signal sequence and activation peptide, was used in the multiple alignments. The phylogenetic analyses were performed using a Bayesian approach as implemented in MrBayes version 3.1.2. Markov Chain Monte Carlo (MCMC) analyses were used to approximate the posterior probabilities of the trees. Analyses were run using the MrBayes manual standard protocol (Ronquist and Huelsenbeck, 2003). The phylogenetic trees were drawn in Figtree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The topologies of all Bayesian phylogenetic trees were supported by posterior probabilities (PP).

Author contributions

Conceptualization, S.A., L.H. and S.W.; validation, S.A. and L.H.; formal analysis, S.A., and L.H.; investigation, S.A., and L.H.; data curation, S.A., and L.H.; writing—original draft preparation, S.A. and L.H.; writing—review and editing, S.W., and FX. A; visualization, S.A., and L. H.; supervision, L.H.; project administration, L.H.; funding acquisition, L.H., S.A. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by grants from Knut and Alice Wallenberg foundation KAW2017.0022. (LH and SW), The Konsul Th C Bergh Foundation (SA), and The Swedish Heart-Lung Foundation (SW).

Funding sources

The Knut and Alice Wallenberg Foundation (LH and SW), The Konsul Th C Bergh Foundation (SA), and The Swedish Heart-Lung Foundation (SW).

Declaration of competing interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dci.2021.104273.

References

Abraham, S.N., St John, A.L., 2010. Mast cell-orchestrated immunity to pathogens. Nat. Rev. Immunol. 10 (6), 440–452.
Akula, S., Thorpe, M., Boinapally, V., Hellman, L., 2015. Granule associated serine proteases of hematopoietic cells - an analysis of their appearance and diversification during vertebrate evolution. PLoS One 10 (11), e0143091.
Avilés, F.X., Vendrell, J., 2013. Carboxypeptidase B. In: Rawlings NDS, G. (Ed.), Handbook of Proteolytic Enzymes. Academic Press, pp. 1324–1329.
Balzar, S., Fajt, M.L., Comhair, S.A., Erzurum, S.C., Bleecker, E., Busse, W.W., et al., 2011. Mast cell phenotype, location, and activation in severe asthma. Data from the Severe Asthma Research Program. Am. J. Respir. Crit. Care Med. 183 (3), 299–309.
Barber, A.K., Fisher, J.R., 1972. A mechanism of action for carboxypeptidase A. Proc. Natl. Acad. Sci. U. S. A. 69 (10), 2970–2974.
Bukrinsky, J.T., Bjerrum, M.J., Kadziola, A., 1998. Native carboxypeptidase A in a new crystal environment reveals a different conformation of the important tyrosine 248. Biochemistry 37 (47), 16555–16564.
Caughey, G.H., 2011. Mast cell proteases as protective and inflammatory mediators. Adv. Exp. Med. Biol. 716, 212–234.
Claesen, K., Mertens, J.C., Leenaerts, D., Hendriks, D., 2021. Carboxypeptidase U (CPU, TAFIa, CPB2) in thromboembolic disease: what do we know three decades after its discovery? Int. J. Mol. Sci. 22 (2).
Dobson, J.T., Seibert, J., Teh, E.M., Da'as, S., Fraser, R.B., Paw, B.H., et al., 2008. Carboxypeptidase A5 identifies a novel mast cell lineage in the zebrafish providing new insight into mast cell fate determination. Blood 112 (7), 2969–2972.
Dougherty, R.H., Sidhu, S.S., Raman, K., Solon, M., Solberg, O.D., Caughey, G.H., et al., 2010. Accumulation of intraepithelial mast cells with a unique protease phenotype in T(H)2-high asthma. J. Allergy Clin. Immunol. 125 (5), 1046–10453 e8.
Fernandez, D., Pallares, I., Vendrell, J., Aviles, F.X., 2010. Progress in metallocarboxypeptidases and their small molecular weight inhibitors. Biochimie 92 (11), 1484–1500.
Fricker, L.D., Gomis-Rüth, F.X., 2013. Carboxypeptidase A4. Handbook of Proteolytic Enzymes, third ed. Academic Press, pp. 1307–1310. third ed.
Fricker, L.D., 2013. Carboxypeptidase A5. In: Rawlings NDS, G. (Ed.), Handbook of Proteolytic Enzymes, third ed. ed. Academic Press, pp. 1315–1317.
Fricker, L.D., Lyons, P.J., 2013. Carboxypeptidase A6. In: Rawlings NDS, G. (Ed.), Handbook of Proteolytic Enzymes, third ed. ed. Academic Press, pp. 1318–1321.

Fricker, M., Gibson, P.G., Powell, H., Simpson, J.L., Yang, I.A., Upham, J.W., et al., 2019. A sputum 6-gene signature predicts future exacerbations of poorly controlled asthma. J. Allergy Clin. Immunol. 144 (1), 51–60 e11.
Galli, S.J., Tsai, M., 2010. Mast cells in allergy and infection: versatile effector and regulatory cells in innate and adaptive immunity. Eur. J. Immunol. 40 (7), 1843–1851.
Galli, S.J., Grimbaldeston, M., Tsai, M., 2008. Immunomodulatory mast cells: negative, as well as positive, regulators of immunity. Nat. Rev. Immunol. 8 (6), 478–486.
Gallwitz, M., Hellman, L., 2006. Rapid lineage-specific diversification of the mast cell chymase locus during mammalian evolution. Immunogenetics 58 (8), 641–654.
Gallwitz, M., Reimer, J.M., Hellman, L., 2006. Expansion of the mast cell chymase locus over the past 200 million years of mammalian evolution. Immunogenetics 58 (8), 655–669.
Garcia-Guerrero, M.C., Garcia-Pardo, J., Berenguer, E., Fernandez-Alvarez, R., Barfi, G. B., Lyons, P.J., et al., 2018. Crystal structure and mechanism of human carboxypeptidase O: insights into its specific activity for acidic residues. Proc. Natl. Acad. Sci. U. S. A. 115 (17), E3932–E3939.
Garcia-Pardo, J., Tanco, S., Garcia-Guerrero, M.C., Dasgupta, S., Aviles, F.X., Lorenzo, J., et al., 2020. Substrate specificity and structural modeling of human carboxypeptidase Z: a unique protease with a frizzled-like domain. Int. J. Mol. Sci. 21 (22), 8663–8687.
Gardell, S.J., Craik, C.S., Clauser, E., Goldsmith, E.J., Stewart, C.B., Graf, M., et al., 1988. A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. J. Biol. Chem. 263 (33), 17828–17836.
Gentek, R., Ghigo, C., Hoeffel, G., Bulle, M.J., Msallam, R., Gautier, G., et al., 2018. Hemogenic endothelial fate mapping reveals dual developmental origin of mast cells. Immunity 48 (6), 1160–11671 e5.
Hallgren, J., Jones, T.G., Abonia, J.P., Xing, W., Humbles, A., Austen, K.F., et al., 2007. Pulmonary CXCR2 regulates VCAM-1 and antigen-induced recruitment of mast cell progenitors. Proc. Natl. Acad. Sci. U. S. A. 104 (51), 20478–20483.
Hellman, L., Thorpe, M., 2014. Granule proteases of hematopoietic cells, a family of versatile inflammatory mediators - an update on their cleavage specificity, in vivo substrates, and evolution. Biol. Chem. 395 (1), 15–49.
Hellman, L.T., Akula, S., Thorpe, M., Fu, Z., 2017. Tracing the origins of IgE, mast cells, and allergies by studies of wild animals. Front. Immunol. 8, 1749.
Heylen, E., Willemse, J., Hendriks, D., 2011. An update on the role of carboxypeptidase U (TAFIa) in fibrinolysis. Front Biosci (Landmark Ed) 16, 2427–2450.
Huang, R., Hellman, L., 1994. Genes for mast-cell serine protease and their molecular evolution. Immunogenetics 40 (6), 397–414.
Irani, A.M., Goldstein, S.M., Wintroub, B.U., Bradford, T., Schwartz, L.B., 1991. Human mast cell carboxypeptidase. Selective localization to MCTC cells. J. Immunol. 147 (1), 247–253.
Kalinina, E., Biswas, R., Berezniuk, I., Hermoso, A., Aviles, F.X., Fricker, L.D., 2007. A novel subfamily of mouse cytosolic carboxypeptidases. Faseb. J. 21 (3), 836–850.
Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. 20 (4), 1160–1166.
Li, Z., Liu, S., Xu, J., Zhang, X., Han, D., Liu, J., et al., 2018. Adult connective tissue-resident mast cells originate from late erythro-myeloid progenitors. Immunity 49 (4), 640–653 e5.
Lutzelschwab, C., Pejler, G., Aveskogh, M., Hellman, L., 1997. Secretory granule proteases in rat mast cells. Cloning of 10 different serine proteases and a carboxypeptidase A from various rat mast cell populations. J. Exp. Med. 185 (1), 13–29.
Lyons, P.J., Fricker, L.D., 2011. Carboxypeptidase O is a glycosylphosphatidylinositol-anchored intestinal peptidase with acidic amino acid specificity. J. Biol. Chem. 286 (45), 39023–39032.
Mao, S.S., Cooper, C.M., Wood, T., Shafer, J.A., Gardell, S.J., 1999. Characterization of plasmin-mediated activation of plasma procarboxypeptidase B. Modulation by glycosaminoglycans. J. Biol. Chem. 274 (49), 35046–35052.
Metcalfe, D.D., Baram, D., Mekori, Y.A., 1997. Mast cells. Physiol. Rev. 77 (4), 1033–1079.
Natsuaki, M., Stewart, C.B., Vanderslice, P., Schwartz, L.B., Natsuaki, M., Wintroub, B.U., et al., 1992. Human skin mast cell carboxypeptidase: functional characterization, cDNA cloning, and genealogy. J. Invest. Dermatol. 99 (2), 138–145.
Pejler, G., 2019. The emerging role of mast cell proteases in asthma. Eur. Respir. J. 54 (4), 1 900 668–1900685.
Pejler, G., Abrink, M., Ringvall, M., Wernersson, S., 2007. Mast cell proteases. Adv. Immunol. 95, 167–255.
Pejler, G., Knight, S.D., Henningsson, F., Wernersson, S., 2009. Novel insights into the biological function of mast cell carboxypeptidase A. Trends Immunol. 30 (8), 401–408.
Pejler, G., Ronnberg, E., Waern, I., Wernersson, S., 2010. Mast cell proteases: multifaceted regulators of inflammatory disease. Blood 115 (24), 4981–4990.
Petrera, A., Lai, Z.W., Schilling, O., 2014. Carboxyterminal protein processing in health and disease: key actors and emerging technologies. J. Proteome Res. 13 (11), 4497–4504.
Piliponsky, A.M., Chen, C.C., Nishimura, T., Metz, M., Rios, E.J., Dobner, P.R., et al., 2008. Neurotensin increases mortality and mast cells reduce neurotensin levels in a mouse model of sepsis. Nat. Med. 14 (4), 392–398.
Quesada, V., Ordonez, G.R., Sanchez, L.M., Puente, X.S., Lopez-Otin, C., 2009. The Degradome database: mammalian proteases and diseases of proteolysis. Nucleic Acids Res. 37 (Database issue), D239–D243.
Rawlings, N.D., 2020. Twenty-five years of nomenclature and classification of proteolytic enzymes. Biochim. Biophys. Acta Protein Proteonomics 1868 (2), 140345.
Reimer, J.M., Samollow, P.B., Hellman, L., 2010. High degree of conservation of the multigene tryptase locus over the past 150-200 million years of mammalian evolution. Immunogenetics 62 (6), 369–382.
Reynolds, D.S., Gurley, D.S., Stevens, R.L., Sugarbaker, D.J., Austen, K.F., Serafin, W.E., 1989. Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison of the protein with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases. Proc. Natl. Acad. Sci. U. S. A. 86 (23), 9480–9484.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574.
Sapio, M.R., Fricker, L.D., 2014. Carboxypeptidases in disease: insights from peptidomic studies. Proteonomics Clin. Appl. 8 (5–6), 327–337.
Schneider, L.A., Schlenner, S.M., Feyerabend, T.B., Wunderlin, M., Rodewald, H.R., 2007. Molecular mechanism of mast cell mediated innate defense against endothelin and snake venom sarafotoxin. J. Exp. Med. 204 (11), 2629–2639.
Schwartz, L.B., Irani, A.M., Roller, K., Castells, M.C., Schechter, N.M., 1987. Quantitation of histamine, tryptase, and chymase in dispersed human T and TC mast cells. J. Immunol. 138 (8), 2611–2615.
Sela, I., Ashkenazy, H., Katoh, K., Pupko, T., 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 43 (W1), W7–W14.
Siddhuraj, P., Clausson, C.M., Sanden, C., Alyamani, M., Kadivar, M., Marsal, J., et al., 2021. Lung mast cells have a high constitutive expression of carboxypeptidase A3 mRNA that is independent from granule-stored CPA3. Cells 10 (2), 292–309.
Skidgel, R.A., Erdos, E.G., 1998. Cellular carboxypeptidases. Immunol. Rev. 161, 129–141.
Tanco, S., Zhang, X., Morano, C., Aviles, F.X., Lorenzo, J., Fricker, L.D., 2010. Characterization of the substrate specificity of human carboxypeptidase A4 and implications for a role in extracellular peptide processing. J. Biol. Chem. 285 (24), 18385–18396.
Tanco, S., Lorenzo, J., Garcia-Pardo, J., Degroeve, S., Martens, L., Aviles, F.X., et al., 2013. Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases. Mol. Cell. Proteomics 12 (8), 2096–2110.
Tanco, S., Aviles, F.X., Gevaert, K., Lorenzo, J., Van Damme, P., 2017. Identification of carboxypeptidase substrates by C-terminal COFRADIC. Methods Mol. Biol. 1574, 115–133.
Waern, I., Taha, S., Lorenzo, J., Montpeyo, D., Covaleda-Cortes, G., Aviles, F.X., et al., 2021. Carboxypeptidase inhibition by NvCI suppresses airway hyperreactivity in a mouse asthma model. Allergy 76 (7), 2234–2237.
Wernersson, S., Pejler, G., 2014. Mast cell secretory granules: armed for battle. Nat. Rev. Immunol. 14 (7), 478–494.